This algorithm calculates the belief distribution bel from measurement and control data.

In pseudo-algorithmic form:

$AlgorithmBayesFilter(bel(x_{t-1}), u_t, z_t)$:
for all x_t do:
$\overline{bel}(x_t) = \int p(x_t \mid u_t, x_{t-1})\ bel(x_{t-1})\ dx_{t-1}$
$bel(x_t) = \eta\, p(z_t \mid x_t)\ \overline{bel}(x_t)$
return $bel(x_t)$

The Bayes Filter is recursive, that is, the belief $bel(x_t)$ at time t is calculated from the belief $bel(x_{t-1})$ at time $t-1$. Its input is the belief bel at time $t-1$, along with the most control u_t and the most recent measurement z_t . Its output is the belief $bel(x_t)$ at time t

This is called the update rule of the Bayes Filter algorithm and this rule is applied recursively.

The belief $\overline{bel}(x_t)$ that the robot assigns the state x_t is obtained by the integral (sum) of the product of two distributions: the prior assigned to x_{t-1} and the probability that control u_t induces a transition from x_{t-1} to x_t

This step is called the prediction.

The second step is called the measurement update. This multiplies the belief $\overline{bel}(x_t)$ by the probability that the measurement z_t may have been observed. The result is normalised by a normalisation constant η.

To compute the posterior belief recursively, the algorithm requires an initial belief $bel(x_0)$ at time $t = 0$ as a boundary condition. If we know the value of x_0 with certainty, $bel(x_0)$ should be initialised with a point mass distribution that centers all probability mass on the correct value of x_0 and against zero probability anywhere else. If not, the belief may be initialised using a uniform distribution over the domain of x_0.

Full derivation

We can prove the correctness of the Bayes filter by induction:

First, we apply Bayes rule to the target posterior:

$$\begin{aligned} p(x_t \mid z_{1:t}, u_{1:t}) &= \frac{p(z_t \mid x_t, z_{1:t-1}, u_{1:t})\ p(x_t \mid z_{1:t-1}, u_{1:t})}{p(z_t \mid x_{t,} z_{1:t-1}, u_{1:t})} \\ &= \eta\, p(z_t \mid x_t, z_{1:t-1}, u_{1:t})\ p(x_t \mid z_{1:t-1}, u_{1:t}) \end{aligned}$$

i.e. applying Bayes rule, the posterior is equal to the product of the prior belief $p(x_t \mid z_{1:t-1}, u_{1:t})$ and the observation model $p(z_t \mid x_t, z_{1:t-1}, u_{1:t})$, normalized by η

Assuming, that our state is complete, we can express the following conditional independence:

$$p(z_t \mid x_t, z_{1:t-1}, u_{1:t}) = p(z_t \mid x_t)$$

i.e. If we are predicting the measurement given the state, no past measurement or control would provide us any additional information since the state is complete.

This means that:

$$p(x_t \mid z_{1:t}, u_{1:t}) = \eta\, p(z_t \mid x_t)\; p(x_t \mid z_{1:t-1}, u_{1:t})$$

i.e. the posterior equals the product of the observation model and the prediction (posterior without z_t), normalized by η

$$bel(x_t) = \eta\, p(z_t \mid x_t)\; \overline{bel}(x_t)$$

Using the theorem of total probability Probability and Expectation algebra > Basic probability rules we can expand the prediction:

$$\begin{aligned}\overline{bel}(x_t) &= p(x_t \mid z_{1:t-1}, u_{1:t}) \\ &= \int p(x_t \mid x_{t-1}, z_{1:t-1}, u_{1:t})\; p(x_{t-1} \mid z_{1:t-1}, u_{1:t})\; dx_{t-1}\end{aligned}$$

Here, the theorem of total probability allows us to marginalize over all possible values of the previous state.

The state depends probabilistically on the previous state, thus to compute $\overline{bel}(x_t)$ we need to take into account all possible values of x_{t-1}. The Theoreom of Total Probability ensures that we handle this marginalization correctly:

$$\overline{bel}(x_t) = \int (\text{transition from } x_{t-1} \text{ to } x_t) \cdot (\text{belief about } x_{t-1})\; dx_{t-1}$$

Exploiting the assumption that our state is complete, we can simplify this:

$$\begin{aligned}\overline{bel}(x_t) &= \int p(x_t \mid x_{t-1}, z_{1:t-1}, u_{1:t})\; p(x_{t-1} \mid z_{1:t-1}, u_{1:t})\; dx_{t-1} \\ &= \int p(x_t \mid x_{t-1}, u_{1:t})\; p(x_{t-1} \mid z_{1:t-1}, u_{1:t})\; dx_{t-1} \\ &= \int p(x_t \mid x_{t-1}, u_{1:t})\; bel(x_{t-1})\end{aligned}$$

The Markov Assumption

The Markov Assumption, or complete state assumption, postulates that past and future data are independent if one knows the current state x_t.

Gaussian Filters

Gaussian filters constitute the earliest tractable implementations of the Bayes filter for continuous spaces.

Gaussian filters share the basic idea that beliefs are represented by multivariate normal distributions.

Kalman Filters

The Kalman Filter implements beliefs by the moments parameterization. At time t, the belief is represented by the mean μ_t and the covariance Σ_t. Posteriors are Gaussian if the following three properties hold, in addition to the Markov assumptions of the Bayes filter:

1. The state transition probability must be a linear function in its arguments with added Gaussian noise, i.e.

   $$x_t = A_t \, x_{t-1} + B_t \, u_t + \epsilon_t$$

   where:

- x_t is the state vector at time t
- u_t is the control vector at time t
- A_t is a square matrix of size $n \times n$ where n is the dimension of the state vector
- B_t is a matrix of size $n \times m$ where n is the dimension of the control vector

By multiplying the state and control vector with the A_t and B_t matrices, respectively, the state transition function becomes linear in its arguments.

The term $A_t x_{t-1}$ represents a matrix-vector multiplication. For example, if x_{t-1} is 2-dimensional:

$$A_t = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad x_{t-1} = \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \end{bmatrix} \$\$Then:$$

\begin{align*} A t x_{t-1} &= \begin{bmatrix} a_{11} x_{t-1}^{(1)} + a_{12} x_{t-1}^{(2)} \\ a_{21} x_{t-1}^{(1)} + a_{22} x_{t-1}^{(2)} \end{bmatrix} \\
\end{align*}
$$ Each element of x_t is a weighted sum of the elements of x_{t-1}, which is a linear operation.

The same holds for $B_t u_t$. The term ϵ_t is simply added to the sum of the linear components. Since addition does not break linearity, the overall transition remains linear. There are no nonlinear terms like squares, products or transcendental functions involved

Thus, Kalman filters are said to assume linear state dynamics.

The random variable ϵ_t is a Gaussian random vector that models the uncertainty introduced by the state transition. Its mean is 0 and its covariance is denoted R_t.

A state transition of this form $x_t = A_t\ x_{t-1} + B_t\ u_t + \epsilon_t$ is called a linear Gaussian, reflect that it is linear in its arguments with additive Gaussian noise.

The state transition probability $x_t = A_t\ x_{t-1} + B_t\ u_t + \epsilon_t$ denotes

$$p(x_t \mid u_t, x_{t-1})$$

This is obtained by plugging it into the definition of the multivariate normal distribution Multivariate normal distribution > Multivariate normal distribution. The mean of the posterior state is given by $x_t = A_t\ x_{t-1} + B_t\ u_t$ and the covariance by R_t:

$$p(x_t \mid u_t, x_{t-1}) = \det(2\pi R_t)^{-1/2} \exp\left\{-\frac{1}{2}(x_t - A_t\ x_{t-1} - B_t\ u_t)^T R_t^{-1}\ (x_t - A_t\ x_{t-1} - B_t\ u_t)\right\}$$

This formulation shows that the transition model combines linear deterministic dynamics (via A_t and B_t) with additive uncertainty (via the Gaussian noise ϵ_t). The result is a linear Gaussian model, where the state transitions follow a multivariate normal distribution characterized by the mean $\mu_t = A_t x_{t-1} + B_t u_t$ and the covariance R_t.

2. The measurement probability must be linear in its arguments with added Gaussian noise

   $$z_t = C_t\ x_t + \delta_t$$

   where:

- z_t is the measurement vector
- C_t is a matrix of size $k \times n$ where k is the dimension of the measurement vector
- δ_t is the measurement noise vector

The distribution of δ_t is a multivariate Gaussian with zero mean and covariance Q_t. The measurement probability is thus given by the following multivariate normal distribution:

$$p(z_t\ |x_t) = \det(2\pi Q_t)^{-1/2} \exp\left\{-\frac{1}{2}(z_t - C_t\ x_t)^T Q_t^{-1}\ (z_t - C_t\ x_t)\right\}$$

This formulation shows that the measurement model is linear in its arguments $C_t\ x_t$ with additive Gaussian noise δ_t The resulting measurement probability $p(z_t\ |x_t)$ is a multivariate Gaussian distribution.

This is the second key assumption of the Kalman Filter and ensures that the observed measurements are probabilistically related to the state in a mathematically tractable way.

3. The initial belief $bel(x_0)$ must be normally distributed, with mean μ_0 and covariance Σ_0

   $$bel(x_0) = p(x_0) = \det(2\pi\Sigma_0)^{-1/2} \exp\left\{-\frac{1}{2}(x_0 - \mu_0)^T \Sigma_0^{-1}\ (x_0 - \mu_0)\right\}$$

   The initial belief is required to be normally distributed to ensure the Gaussian assumptions of the filter hold throughout the process. It ensures that the Kalman Filter

can initialise its recursive process in a mathematically tractable way and that all subsequent beliefs (posteriors) remain Gaussian (given the assumptions of linear dynamics and Gaussian noise).

The three above assumptions are sufficient to ensure that the posterior $bel(x_t)$ is always a Gaussian for any point in time t.

Deriving the KF algorithm

KFs represent the belief $bel(x_t)$ at time t by the mean μ_t and covariance Σ_t. The input to the algorithm is the belief at time $t-1$, represented by by the mean μ_{t-1} and covariance Σ_{t-1}.

To update these parameters we require the control u_t and measurement Σ_t.

The predicted belief $\overline{\mu}$ and $\overline{\Sigma}$ is calculated representing $\overline{bel}(x_t)$ one time step later, but before incorporating the measurement z_t. It is obtained by incorporating the control u_t. The mean is updated using the deterministic version of the state transition function with the mean μ_{t-1} substituted for the state x_{t-1}.

$$\overline{\mu}_t = A_t\ \mu_{t-1} + B_t\ u_t + \epsilon_t$$

The update of the covariance considers the fact that states depend on previous states through the linear matrix A_t. The matrix is multiplied twice into the covariance, since the covariance is a quadratic matrix.

$$\overline{\Sigma}_t = A_t\ \Sigma_{t-1}\ A_t^T + R_t$$

We the transform $\overline{bel}$ into the bel by incorporating the measurement z_t.

$$\mu_t = \overline{\mu}_t + K_t(z_t - C_t\ \overline{\mu}_t)$$

$$K_t$$

represents the Kalman gain. It specifies the degree to which the measurement is incorporated into the new state estimate.

A high Kalman Gain, means that the new state estimate would be closer to the measurement.

$$(z_t - C_t\ \overline{\mu}_t)$$

represents the innovation. This is the difference between the actual measurement z_t and the predicted measurement $C_t\ \overline{\mu}_t$

The new covariance of the posterior belief is then calculated, adjusting for the information gain resulting from the measurement.

Thus, the Kalman Filter algorithm for linear Gaussian state transitions and measurements becomes:

$Algorithm Kalman Filter(\mu_{t-1}, \Sigma_{t-1}, u_t, z_t)$:

```
$\overline{\mu}_{t} = A_{t} \ \mu_{t-1} + B_{t } \ u_{t} + \epsilon_{t}$
$\overline{\Sigma}_{t} = A_{t} \ \Sigma_{t-1} \ A_{t}^T + R_{t}$

$K_{t} = \overline{\Sigma}_{t} \ C_{t}^T (C_{t} \ \overline{\Sigma}_{t} \
C_{t}^T + Q_{t})-1$

$\mu_{t} = \overline{\mu}_{t} + K_{t} (z_{t} - C_{t} \
\overline{\mu}_{t})$
$\Sigma_{t} = (I - K_{t} \ C_{t}) \ \overline{\Sigma}_{t}$
$\text{return } \mu_{t}, \Sigma_{t}$
```

Proving the Kalman Filter from the Bayes Filter

Proving the prediction step

The prediction step of the Bayes filter is:

$$\overline{bel}(x_t) = \int p(x_t \mid u_t, x_{t-1}) \; bel(x_{t-1}) \; dx_{t-1}$$

This is derived from the theory of total probability. We say that the probability distribution of our prediction is the infinite sum of (probability of a state given the previous state and the new control input) ANDed with (probability of the previous state), varying over probabilities of the previous state. So we consider how the previous state + new control can lead us to a prediction while taking into account that the previous state is also given by a probability distribution.

For the Kalman Filter we have:

$$p(x_t \mid u_t, x_{t-1}) \sim \mathcal{N}(x_t; A_t \; x_{t-1} + B_t \; u_t, R_t)$$

$$bel(x_{t-1}) \sim \mathcal{N}(x_{t-1}; \mu_{t-1}, \Sigma_{t-1})$$

i.e. the probability of the prediction is a multivariate Gaussian distribution, with the mean given by a linear combination of the previous state and the control, and with covariance given by R_t

We know that since $\overline{bel}(x_t)$ is a Gaussian with mean $\overline{\mu}_t$ and covariance $\overline{\Sigma}_t$, we can write it as, using the definition for the multivariate Gaussian distribution

$$\overline{bel}(x_t) = \eta \int \exp \left\{ -\frac{1}{2} (x_t - A_t \; x_{t-1} - B_t \; u_t)^T R_t^{-1} \; (x_t - A_t \; x_{t-1} - B_t \; u_t) \right\}$$
$$\exp \left\{ -\frac{1}{2} (x_{t-1} - \mu_{t-1})^T \Sigma_t^{-1} \; (x_{t-1} - \mu_{t-1}) \right\} dx_{t-1}$$

which we simplify to:

$$\overline{bel}(x_t) = \eta \int \exp\{-L_t\}\, dx_{t-1}$$

with

$$L_t = -\frac{1}{2}(x_t - A_t\, x_{t-1} - B_t\, u_t)^T R_t^{-1}\, (x_t - A_t\, x_{t-1} - B_t\, u_t) + \frac{1}{2}(x_{t-1} - \mu_{t-1})^T \Sigma_t^{-1}\, (x_{t-1} - \mu_{t-1})$$

We can simplify this, by recognising that L_t is a function with terms dependent on both x_t and x_{t-1} and terms that only depend on x_t:

$$\begin{aligned} L_t &= L_t(x_{t-1}, x_t) + L_t(x_t) \\ \implies \overline{bel}(x_t) &= \eta \int \exp\left\{-L_t(x_{t-1}, x(t)) - L_t(x_t)\right\} dx_{t-1} \\ &= \eta \exp\left\{-L_t(x_t)\right\} \int \exp\left\{-L_t(x_{t-1}, x_t)\right\} dx_{t-1} \end{aligned}$$

The integral $\int \exp\left\{-L_t(x_{t-1}, x_t)\right\} dx_{t-1}$ represents a Gaussian integral over x_{t-1} which evaluates to a constant independent of x_t which can be absorbed into a new normalization factor η, different from the one above

$$\overline{bel}(x_t) = \eta \exp\{-L_t(x_t)\}$$

Calculating the first and second derivatives of L_t gives us an expression for $L_t(x_{t-1}, x_t)$ which is a quadratic in x_{t-1}:

$$\begin{aligned} \frac{\partial L_t}{\partial x_{t-1}} &= -A_t^T R_t^{-1}(x_t - A_t\, x_{t-1} - B_t\, u_t) + \Sigma_{t-1}^{-1}(x_{t-1} - \mu_{t-1}) \\ \frac{\partial^2 L_t}{\partial x_{t-1}^2} &= -A_t^T R_t^{-1} A_t + \Sigma_{t-1}^{-1} =: \Psi_t^{-1} \end{aligned}$$

Ψ_t defines the curvature of $L_t(x_{t-1}, x_t)$

and setting the first derivative to 0, gives us a an expression for the mean:

$$A_t^T R_t^{-1}(x_t - A_t\, x_{t-1} - B_t\, u_t) = \Sigma_{t-1}^{-1}(x_{t-1} - \mu_{t-1})$$

If L_t is the (probability of a state given the previous state and the new control input) ANDed with (probability of the previous state), then its first derivative with respect to x_{t-1} represents how much this probability changes as we vary x_{t-1}. Where this derivative is 0, we have the mean, i.e. the most likely estimate of x_t given the previous state and the control input.

The first derivative tells us the direction in which the probability of our estimate for x_t increases given the previous state and control input.

The second derivative tells us how the first derivative changes with respect to x_{t-1}, a large second derivative, i.e. one with a sharp peak, means we are confident about our estimate based on the previous state and control input. This makes sense, as if we really know the x_t is a particular value given a control and a previous state then it makes sense that the probabilities around it would be small but at that point the probability would be high and we would have sharp peak.

The second derivative, also called the inverse of the Hessian Ψ_t thus describes the spread of uncertainty about our prediction, which we can inerpret as the curvature of the probability distribution.

The inverse of the Hessian depends on the weight we give to the previous state in the prediction step A_t , the covariance of the probability of the state estimate given the previous state and a control input R_t and the covariance of the belief in the previous state estimate Σ_{t-1}

Through manipulation of our expression for the mean we can derive that:

$$\begin{aligned} L_t(x_{t-1}, x_t) =& \frac{1}{2}\left(x_{t-1} - \Psi_t \left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right]\right)^T \Psi^{-1} \\ & \left(x_{t-1} - \Psi_t \left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right]\right) \end{aligned}$$

So we can determine:

$$\begin{aligned} L_t(x_t) =& L_t - L_t(x_{t-1}, x_t) \\ =& -\frac{1}{2}(x_t - A_t\, x_{t-1} - B_t\, u_t)^T R_t^{-1}\, (x_t - A_t\, x_{t-1} - B_t\, u_t) + \frac{1}{2}(x_{t-1} - \mu_{t-1})^T \Sigma_t^{-1}\, (x_{t-1} - \mu_{t-1} \\ & - \frac{1}{2}\left(x_{t-1} - \Psi_t \left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right]\right)^T \Psi^{-1} \left(x_{t-1} - \Psi_t \left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right]\right) \end{aligned}$$

By substituting back $\Psi_t = (-A_t^T R_t^{-1} A_t + \Sigma_{t-1}^{-1})^{-1}$ we can show that, as expected from our construction of $L_t(x_{t-1}, x_t)$, L_t is not dependent on x_{t-1}

$$\begin{aligned} L_t(x_t) =& \frac{1}{2}(x_t - B_t\, u_t)^T R_t^{-1}\, (x_t - B_t\, u_t) + \frac{1}{2}\mu_{t-1}^T \Sigma_t^{-1} \mu_{t-1} \\ & - \frac{1}{2}\left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right]^T \left(A_t^T R_t^{-1} A_t + \Sigma_{t-1}^{-1}\right)^{-1} \\ & \left[A_t^T R_t^{-1}(x_t - B_t u_t) + \Sigma_{t-1}^{-1}\, \mu_{t-1}\right] \end{aligned}$$

We can see that $L_t(x_t)$ is quadratic in x_t, meaning that the prediction is indeed normally distributed as we required with the mean and covariance of this distribution being given by the minimum and curvature of $L_t(x_t)$.

$$\begin{aligned} \frac{\partial L_t}{\partial x_t} =& \left(R_t + A_t \Sigma_{t-1} A_t^T\right)^{-1} (x_t - B_t\, u_t) \\ & - R_t^{-1} A_t (A_t^T R_t^{-1} A_t + \Sigma_{t-1}^{-1})^{-1}\, \Sigma_{t-1}^{-1} \mu_{t-1} = 0 \end{aligned}$$

$$\left(R_t + A_t \Sigma_{t-1} A_t^T\right)^{-1} (x_t - B_t\, u_t) = R_t^{-1} A_t (A_t^T R_t^{-1} A_t + \Sigma_{t-1}^{-1})^{-1}\, \Sigma_{t-1}^{-1} \mu_{t-1}$$

Solving for x_t through some easy cancellations, we get:

$$x_t = B_t\, u_t + A_t\, \mu_{t-1}$$

i.e. the mean of the belief after incorporating the control is $B_t\, u_t + A_t\, \mu_{t-1}$

This proves the correctness of line 2 of the Kalman filter algorithm

$$\frac{\partial^2 L_t}{\partial x_t^2} = (A_t\, \Sigma_{t-1}\, A_t^T + R_t)^{-1}$$

This gives the curvature of L_t, whose inverse is the covariance of the belief, proving line 3 of the algorithm.

Proving the measurement update step

The measurement update step of the Kalman Filter is:

$$bel(x_t) = \eta\, p(z_t \mid x_t)\, \overline{bel}(x_t)$$

We multiply the belief $\overline{bel}(x_t)$ by the probability that the measurement z_t may have been observed. The result is normalised by a normalisation constant η.

For the Kalman Filter we have:

$$p(z_t \mid x_t) \sim \mathcal{N}(z_t;\, C_t\, x_t,\, Q_t)$$

$$\overline{bel}(x_t) \sim \mathcal{N}(x_t;\, \overline{\mu}_t,\, \overline{\Sigma}_t)$$

The mean and covariance of the prediction are given by $\overline{\mu}_t \text{ and } \overline{\Sigma}_t$. The measurement probability is defined to be normal as well, with mean $C_t\, x_t$ and covariance Q_t

Thus, we have:

$$belx_t = \eta \exp\{-J_t\}$$

$$J_t = \frac{1}{2}(z_t - C_t\, x_t)\, Q_t^{-1}\, (z_t - C_t\, x_t) + \frac{1}{2}(x_t - \overline{\mu}_t)^T\, \overline{\Sigma}_t^{-1}\, (x_t - \overline{\mu}_t)$$

We can see that this is a quadratic in x_t hence the belief is a Gaussian.

The minimum and curvature of J_t will give us the mean and covariance of this probability distribution.

$$\frac{\partial J}{\partial x_t} = -C_t^T Q_t^{-1}(z_t - C_t x_t) + \overline{\Sigma}_t^{-1}(x_t - \overline{\mu}_t) = 0$$

$$\begin{aligned}
C_t^T Q_t^{-1}(z_t - C_t x_t) &= \overline{\Sigma}_t^{-1}(\mu_t - \overline{\mu}_t) \\
C_t^T Q_t^{-1}(z_t - C_t x_t + C_t\, \overline{\mu}_t - C_t\, \overline{\mu}_t) &= \overline{\Sigma}_t^{-1}(\mu_t - \overline{\mu}_t) \\
C_t^T Q_t^{-1}(z_t - C_t x_t + C_t\, \overline{\mu}_t - C_t\, \overline{\mu}_t) &= \overline{\Sigma}_t^{-1}(\mu_t - \overline{\mu}_t) \\
C_t^T Q_t^{-1}\,(z_t - C_t\, \overline{\mu}_t) - C_t^T Q_t^{-1} C_t\,(\mu_t - \overline{\mu}_t) &= \overline{\Sigma}_t^{-1}(\mu_t - \overline{\mu}_t) \\
C_t^T Q_t^{-1}\,(z_t - C_t\, \overline{\mu}_t) &= \left(C_t^T Q_t^{-1} C_t + \overline{\Sigma}_t^{-1}\right)(\mu_t - \overline{\mu}_t) \\
C_t^T Q_t^{-1}\,(z_t - C_t\, \overline{\mu}_t) &= \left(\Sigma_t^{-1}\right)(\mu_t - \overline{\mu}_t) \\
\Sigma_t C_t^T Q_t^{-1}\,(z_t - C_t\, \overline{\mu}_t) &= (\mu_t - \overline{\mu}_t)
\end{aligned}$$

$$\frac{\partial^2 J}{\partial x_t^2} = C_t^T\, Q_t^{-1}\, C_t + \overline{\Sigma}_t^{-1}$$

we define the Kalman Gain as

$$K_t = \Sigma_t C_t^T Q_t^{-1}$$

$$\begin{aligned} \implies K_t \left(z_t - C_t \, \overline{\mu}_t\right) &= \left(\mu_t - \overline{\mu}_t\right) \\ \mu_t &= \overline{\mu}_t + K_t \left(z_t - C_t \, \overline{\mu}_t\right) \end{aligned}$$

This proves line 5 of the Kalman filter algorithm

The Kalman gain is defined in terms of Σ_t but we use the Kalman gain to calculate Σ_t so we should transform the Kalman gain to define it in terms of other covariances:

$$K_t = \overline{\Sigma}_t \, C_t^T \left(C_t \, \overline{\Sigma}_t \, C_t^T + Q_t\right)^{-1}$$

With this expression we also avoid inverting the state covariance matrix, which is essential to Kalman filters for high-dimensional state spaces

According to the Linear algebra > Inversion Lemma, we can:

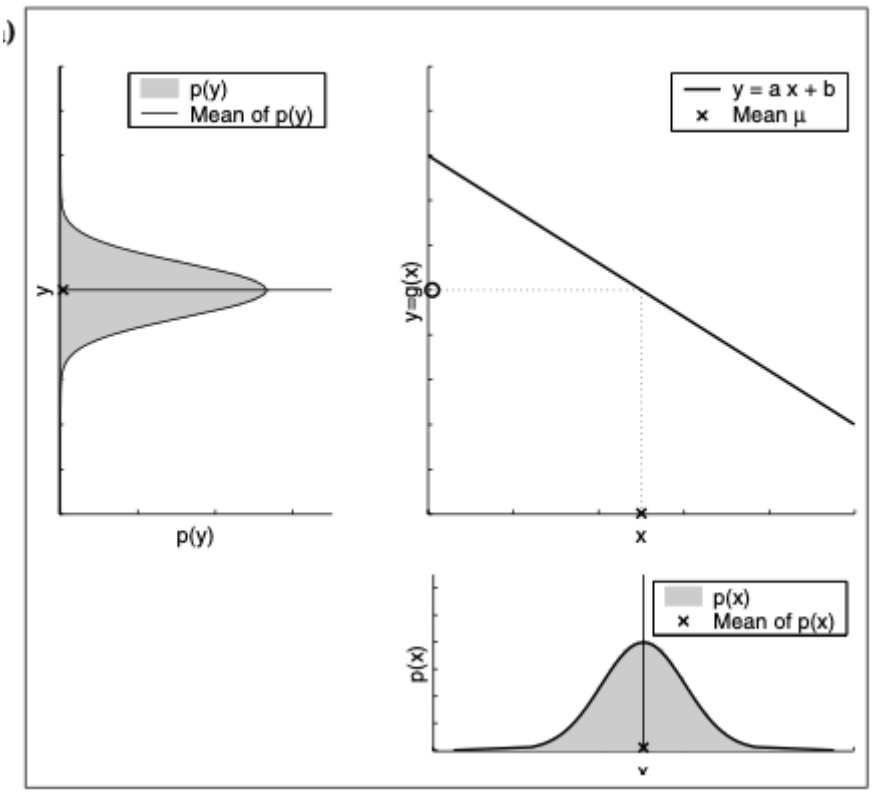
$$\begin{aligned} \left(\frac{\partial^2 L_t}{\partial x_t^2}\right)^{-1} = \Psi^{-1} = \Sigma_t &= \left(C_t^T \, Q_t^{-1} \, C_t + \overline{\Sigma}_t^{-1}\right)^{-1} \\ &= \overline{\Sigma}_t - \overline{\Sigma}_t \, C_t^T \left(Q_t + C_t \, \overline{\Sigma}_t \, C_t^T\right)^{-1} C_t \, \overline{\Sigma}_t \\ &= \left[I - \overline{\Sigma}_t \, C_t^T \left(Q_t + C_t \, \overline{\Sigma}_t \, C_t^T\right)^{-1} C_t\right] \overline{\Sigma}_t \\ &= \left(I - K_t \, C_t\right) \overline{\Sigma}_t \end{aligned}$$

which proves the correctness of line 6 of the Kalman Filter

Extended Kalman Filters

The assumptions that observations are linear functions of the state and that the next state is a linear function of the previous state are crucial for the correctness of the Kalman filter.

Lets say we have a Gaussian random variable $X \sim \mathcal{N}(x;\ \mu,\ \sigma^2)$. If we pass X through a linear function $y = ax + b$. The resulting variable is a Gaussian $Y \sim \mathcal{N}(y;\ a\mu + b,\ a^2\sigma^2)$. In graphical form this can be seen below

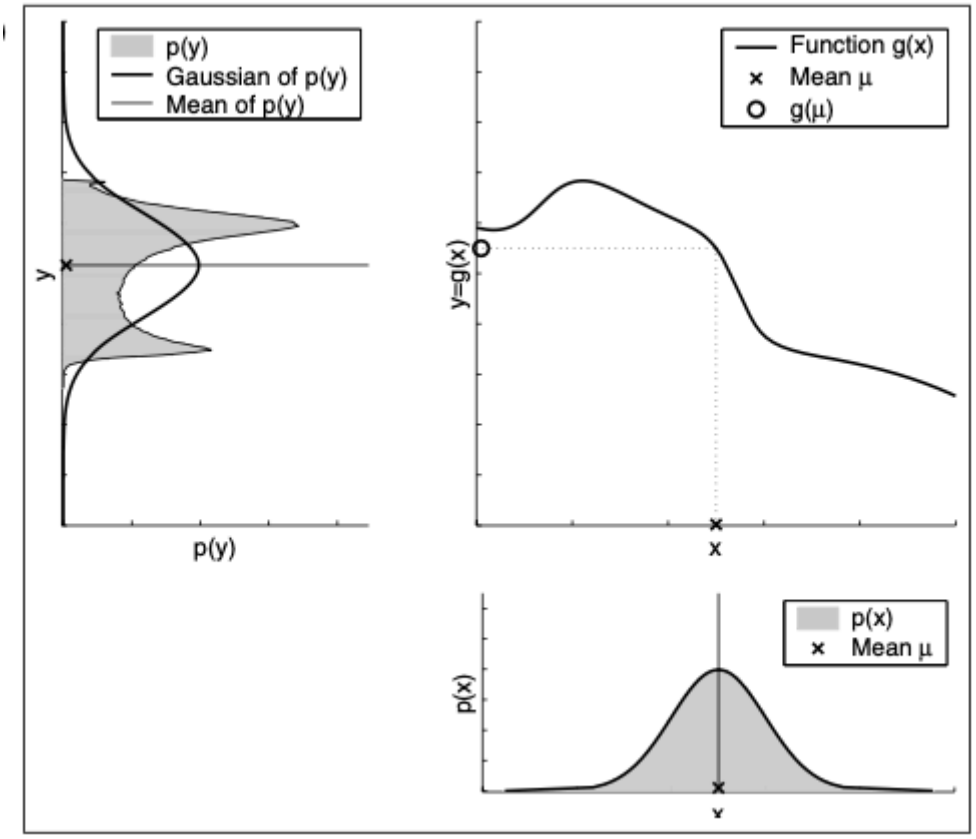


Unfortunately, state transitions and measurements are rarely linear in practice, rendering plain Kalman filters innaplicable to all but the most trivial robotics problems.

The EKF relaxes the linearity assumptions. It assumes that the state transition probability and the measurement probabilities are governed by nonlinear functions g and h:

$$x_t = g(u_t, x_{t-1}) + \epsilon_t$$
$$z_t = h(x_t) + \delta_t$$

With arbitrary functions g and h, the belief is no longer a Gaussian. Performing the belief update exactly is usually impossible for nonlinear functions g and h, and the Bayes filter does not possess a closed-form solution.

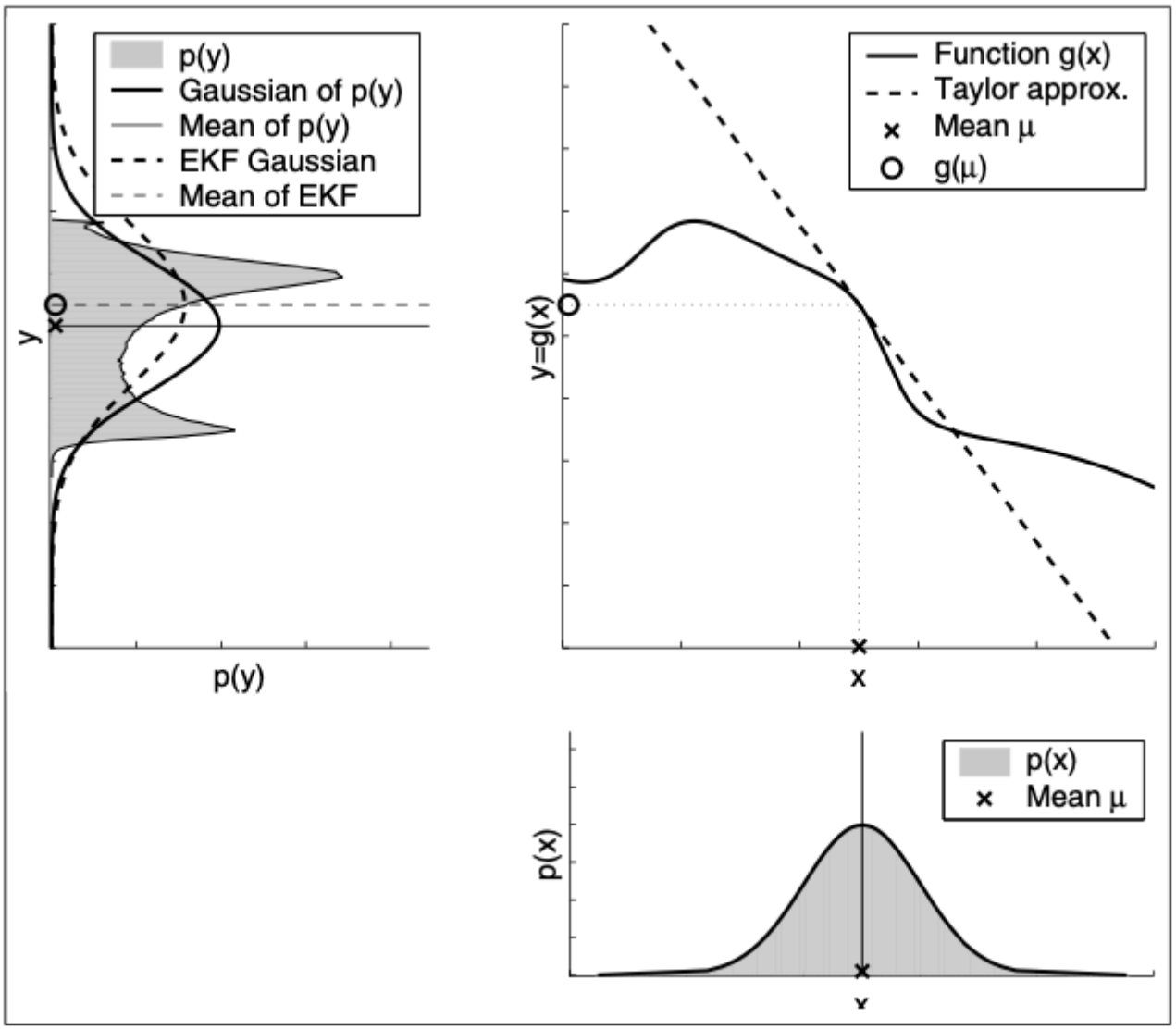


Thus the EKF calculates a Gaussian approximation to the true belief. Thus, the EKF inherits from the Kalman filter the basic belief representation, but it differs in that this belief is only approximate, not exact. The goal of the EKF thus shifts from computing the exact posterior to efficiently estimate its mean and covariance. However, since these cannot be computed in closed form, we use an approximation.

Linearization Via Taylor Expansion

Linearization is the key idea behind EKFs. The nonlinear functions representing the state transition and measurement probabilities are approximated by a linear function tangent to them.

g is approximated by a linear function that is tangent to g at the mean of the Gaussian. Projecting the Gaussian through this linear approximation results in a Gaussian density. The solid line in the graph below represents the mean and covariance of the Monte-Carlo approximation. The mismatch between these two Gaussians indicated the error caused by the linear approximation of g



EKFs use first order Taylor expansion to construct a linear approximation to a function g from g's value and slope. The slope is given by the partial derivative:

$$g'(u_t, x_{t-1}) = \frac{\partial g(u_t, x_{t-1})}{\partial x_{t-1}}$$

For plain KFs for state transition and measurement probabilities we use the most likely previous state to calculate the next most likely next state. And this most likely previous state is the posterior μ_{t-1}. So for our Taylor approximation, we will approximate g by its value at μ_{t-1} and at u_t.

We achieve the linear extrapolation by a term proportional to the gradient of g at μ_{t-1} and u_t:

$$\begin{aligned} g(u_t, x_{t-1}) &\approx g(u_t, \mu_{t-1}) + g'(u_t, \mu_{t-1})\,(x_{t-1} - \mu_{t-1}) \\ &= g(u_t, \mu_{t-1}) + G_t\,(x_{t-1} - \mu_{t-1}) \end{aligned}$$

As a Gaussian, the state transition probability becomes:

$$p(x_t \mid u_t, x_{t-1}) \approx \det(2\pi R_t)^{1/2} \exp\left\{ -\frac{1}{2}[x_t - g(u_t, \mu_{t-1}) - G_t(x_{t-1} - \mu_{t-1})]^T R_t^{-1} \left[x_t - g(u_t, \mu_{t-1}) - \right.\right.$$

Note that G_t is a $n \times n$ matrix where n is the size of the state vector. G_t is the Jacobian.

The same linearization is used for the measurement function h. The Taylor expansion is developed around $\overline{\mu}_t$, the state deemed most likely at the time we linearize h.

$$\begin{aligned} h(x_t) &\approx h(\overline{\mu}_t) + h'(\overline{\mu}_t)\,(x_t - \overline{\mu}_t) \\ &= h(\overline{\mu}_t) + H_t\,(x_t - \overline{\mu}_t) \end{aligned}$$

with

$$h'(x_t) = \frac{\partial h(x_t)}{\partial x_t}$$

Written as a Gaussian we have:

$$p(z_t \mid x_t) = \det(2\pi\, Q_t)^{-1/2} \exp\left\{-\frac{1}{2}[z_t - h(\overline{\mu}_t) - H_t(x_t - \overline{\mu}_t)]^T Q_t^{-1}\,[z_t - h(\overline{\mu}_t) - H_t(x_t - \overline{\mu}_t)]\right\}$$

$Algorithm Extended Kalman Filter(\mu_{t-1}, \Sigma_{t-1}, u_t, z_t)$:

```
$\overline{\mu}_{t} = g(u_{t}, \mu_{t-1})$
$\overline{\Sigma}_{t} = G_{t} \ \Sigma_{t-1} \ G_{t}^T + R_{t}$

$K_{t} = \overline{\Sigma}_{t} \ H_{t}^T (H_{t} \ \overline{\Sigma}_{t} \
H_{t}^T + Q_{t})-1$

$\mu_{t} = \overline{\mu}_{t} + K_{t} (z_{t} -  h(\overline{\mu}_{t}))$
$\Sigma_{t} = (I - K_{t} \ H_{t}) \ \overline{\Sigma}_{t}$
$\text{return } \mu_{t}, \Sigma_{t}$
```

This is the same as the KF algorithm except the linear predictions are replaced by their nonlinear generalizations.