Kalman Filter and Extended Kalman Filter (EKF)

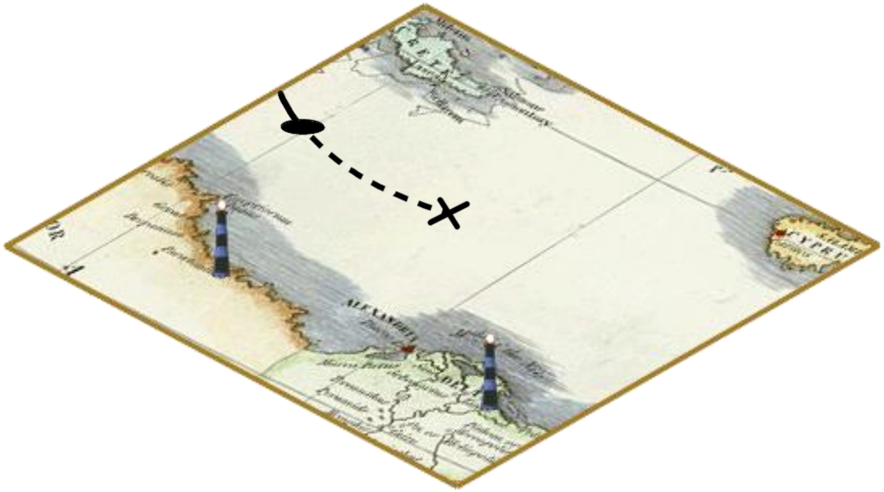
Nived Chebrolu

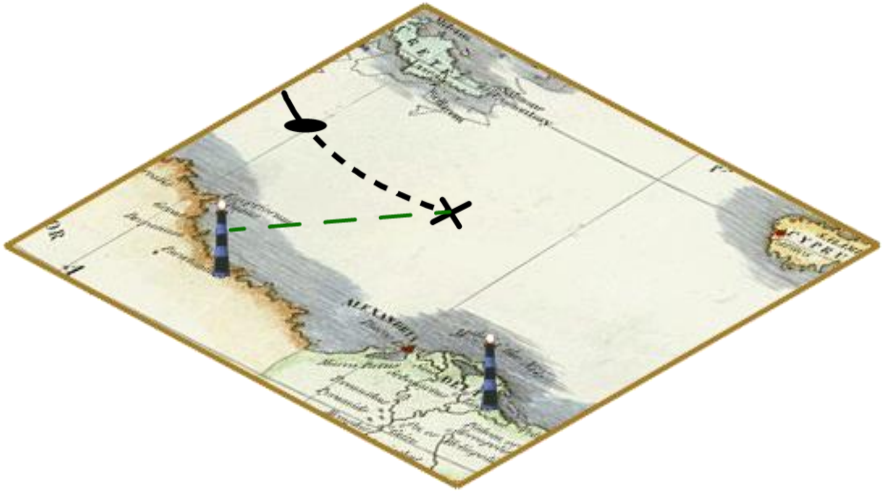
Slides adapted from Cyrill Stachniss at Uni. Bonn

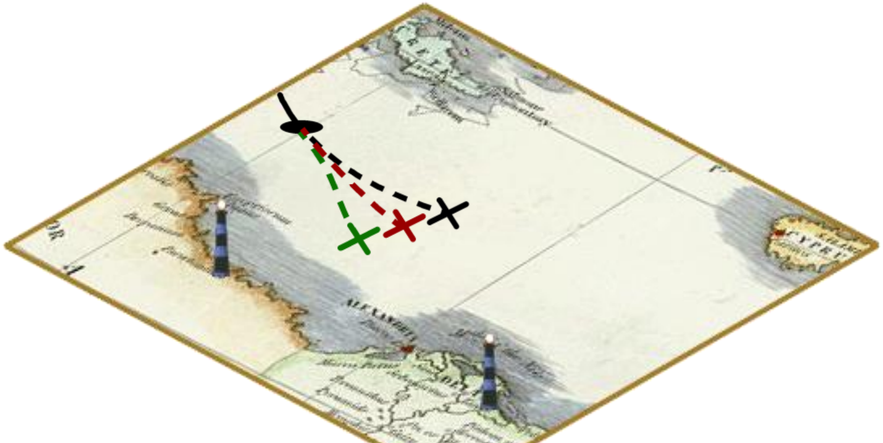
Kalman Filter

- **Kalman Filter is the Bayes filter for the Gaussian linear case**
- Performs recursive state estimation
- Prediction step to exploit the controls
- Correction step to exploit the observations

Kalman Filter Example







Courtesy: Thrun

Bayes Filter for State Estimation

- Bayes filter is one tool for state estimation
- **Prediction**

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

- **Correction**

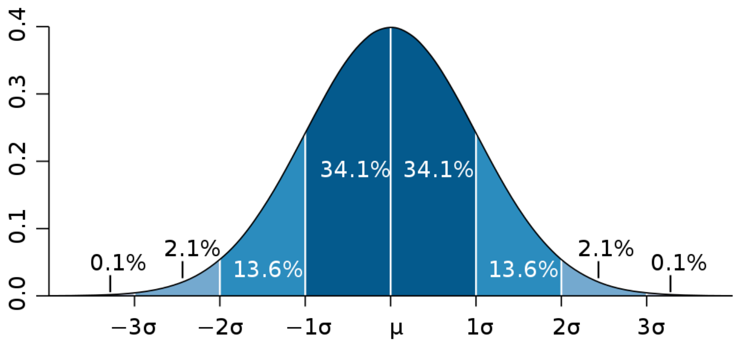
$$bel(x_t) = \eta\, p(z_t \mid x_t)\ \overline{bel}(x_t)$$

Kalman Filter

- KF is a Bayes filter
- Everything is Gaussian

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

- Optimal solution for linear models and Gaussian distributions



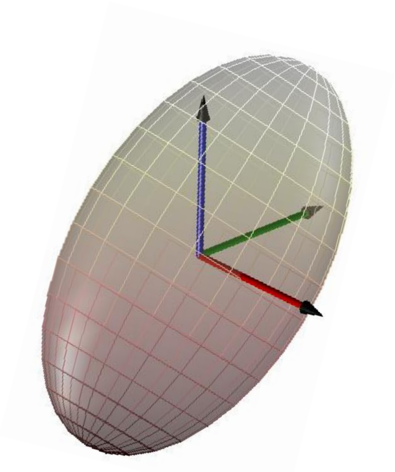


Image courtesy: K. Arras

How to Update a Gaussian Belief Based on Motions and Observations?

Gaussians: Marginalization and Conditioning

- Given

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \qquad p(x) = \mathcal{N}$$

- The marginals are Gaussians

$$p(x_a) = \mathcal{N} \qquad p(x_b) = \mathcal{N}$$

- as well as the conditionals

$$p(x_a \mid x_b) = \mathcal{N} \qquad p(x_b \mid x_a) = \mathcal{N}$$

Linear Model for
Motions and Observations

Linear Models

- Both models can be expressed through a linear function

$$f(x) = A\,x + b$$

Linear Models

- Both models can be expressed through a linear function

$$f(x) = A\,x + b$$

- A **Gaussian** that is transformed trough a linear function **stays Gaussian**

Linear Model

- The Kalman filter assumes a linear transition and observation model
- Zero mean Gaussian noise

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

$$z_t = C_t x_t + \delta_t$$

Components of a Kalman Filter

A_t Matrix $(n \times n)$ that describes how the state evolves from $t-1$ to t without controls or noise.

B_t Matrix $(n \times l)$ that describes how the control u_t changes the state from $t-1$ to t.

C_t Matrix $(k \times n)$ that describes how to map the state x_t to an observation z_t.

ϵ_t

δ_t

Random variables representing the process and measurement noise that are assumed to be independent and normally distributed with covariance R_t and Q_t respectively.

Linear Motion Model

- Motion under Gaussian noise leads to

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

Linear Motion Model

- Motion under Gaussian noise leads to

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

- R_t describes the noise of the motion

Linear Observation Model

- Measuring under Gaussian noise leads to

$$p(z_t \mid x_t) = ?$$

Linear Observation Model

- Measuring under Gaussian noise leads to

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

- Q_t describes the measurement noise

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

$$bel(x_t) = \eta \, \underbrace{p(z_t \mid x_t)}_{\text{Gaussian}} \, \underbrace{\overline{bel}(x_t)}_{?}$$

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

$$bel(x_t) = \eta \; \underbrace{p(z_t \mid x_t)}_{\text{Gaussian}} \; \underbrace{\overline{bel}(x_t)}_{?}$$

- The product of two Gaussian is again a Gaussian
- We need to show that $\overline{bel}(x_t)$ is Gaussian so that $bel(x_t)$ is Gaussian

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

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

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

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

- Is that sufficient so that $\overline{bel}(x_t)$ is Gaussian?

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

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

- Is that sufficient so that $\overline{bel}(x_t)$ is Gaussian?
- Convolution of two Gaussians stays Gaussian

Everything stays Gaussian

- We can write

$$
\begin{aligned}
&\overline{bel}(x_t) \\
&= \int p(x_t \mid u_t, x_{t-1})\, bel(x_{t-1})\, dx_{t-1} \\
&= \eta \int \exp\left(-\frac{1}{2}\,(x_t - A_t\, x_{t-1} - B_t\, u_t)^T\, R_t^{-1}\,(x_t - A_t\, x_{t-1} - B_t\, u_t)\right) \\
&\qquad\qquad \exp\left(-\frac{1}{2}\,(x_{t-1} - \mu_{t-1})^T \Sigma_{t-1}^{-1}(x_{t-1} - \mu_{t-1})\right)\, dx_{t-1}
\end{aligned}
$$

Everything stays Gaussian

- We can write

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

- and thus

$$\begin{aligned} \overline{bel}(x_t) &= \eta \int \exp\left(-L_t\right)\,dx_{t-1} \\ L_t &= \frac{1}{2}\,(x_t - A_t\,x_{t-1} - B_t\,u_t)^T\,R_t^{-1}\,(x_t - A_t\,x_{t-1} - B_t\,u_t) \\ &\quad + \frac{1}{2}\,(x_{t-1} - \mu_{t-1})^T\,\Sigma_{t-1}^{-1}\,(x_{t-1} - \mu_{t-1}) \end{aligned}$$

Everything stays Gaussian

- We can split up L_t in a part that depends on x_t and on x_t, x_{t-1}

$$L_t = L_t(x_{t-1}, x_t) + L_t(x_t)$$

- Thus

$$\begin{aligned} \overline{bel}(x_t) &= \eta \int \exp\left(-L_t(x_{t-1}, x_t) - L_t(x_t)\right)\, dx_{t-1} \\ &= \eta\ \underbrace{\exp\left(-L_t(x_t)\right)}_{\text{Gaussian}} \underbrace{\int \exp\left(-L_t(x_{t-1}, x_t)\right)\, dx_{t-1}}_{\text{Marginalization}} \end{aligned}$$

- Details: Probabilistic Robotics, Ch. 3.2 (p. 46-49)

Everything stays Gaussian

- Given an initial Gaussian belief, the belief stays Gaussian

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

$$\underbrace{bel(x_t)}_{\text{Gaussian}} = \eta \; \underbrace{p(z_t \mid x_t)}_{\text{Gaussian}} \; \underbrace{\overline{bel}(x_t)}_{\text{Gaussian}}$$

Everything is and stays Gaussian!

How Do We Typically Represent Gaussians?

$$\mu \quad \Sigma$$

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

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

$$\Rightarrow \quad \mu_t = ? \qquad \Sigma_t = ?$$

To Derive the Kalman Filter Algorithm, One Exploits…

- Product of two Gaussians is a Gaussian
- Gaussians stays Gaussians under linear transformations
- Marginal and conditional distribution of a Gaussian stays a Gaussian
- Computing mean and covariance of the marginal and conditional of a Gaussian
- Matrix inversion lemma
- …

This leads us to…

Kalman Filter Algorithm

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

$$\bar{\mu}_t = A_t \, \mu_{t-1} + B_t \, u_t$$

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

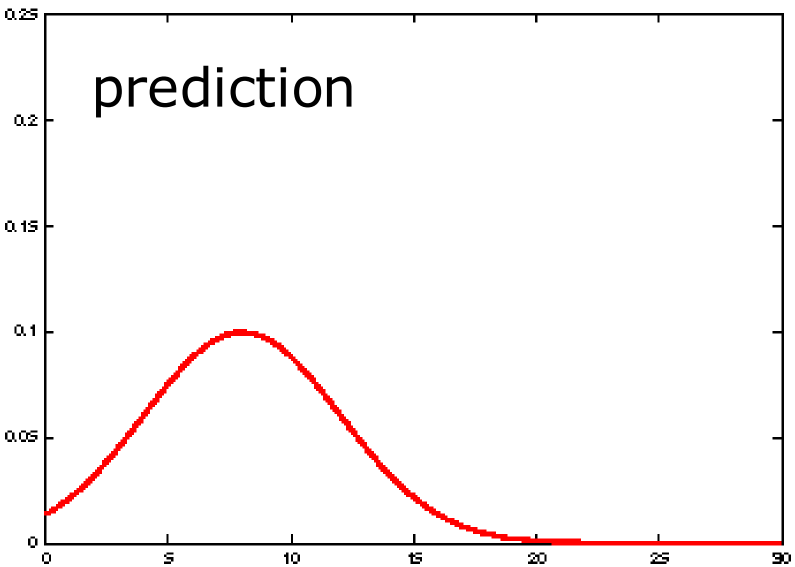
$$K_t = \bar{\Sigma}_t \, C_t^T (C_t \, \bar{\Sigma}_t \, C_t^T + Q_t)^{-1}$$

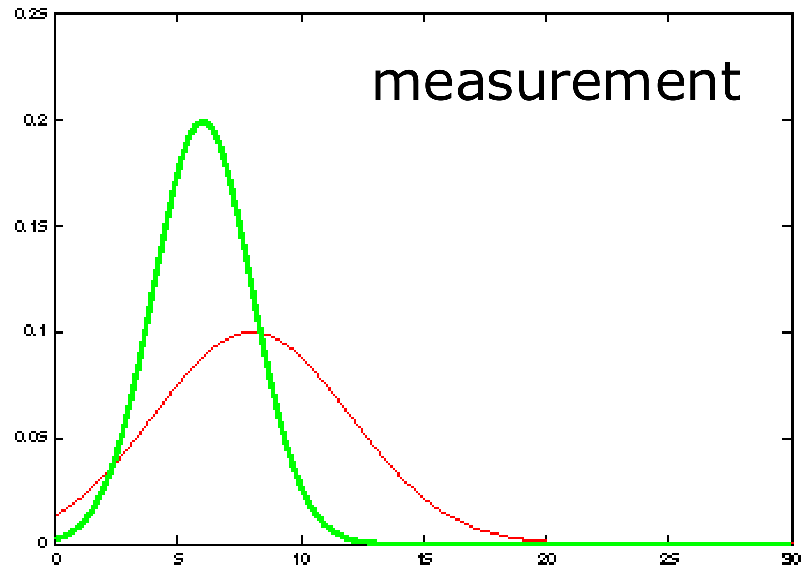
$$\mu_t = \bar{\mu}_t + K_t (z_t - C_t \, \bar{\mu}_t)$$

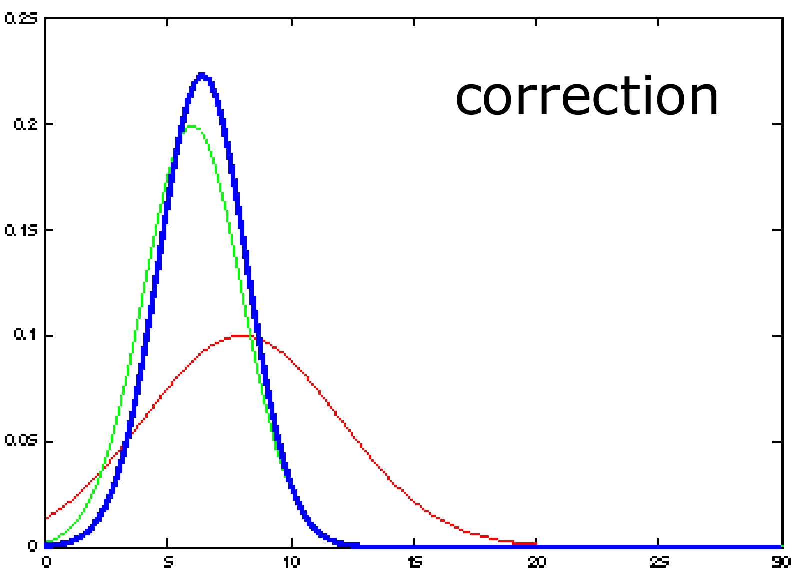
$$\Sigma_t = (I - K_t \, C_t) \, \bar{\Sigma}_t$$

$$\text{return } \mu_t, \Sigma_t$$

1D Kalman Filter Example (1)

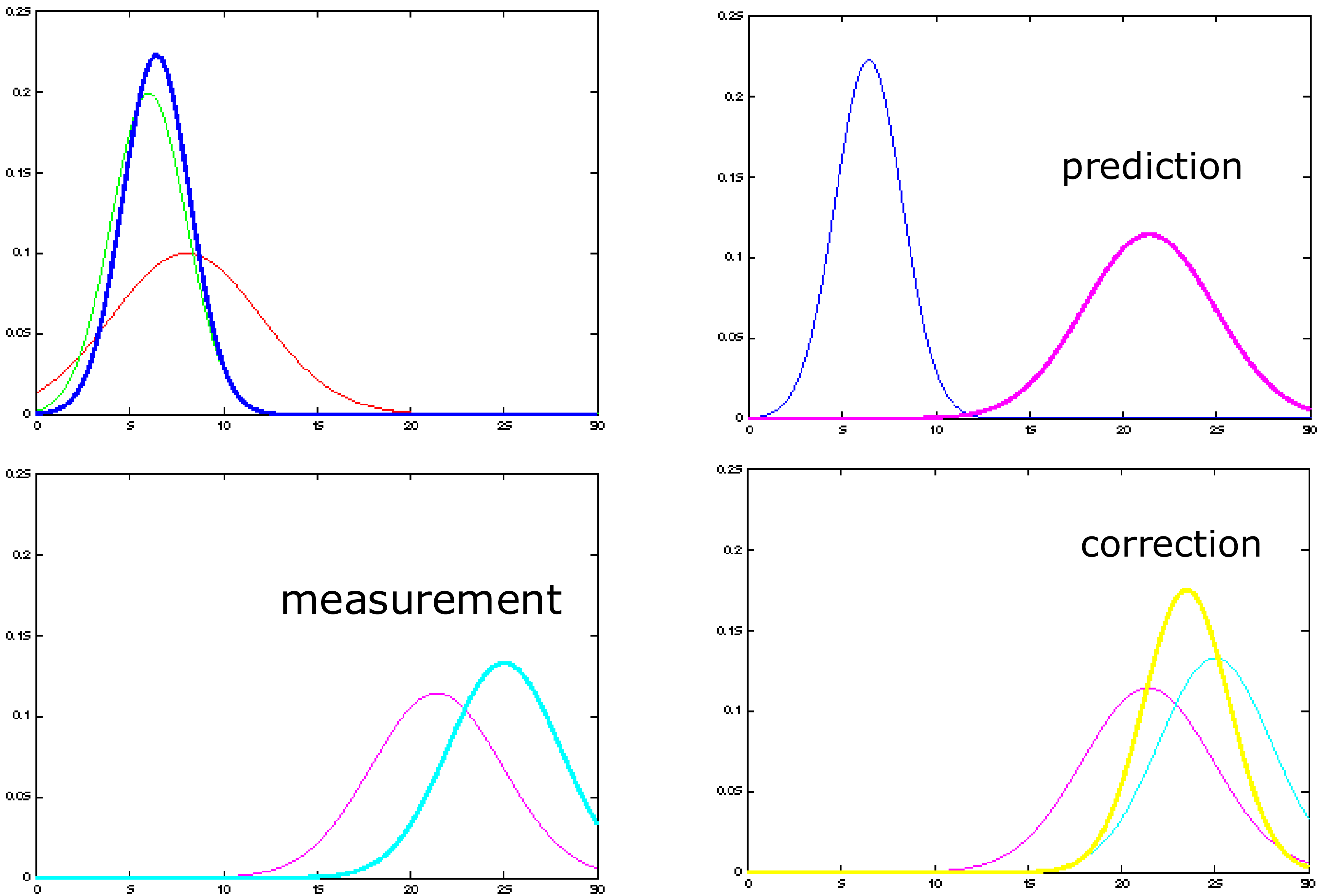






It's a weighted mean!

1D Kalman Filter Example (2)



Kalman Filter Algorithm

Kalman_filter$(\mu_{t-1}, \Sigma_{t-1}, u_t, z_t)$**:**

$$\bar{\mu}_t = A_t \, \mu_{t-1} + B_t \, u_t$$
$$\bar{\Sigma}_t = A_t \, \Sigma_{t-1} \, A_t^T + R_t$$

$$K_t = \bar{\Sigma}_t \, C_t^T (C_t \, \bar{\Sigma}_t \, C_t^T + Q_t)^{-1}$$
$$\mu_t = \bar{\mu}_t + K_t (z_t - C_t \, \bar{\mu}_t)$$
$$\Sigma_t = (I - K_t \, C_t) \, \bar{\Sigma}_t$$
$$return \; \mu_t, \Sigma_t$$

A Useless Sensor...

- Infinite noise: $Q_t = \mathrm{diag}(\infty)$
- Leads to

$$4: \quad K_t = \bar{\Sigma}_t\, C_t^T (\cancel{C_t\, \bar{\Sigma}_t\, C_t^T + Q_t})^{-1} \quad \rightarrow 0$$

$$5: \quad \mu_t = \bar{\mu}_t + K_t(z_t - C_t\, \bar{\mu}_t)$$

$$5: \quad \mu_t = \bar{\mu}_t + \underbrace{K_t}_{\rightarrow 0}(z_t - C_t\, \bar{\mu}_t)$$

$$5: \quad \mu_t = \bar{\mu}_t$$

- Observation will be ignored

A Perfect Sensor...

- No noise: $Q_t = \mathrm{diag}(0)$
- Leads to

$$4: \quad K_t = \bar{\Sigma}_t\, C_t^T (C_t\, \bar{\Sigma}_t\, C_t^T + \cancel{Q_t})^{-1} \quad \rightarrow C_t^{-1}$$

$$5: \quad \mu_t = \bar{\mu}_t + K_t (z_t - C_t\, \bar{\mu}_t)$$

$$5: \quad \mu_t = \bar{\mu}_t + K_t (z_t - C_t\, \bar{\mu}_t)$$

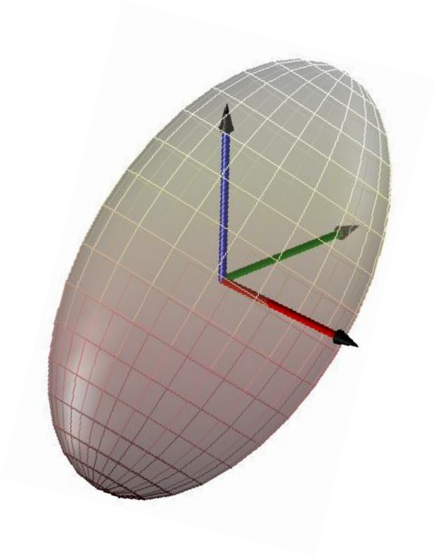
$$\rightarrow C_t^{-1} z_t - \bar{\mu}_t$$

$$5: \quad \mu_t = C_t^{-1} z_t$$

- Prediction will be ignored

Kalman Filter Assumptions

- Gaussian distributions and noise
- Linear motion and observation model



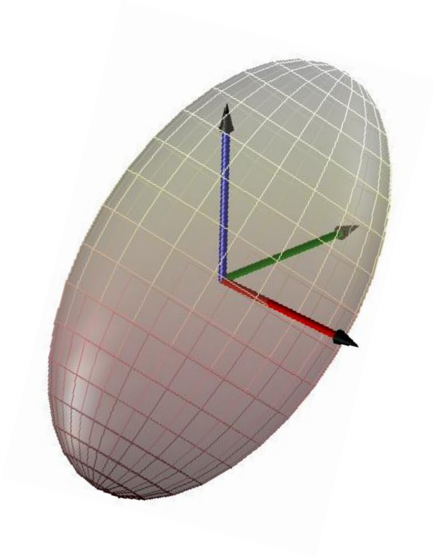
$$x_t = A_t x_{t-1} + B_t u_t + \epsilon_t$$

$$z_t = C_t x_t + \delta_t$$

Extended Kalman Filter

Kalman Filter Assumptions

- Gaussian distributions and noise
- Linear motion and observation model



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

$$z_t = C_t x_t + \delta_t$$

What if this is not the case?

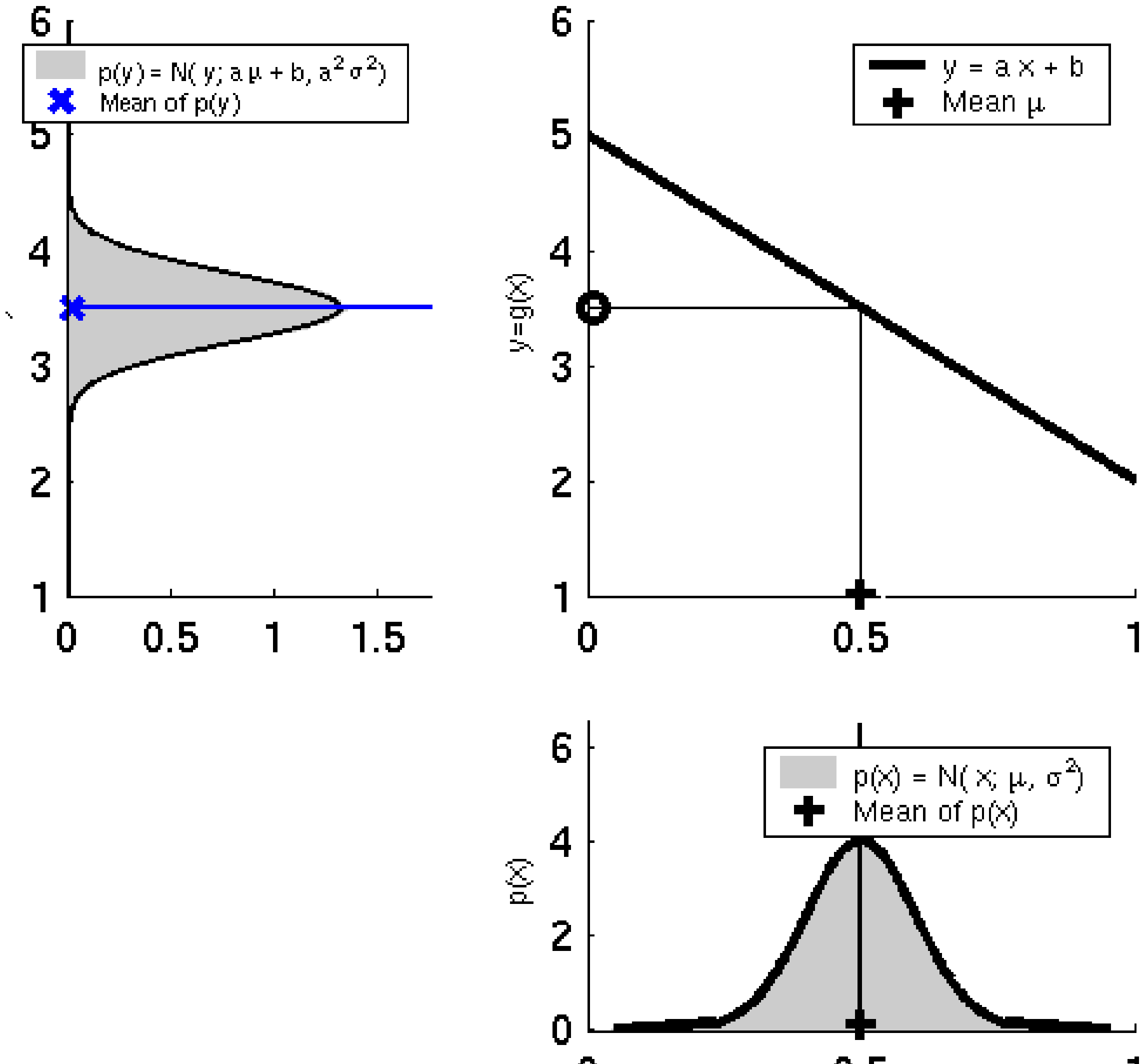
Non-Linear Dynamic Systems

- Most realistic problems involve nonlinear functions

$$\cancel{x_t = A_t x_{t-1} + B_t u_t + \epsilon_t} \qquad \cancel{z_t = C_t x_t + \delta_t}$$

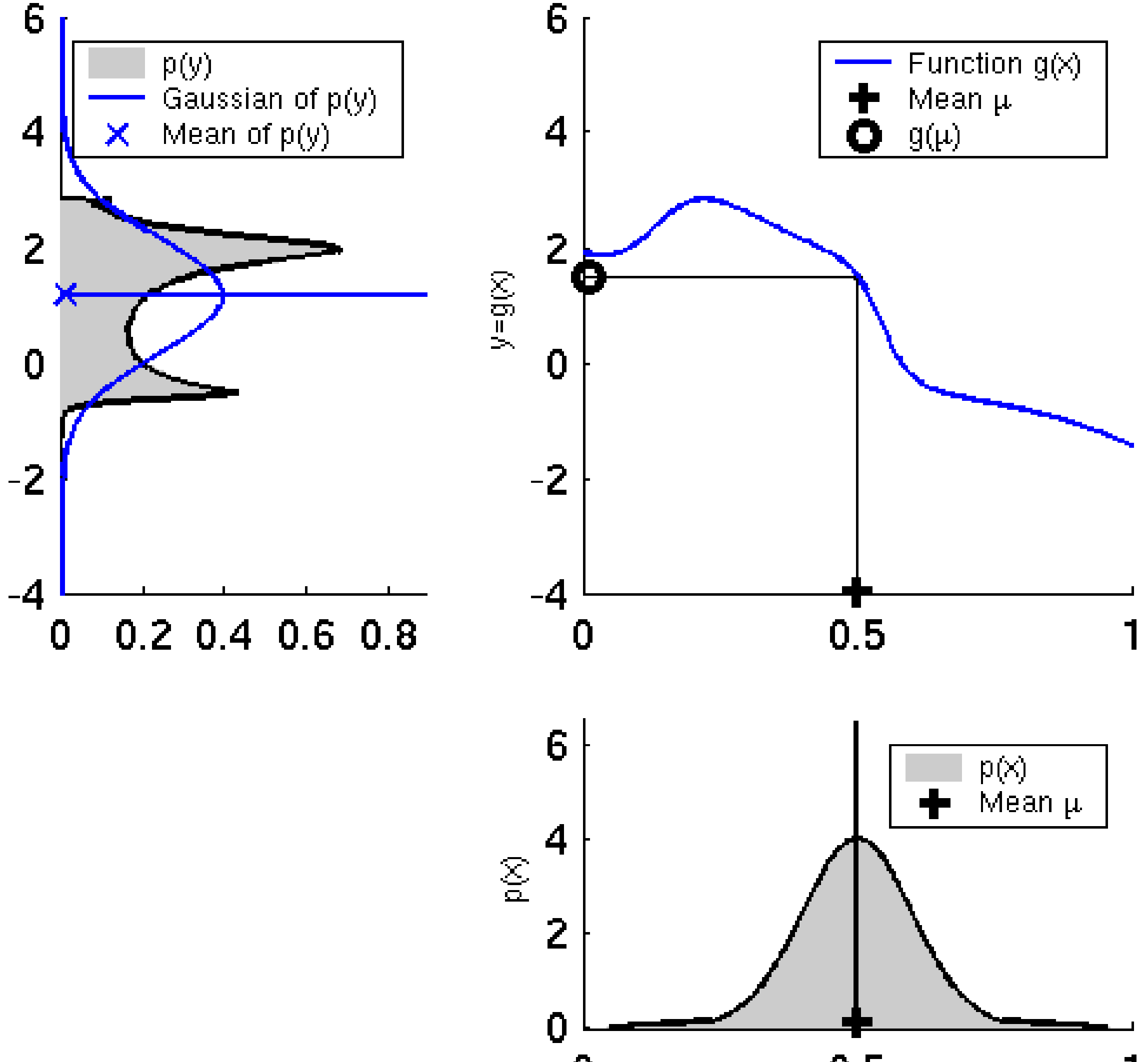
$$x_t = g(u_t, x_{t-1}) + \epsilon_t \qquad z_t = h(x_t) + \delta_t$$

Linearity Assumption Revisited



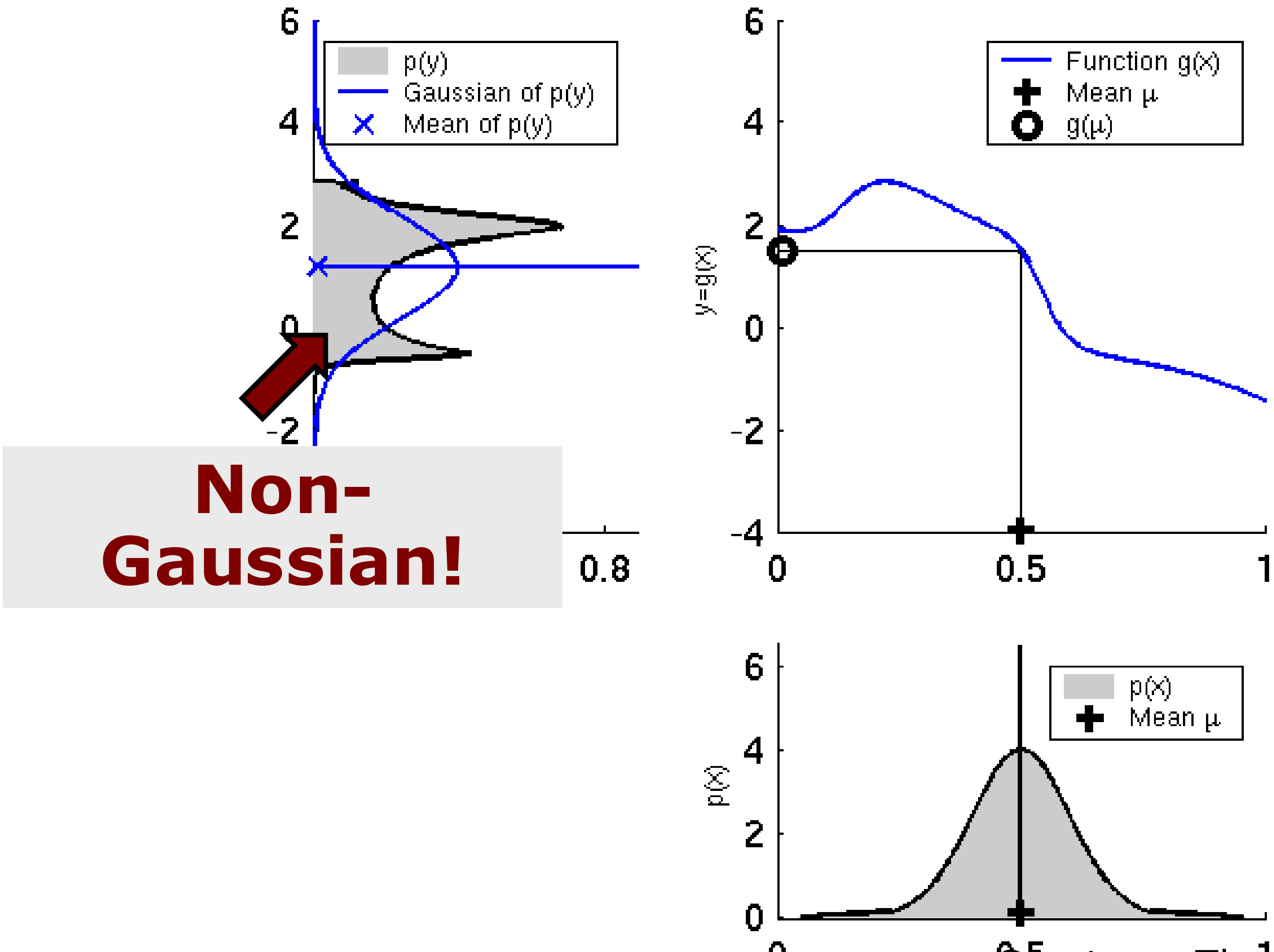
Courtesy: Thrun, Burgard, Fox

Non-Linear Function



Courtesy: Thrun, Burgard, Fox

Non-Linear Function



Courtesy: Thrun, Burgard, Fox

Non-Gaussian Distributions

- The non-linear functions lead to non-Gaussian distributions
- Kalman filter is not applicable anymore!

What can be done to resolve this?

Non-Gaussian Distributions

- The non-linear functions lead to non-Gaussian distributions
- Kalman filter is not applicable anymore!

What can be done to resolve this?

Local linearization!

EKF Linearization: First Order Taylor Expansion

- Prediction:

$$g(u_t, x) \approx g(u_t, \mu_{t-1}) + \underbrace{\frac{\partial g(u_t, x)}{\partial x}}_{=:\, G_t} (x - \mu_{t-1})$$

- Correction:

$$h(x_t) \approx h(x) + \underbrace{\frac{\partial h(x)}{\partial x_t}}_{=:\, H_t} (x - \bar{\mu}_t)$$

Jacobians

Reminder: Jacobian

- It is a non-square matrix $m \times n$ in general
- Given a vector-valued function

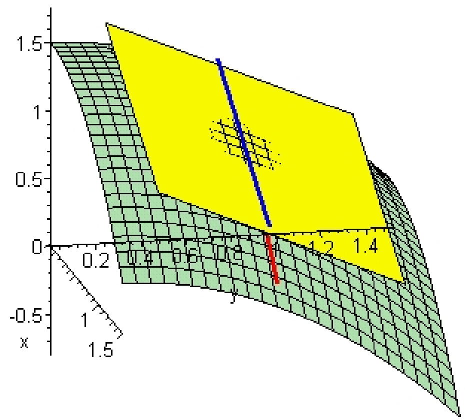
$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

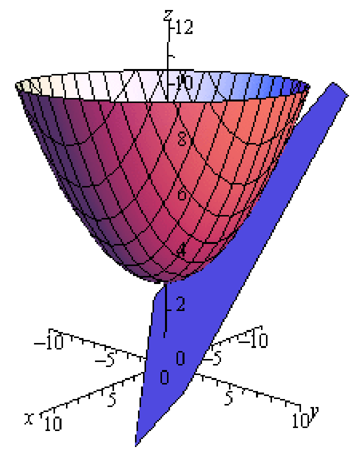
- The Jacobian is defined as

$$G_x = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix}$$

Reminder: Jacobian

- It is the orientation of the tangent plane to the vector-valued function at a given point





Courtesy: K. Arras

- Generalizes the gradient of a scalar valued function

EKF Linearization: First Order Taylor Expansion

- Prediction:

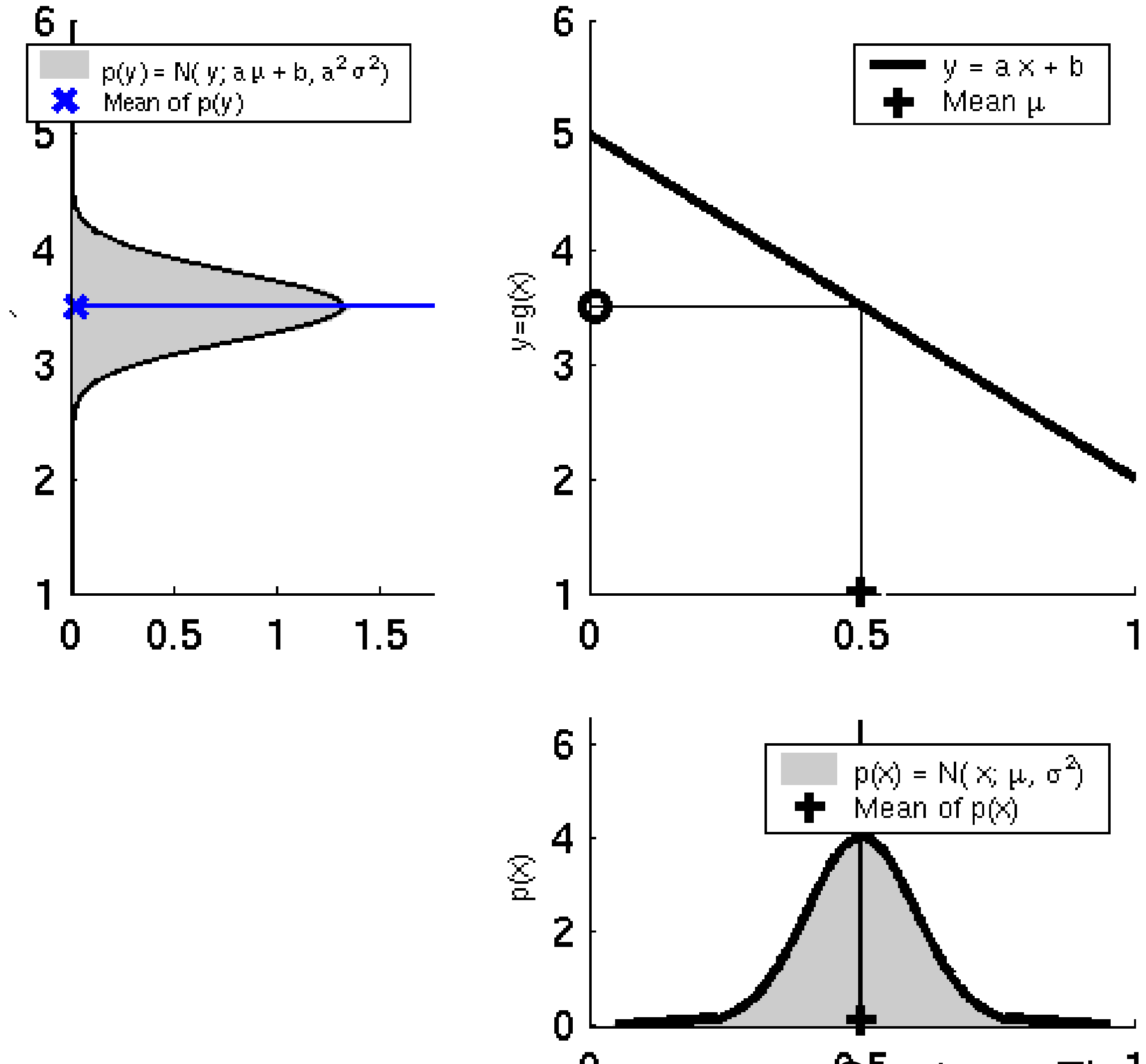
$$g(u_t, x) \approx g(u_t, \mu_{t-1}) + \underbrace{\frac{\partial g(u_t, x)}{\partial x}}_{=: G_t} (x - \mu_{t-1})$$

- Correction:

$$h(x_t) \approx h(x) + \underbrace{\frac{\partial h(x)}{\partial x_t}}_{=: H_t} (x - \bar{\mu}_t)$$

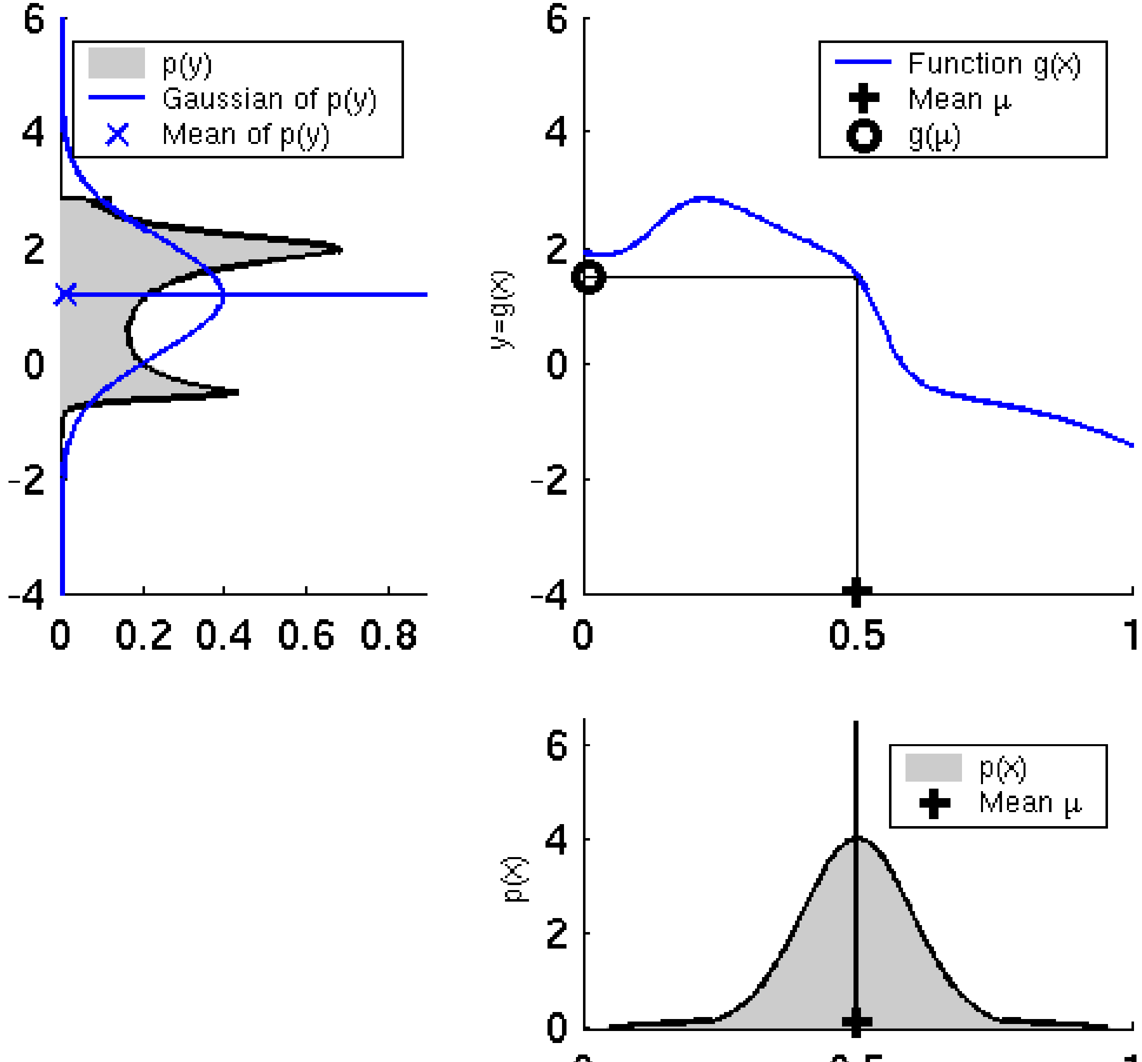
Linear functions!

Linearity Assumption Revisited



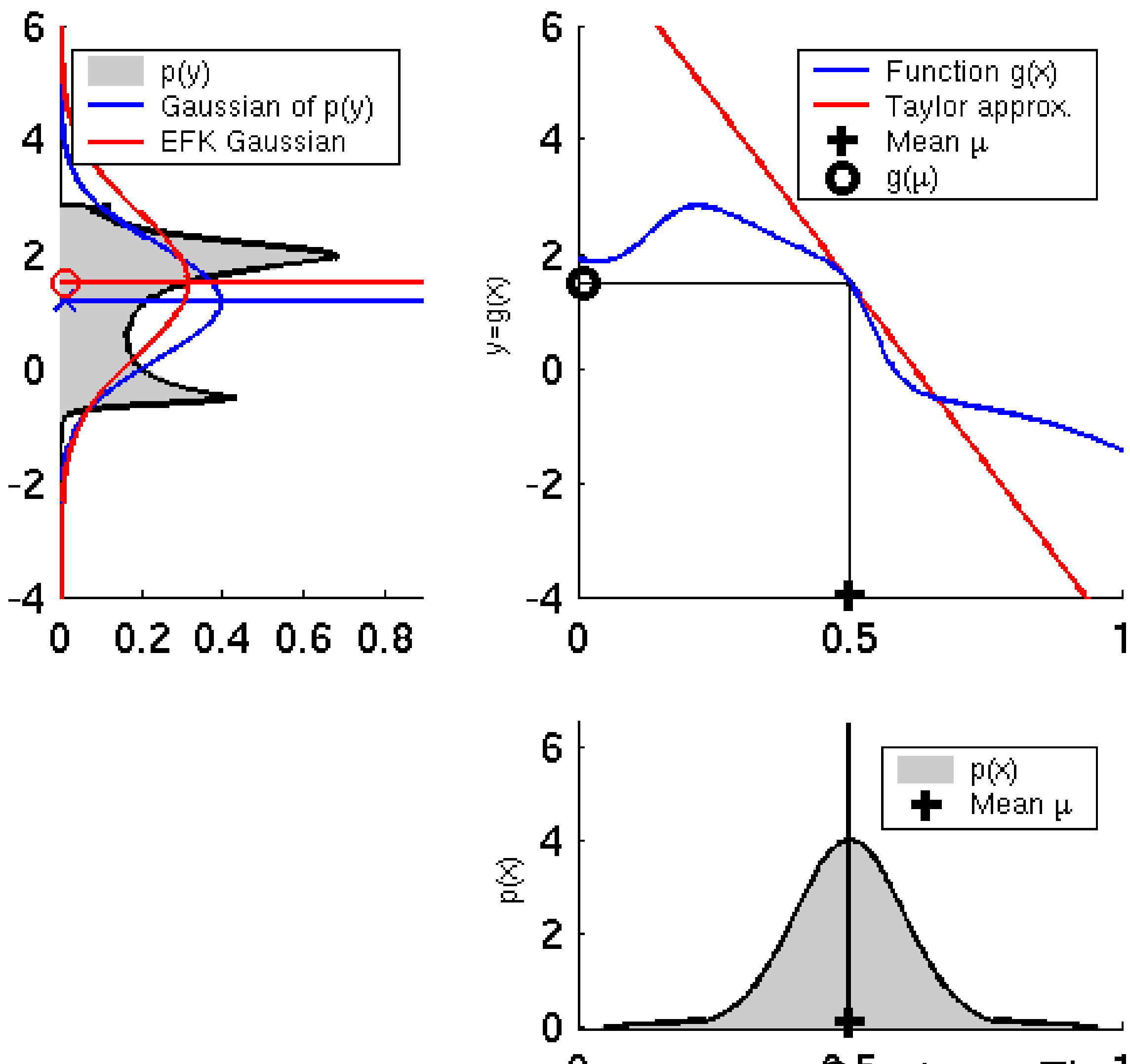
Courtesy: Thrun, Burgard, Fox

Non-Linear Function



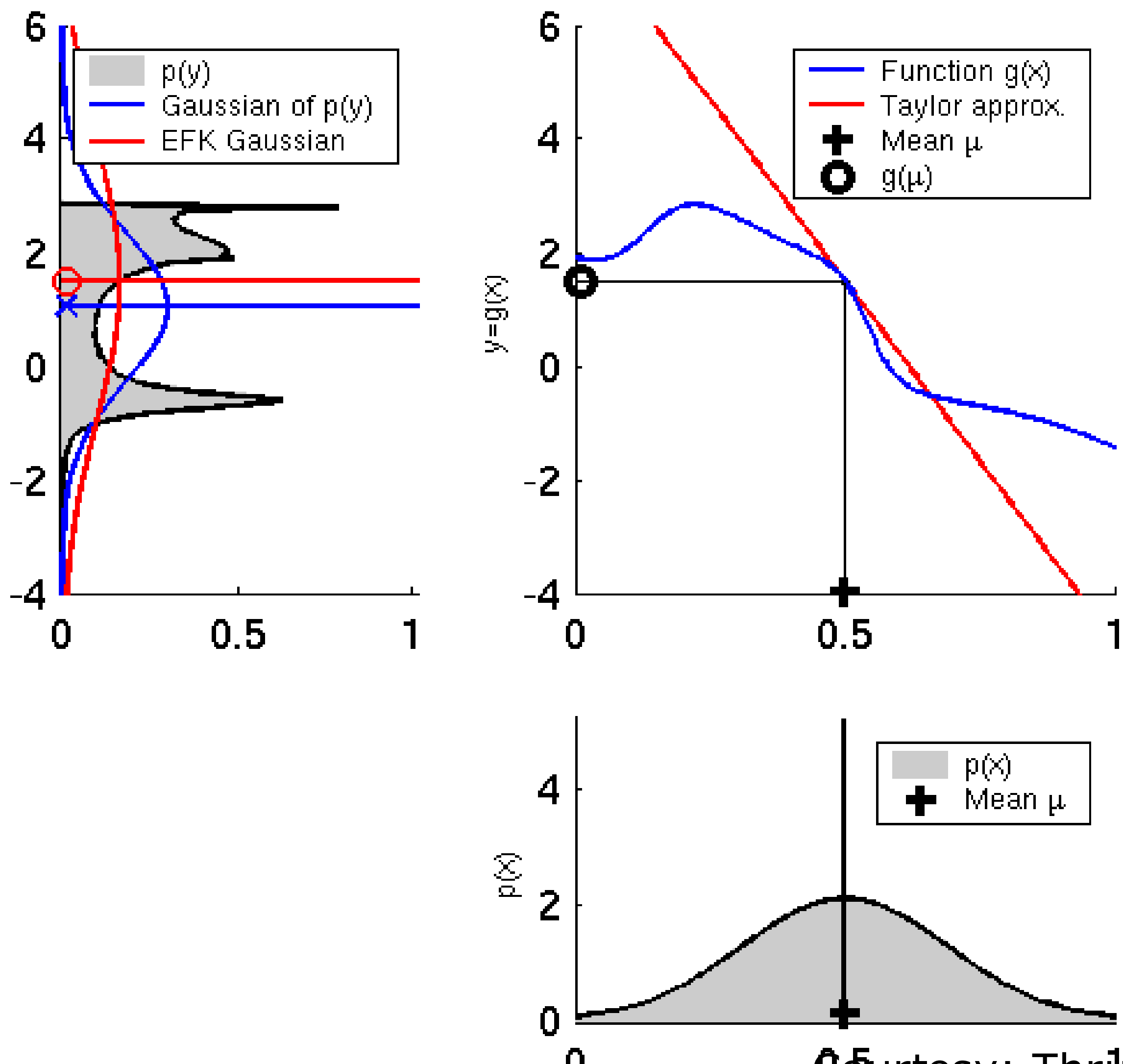
Courtesy: Thrun, Burgard, Fox

EKF Linearization (1)



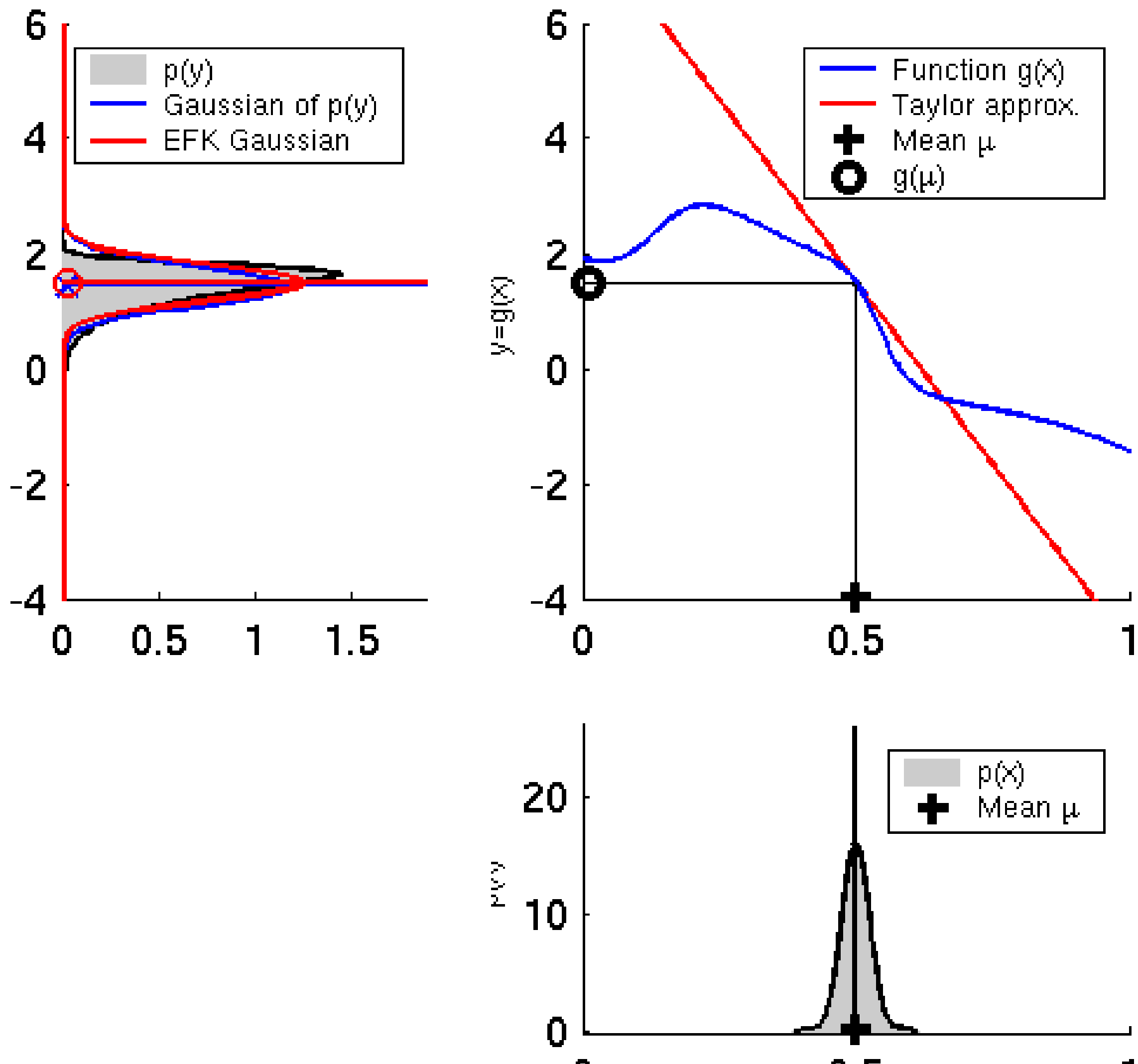
Courtesy: Thrun, Burgard, Fox

EKF Linearization (2)



Courtesy: Thrun, Burgard, Fox

EKF Linearization (3)



Courtesy: Thrun, Burgard, Fox

Linearized Motion Model

- The linearized model leads to

$$p(x_t \mid u_t, x_{t-1}) \approx \det\left(2\pi R_t\right)^{-\frac{1}{2}}$$
$$\exp\Big(-\frac{1}{2}\left(x_t - g(u_t, \mu_{t-1}) - G_t\left(x_{t-1} - \mu_{t-1}\right)\right)^T$$
$$R_t^{-1}\left(x_t - \underbrace{g(u_t, \mu_{t-1}) - G_t\left(x_{t-1} - \mu_{t-1}\right)}_{\text{linearized model}}\right)\Big)$$

- R_t describes the noise of the motion

Linearized Observation Model

- The linearized model leads to

$$p(z_t \mid x_t) = \det\left(2\pi Q_t\right)^{-\frac{1}{2}}$$
$$\exp\left(-\frac{1}{2}\left(z_t - h(\bar{\mu}_t) - H_t\left(x_t - \bar{\mu}_t\right)\right)^T \right.$$
$$\left. Q_t^{-1}\left(z_t - \underbrace{h(\bar{\mu}_t) - H_t\left(x_t - \bar{\mu}_t\right)}_{\text{linearized model}}\right)\right)$$

- Q_t describes the measurement noise

Extended Kalman Filter Algorithm

1: $\textbf{Extended\_Kalman\_filter}(\mu_{t-1}, \Sigma_{t-1}, u_t, z_t)$**:**

2: $\bar{\mu}_t = \underline{g(u_t, \mu_{t-1})}$

3: $\bar{\Sigma}_t = G_t \, \Sigma_{t-1} \, G_t^T + R_t$

4: $K_t = \bar{\Sigma}_t \, H_t^T (H_t \, \bar{\Sigma}_t \, H_t^T + Q_t)^{-1}$

5: $\mu_t = \bar{\mu}_t + K_t(z_t - \underline{h(\bar{\mu}_t)})$

6: $\Sigma_t = (I - K_t \, H_t) \, \bar{\Sigma}_t$

7: $return \ \mu_t, \Sigma_t$

$$\boxed{\begin{array}{c} A_t \leftrightarrow G_t \\ C_t \leftrightarrow H_t \end{array}}$$

KF vs. EKF

Extended Kalman Filter Algorithm

$$\textbf{Extended\_Kalman\_filter}(\mu_{t-1}, \Sigma_{t-1}, u_t, z_t)\textbf{:}$$

prediction

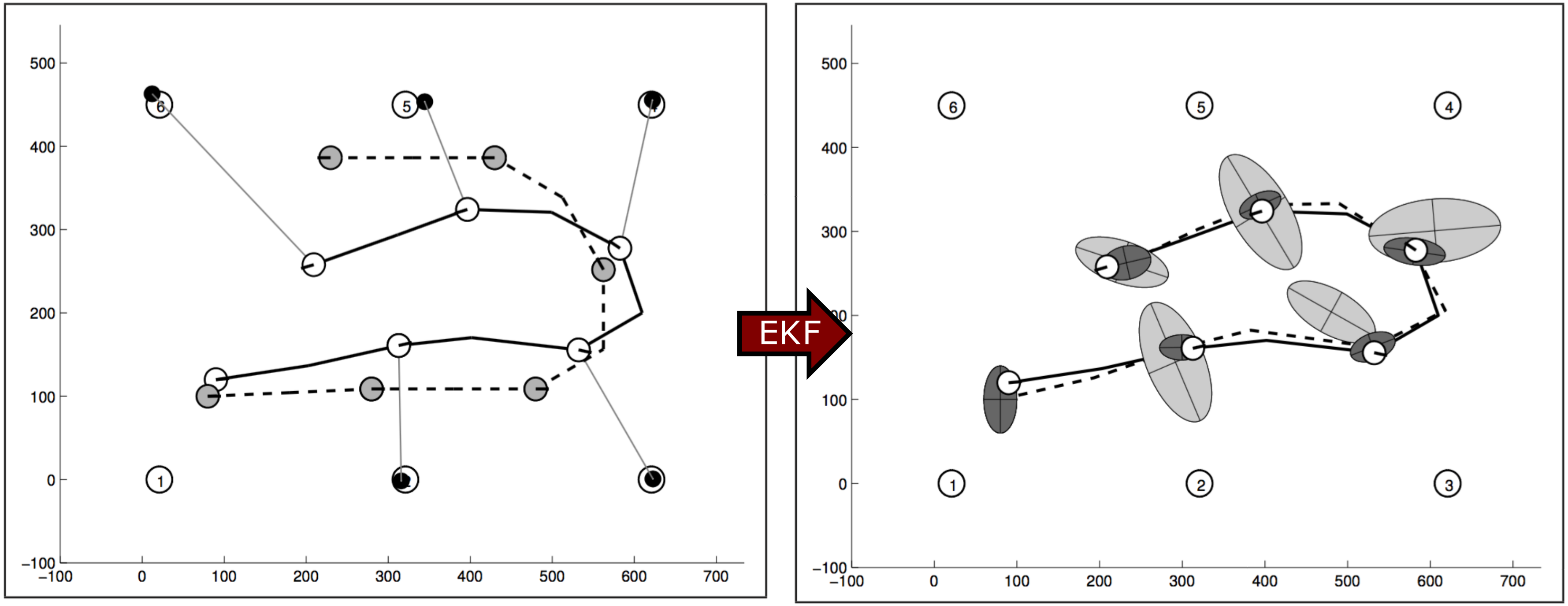
$$\bar{\mu}_t = g(u_t, \mu_{t-1})$$
$$\bar{\Sigma}_t = G_t\, \Sigma_{t-1}\, G_t^T + R_t$$

correction

$$K_t = \bar{\Sigma}_t\, H_t^T (H_t\, \bar{\Sigma}_t\, H_t^T + Q_t)^{-1}$$
$$\mu_t = \bar{\mu}_t + K_t(z_t - h(\bar{\mu}_t))$$
$$\Sigma_t = (I - K_t\, H_t)\, \bar{\Sigma}_t$$

$$return\ \mu_t, \Sigma_t$$

EKF Localization Example



weighted sum of predictions and observations

Image courtesy: Thrun, Burgard, Fox

Kalman Filter Summary

- Kalman filter is a Bayes filter for the linear Gaussian case
- All models/beliefs must be Gaussian
- Linear motion and observation model
- Optimal estimator
- Linear models are a serious limitation in real-world applications

Extended Kalman Filter Summary

- Extension of the Kalman filter
- One way to handle the non-linearities
- Performs local linearizations (Taylor)
- Works well in practice for moderate non-linearities
- Large uncertainty leads to increased approximation error

Literature

Kalman Filter and EKF

- Thrun et al.: “Probabilistic Robotics”, Chapter 3
- Schön and Lindsten: “Manipulating the Multivariate Gaussian Density”
- Thrun et al.: “Probabilistic Robotics”, Chapter 7