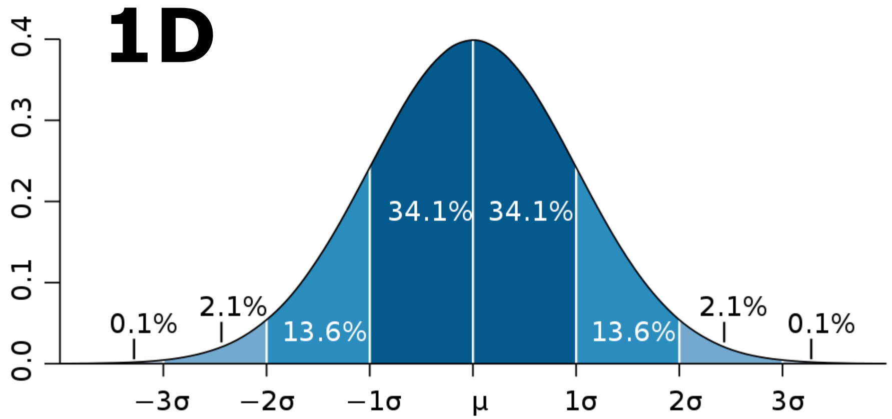
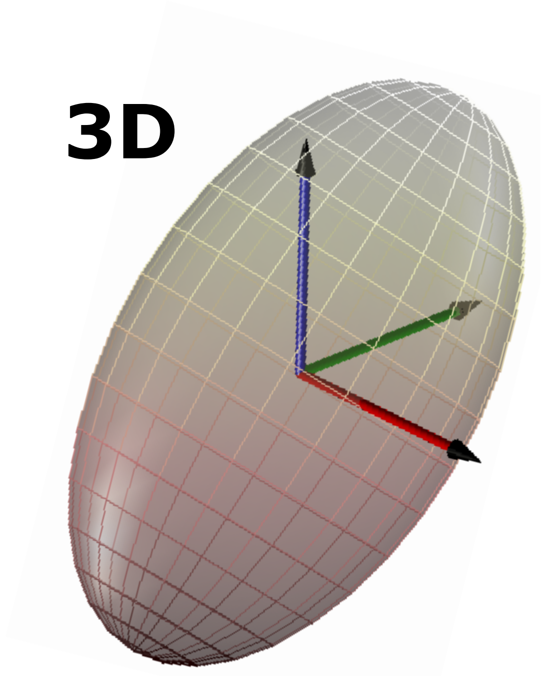
# Information Filter

Dr. Gaurav Pandey

# Gaussian Distribution

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




Slide Credit: Cyrill Stachniss

# Canonical Parametrization

- Alternate parametrization of Gaussian Distribution
- Parametrized by **Information Matrix** and **Information Vector**

$$\Omega = \Sigma^{-1} \qquad\qquad \xi = \Sigma^{-1}\mu$$

Information Matrix

Information Vector

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

Moments Parametrization

# Moments vs Canonical Parametrization

$$
\begin{aligned}
p(x) & \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big) \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu - \frac{1}{2}\mu^T\Sigma^{-1}\mu\Big) \\
&= \boxed{\det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\mu^T\Sigma^{-1}\mu\Big)} \\
&\qquad\qquad \exp\Big(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu\Big) \\
&= \eta \exp\Big(-\frac{1}{2}x^T\underline{\Sigma^{-1}}x + x^T\underline{\Sigma^{-1}\mu}\Big)
\end{aligned}
$$

# Dual Parametrization

- Moments Parametrization

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

- Canonical Parametrization

$$p(x) = \frac{\exp(-\frac{1}{2}\mu^T\xi)}{\det(2\pi\Omega^{-1})^{\frac{1}{2}}} \exp\Big( -\frac{1}{2}x^T\Omega x + x^T\xi\Big)$$

# Information Filter

- Information Filter is Kalman filter with canonical parametrization i.e. beliefs are parametrized by **information matrix** and **information vector** instead of covariance matrix and mean vector.

$$\Omega = \Sigma^{-1}$$
$$\xi = \Sigma^{-1} \mu$$

Information Form

$$\Sigma = \Omega^{-1}$$
$$\mu = \Omega^{-1} \xi$$

Moments Form

# Kalman Filter vs Information 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 $\mu_t, \Sigma_t$

**Algorithm Information_filter**($\xi_{t-1}, \Omega_{t-1}, u_t, z_t$)**:**

$\bar{\Omega}_t = (A_t\ \Omega_{t-1}^{-1}\ A_t^T + R_t)^{-1}$
$\bar{\xi}_t = \bar{\Omega}_t (A_t\ \Omega_{t-1}^{-1}\ \xi_{t-1} + B_t\ u_t)$
$\Omega_t = C_t^T\ Q_t^{-1}\ C_t + \bar{\Omega}_t$
$\xi_t = C_t^T\ Q_t^{-1}\ z_t + \bar{\xi}_t$
return $\xi_t, \Omega_t$

Note: No need to find Kalman Gain in Information Form !!

# Information Filter Derivation

- Lets consider a **linear Gaussian system**

$$
\begin{aligned}
x_t &= A_t x_{t-1} + B_t u_t + \varepsilon_t \\
z_t &= C_t x_t + \delta_t
\end{aligned}
$$

- Information filter tracks the information matrix and the information vector and can be estimated in two steps (similar to Kalman Filter):
  - Prediction
  - Correction

# Prediction

- Predicted Belief (Moments form):

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

- Substitute:

$$\begin{aligned}\mu_{t-1} &= \Omega_{t-1}^{-1}\ \xi_{t-1} \\ \Sigma_{t-1} &= \Omega_{t-1}^{-1}\end{aligned}$$

- Predicted Belief (Information Form):

$$\begin{aligned}\bar{\Omega}_t &= (A_t\ \Omega_{t-1}^{-1}\ A_t^T + R_t)^{-1} \\ \bar{\xi}_t &= \bar{\Omega}_t(A_t\ \Omega_{t-1}^{-1}\ \xi_{t-1} + B_t\ u_t)\end{aligned}$$

# Correction

- Recall Bayes Filter (Measurement Update)

$$
\begin{aligned}
bel(x_t) &= \eta \quad \underbrace{p(z_t \mid x_t)}_{\sim\mathcal{N}(z_t;C_t x_t,Q_t)} \quad \underbrace{\overline{bel}(x_t)}_{\sim\mathcal{N}(x_t;\bar{\mu}_t,\bar{\Sigma}_t)} \\
&= \eta \ \exp\left(-\frac{1}{2}\ (z_t - C_t x_t)^T\ Q_t^{-1}\ (z_t - C_t x_t)\right)\ \exp\left(-\frac{1}{2}\ (x_t - \bar{\mu}_t)^T\ \bar{\Sigma}_t^{-1}\ (x_t - \bar{\mu}_t)\right) \\
&= \eta \ \exp\left(-\frac{1}{2}\ (z_t - C_t x_t)^T\ Q_t^{-1}\ (z_t - C_t x_t) - \frac{1}{2}\ (x_t - \bar{\mu}_t)^T\ \bar{\Sigma}_t^{-1}\ (x_t - \bar{\mu}_t)\right) \\
&= \eta \ \exp\left(-\frac{1}{2}\ x_t^T\ C_t^T\ Q_t^{-1}\ C_t\ x_t + x_t^T\ C_t^T\ Q_t^{-1}\ z_t - \frac{1}{2}\ x_t^T\ \bar{\Omega}_t x_t + x_t^T \bar{\xi}_t\right) \\
&= \eta \ \exp\left(-\frac{1}{2}\ x_t^T\ \underbrace{[C_t^T\ Q_t^{-1}\ C_t + \bar{\Omega}_t]}_{\Omega_t}\ x_t + x_t^T\ \underbrace{[C_t^T\ Q_t^{-1}\ z_t + \bar{\xi}_t]}_{\xi_t}\right) \Rightarrow
\end{aligned}
$$

$$
\begin{aligned}
\Omega_t &= C_t^T\ Q_t^{-1}\ C_t + \bar{\Omega}_t \\
\xi_t &= C_t^T\ Q_t^{-1}\ z_t + \bar{\xi}_t
\end{aligned}
$$

# Information Filter

**Algorithm Information_filter($\xi_{t-1}, \Omega_{t-1}, u_t, z_t$):**

$$\bar{\Omega}_t = (A_t \; \Omega_{t-1}^{-1} \; A_t^T + R_t)^{-1}$$

$$\bar{\xi}_t = \bar{\Omega}_t (A_t \; \Omega_{t-1}^{-1} \; \xi_{t-1} + B_t \; u_t)$$

$$\Omega_t = C_t^T \; Q_t^{-1} \; C_t + \bar{\Omega}_t$$

$$\xi_t = C_t^T \; Q_t^{-1} \; z_t + \bar{\xi}_t$$

return $\xi_t, \Omega_t$

# Kalman Filter vs Information 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 $\mu_t, \Sigma_t$

**Algorithm Information_filter($\xi_{t-1}, \Omega_{t-1}, u_t, z_t$):**

$\bar{\Omega}_t = (A_t\ \Omega_{t-1}^{-1}\ A_t^T + R_t)^{-1}$
$\bar{\xi}_t = \bar{\Omega}_t(A_t\ \Omega_{t-1}^{-1}\ \xi_{t-1} + B_t\ u_t)$
$\Omega_t = C_t^T\ Q_t^{-1}\ C_t + \bar{\Omega}_t$
$\xi_t = C_t^T\ Q_t^{-1}\ z_t + \bar{\xi}_t$
return $\xi_t, \Omega_t$

- Prediction is computationally easy in Kalman filter
- Correction is computationally easy in Information filter
- Information matrix is <u>naturally sparse </u>(We will see the consequences of this when we talk about SLAM)

# Extended Information Filter

- Very similar to EKF: Linearization via Taylor series

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

$$\bar{\mu}_t = g(u_t, \mu_{t-1})$$
$$\bar{\Sigma}_t = G_t\ \Sigma_{t-1}\ G_t^T + R_t$$
$$K_t = \bar{\Sigma}_t\ H_t^T (H_t\ \bar{\Sigma}_t\ H_t^T + Q_t)^{-1}$$
$$\mu_t = \bar{\mu}_t + K_t(z_t - h(\bar{\mu}_t))$$
$$\Sigma_t = (I - K_t\ H_t)\ \bar{\Sigma}_t$$

return $\mu_t, \Sigma_t$

**Algorithm Extended_information_filter**($\xi_{t-1}, \Omega_{t-1}, u_t, z_t$):

$$\mu_{t-1} = \Omega_{t-1}^{-1}\ \xi_{t-1}$$
$$\bar{\Omega}_t = (G_t\ \Omega_{t-1}^{-1}\ G_t^T + R_t)^{-1}$$
$$\bar{\xi}_t = \bar{\Omega}_t\ g(u_t, \mu_{t-1})$$
$$\bar{\mu}_t = g(u_t, \mu_{t-1})$$
$$\Omega_t = \bar{\Omega}_t + H_t^T\ Q_t^{-1}\ H_t$$
$$\xi_t = \bar{\xi}_t + H_t^T\ Q_t^{-1}\ [z_t - h(\bar{\mu}_t) - H_t\ \bar{\mu}_t]$$

return $\xi_t, \Omega_t$

# Prediction

- EKF

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

- EIF

$$
\begin{aligned}
\bar{\Omega}_t &= (G_t \, \Omega_{t-1}^{-1} \, G_t^T + R_t)^{-1} \\
\bar{\xi}_t &= \bar{\Omega}_t \, g(u_t, \Omega_{t-1}^{-1} \, \xi_{t-1})
\end{aligned}
$$

# Correction

$$
\begin{aligned}
bel(x_t) \quad = \quad & \eta \ \exp\left(-\frac{1}{2}\ (z_t - h(\bar{\mu}_t) - H_t\ (x_t - \bar{\mu}_t))^T\ Q_t^{-1}\right. \\
& \left.(z_t - h(\bar{\mu}_t) - H_t\ (x_t - \bar{\mu}_t)) - \frac{1}{2}(x_t - \bar{\mu}_t)^T \bar{\Sigma}_t^{-1} (x_t - \bar{\mu}_t)\right)
\end{aligned}
$$

$$
\begin{aligned}
\Omega_t \quad &= \quad \bar{\Omega}_t + H_t^T\ Q_t^{-1}\ H_t \\
\xi_t \quad &= \quad \bar{\xi}_t + H_t^T\ Q_t^{-1}\ (z_t - h(\bar{\mu}_t) + H_t\ \bar{\mu}_t)
\end{aligned}
$$

# EIF vs EKF

- EIF is EKF in canonical form
- Prediction and correction step complexities differ in the two formulations
- Theoretically both are same
- EIF is reportedly more stable numerically
- Information matrix is naturally sparse this makes EIF more suitable for some SLAM problems especially in multi-agent SLAM where measurement updates are obtained from other robots as well.
- Easy to assign an initial error:

$$\Sigma = \infty \quad \Rightarrow \quad \Omega = 0$$

# Revision

- A robot uses a range sensor that can measure ranges from 0m to 3m. For simplicity, assume that the actual ranges are uniformly distributed in this interval. Unfortunately, the sensor can be faulty. When the sensor is faulty it constantly outputs a range below 1m, regardless of the actual range in the sensor's measurement cone. We know that the prior probability of a sensor to be faulty is $p = 0.01$. Suppose the robot queried its sensor N times and every single time the measurement value is below 1m. What is the posterior probability that the sensor is faulty, for $N = 1, 5$.