Probabilistic Robotics Course

Kalman Filters

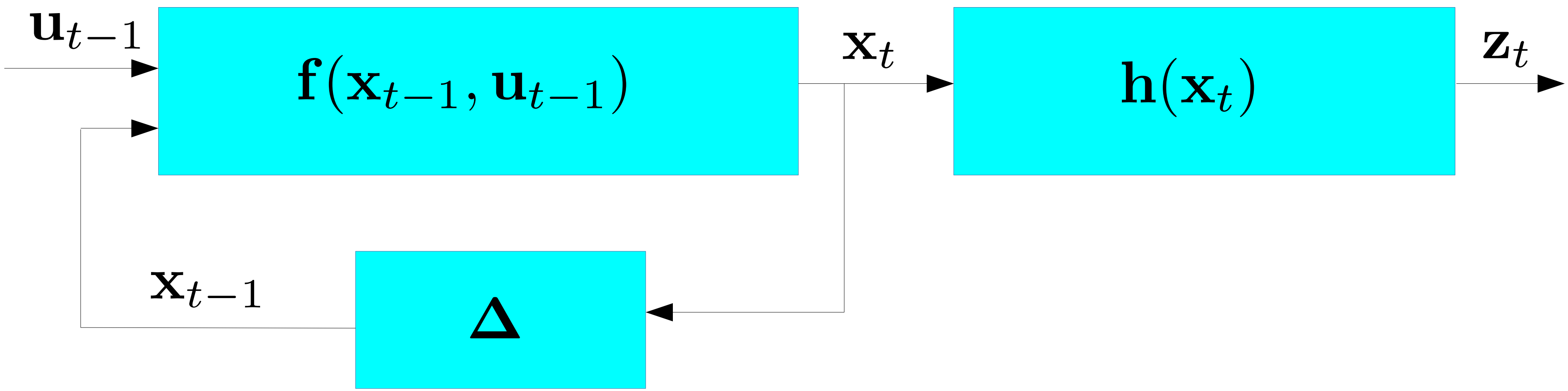
Giorgio Grisetti
grisetti@dis.uniroma1.it

Dept of Computer Control and Management Engineering
Sapienza University of Rome

Outline

- (Stochastic) System Models
- Linear Systems and Gaussian Noise
- Kalman Filter
- Non-Linear Systems and Gaussian Noise
- Extended Kalman Filter

System Model



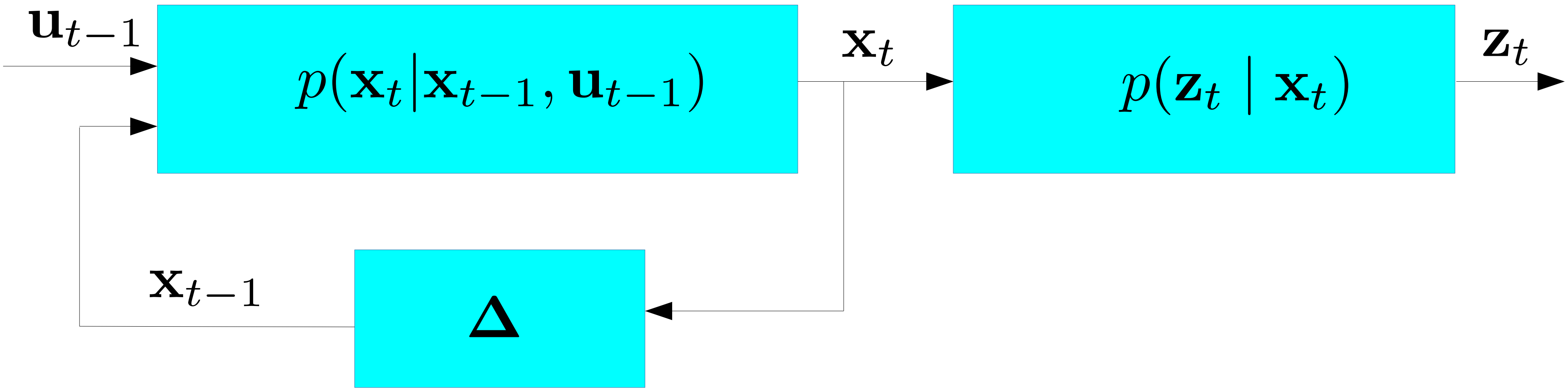
Notation:

- **u**: controls
- **x**: state
- **z**: measurement
- **f**: transition function
- **h**: observation function

Perfect knowledge of

- inputs
- measurements
- transition model
- observation model

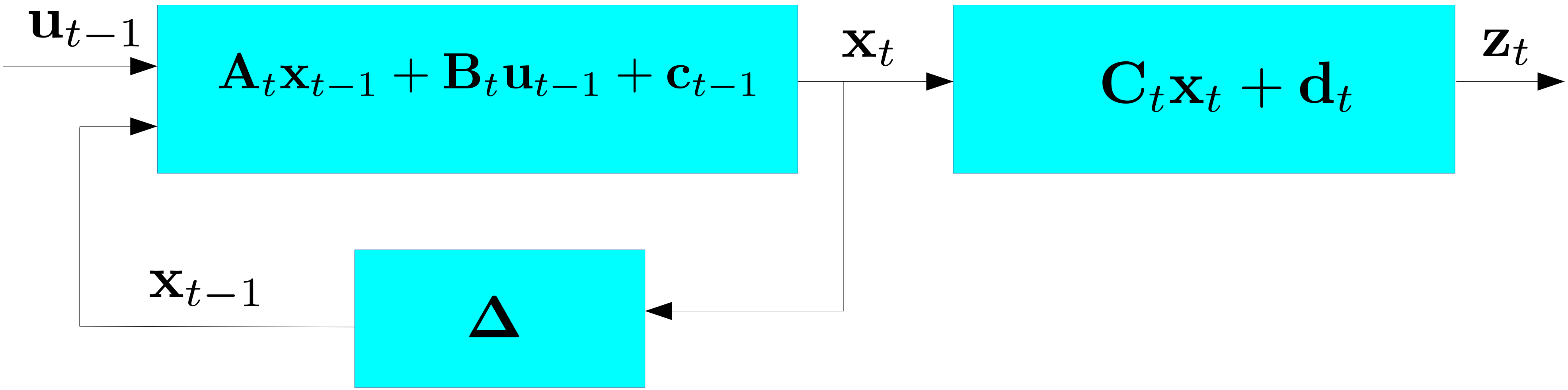
Stochastic System Model



Transition and observation functions are replaced by conditional distributions

- Transition Model $p(\mathbf{x}_t|\mathbf{x}_{t-1},\mathbf{u}_{t-1})$
- Observation Model $p(\mathbf{z}_t \mid \mathbf{x}_t)$

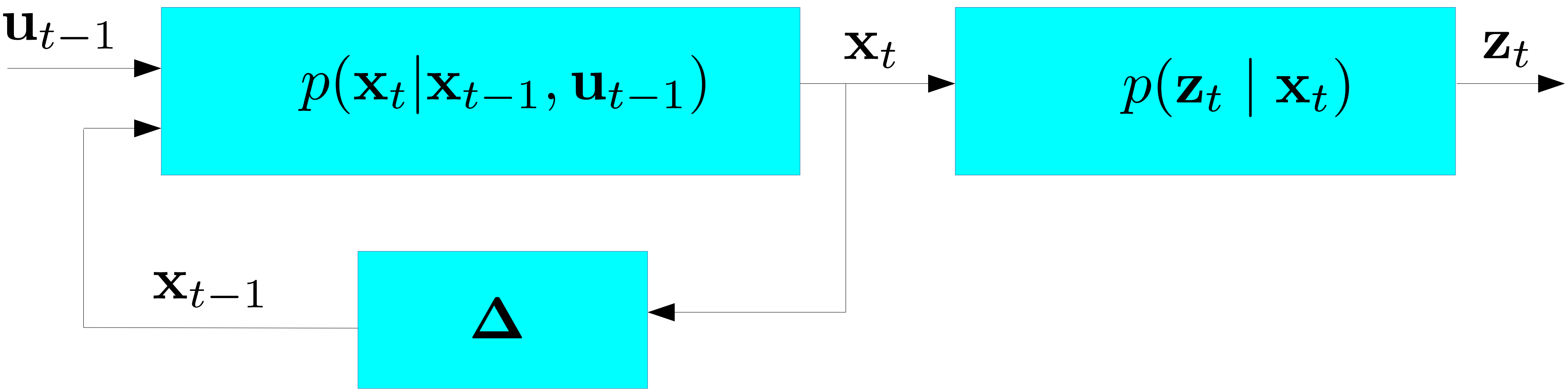
Linear System



Transition and observation functions are affine transforms

- **A**: state transition matrix
- **B**: input matrix
- **C**: observation matrix

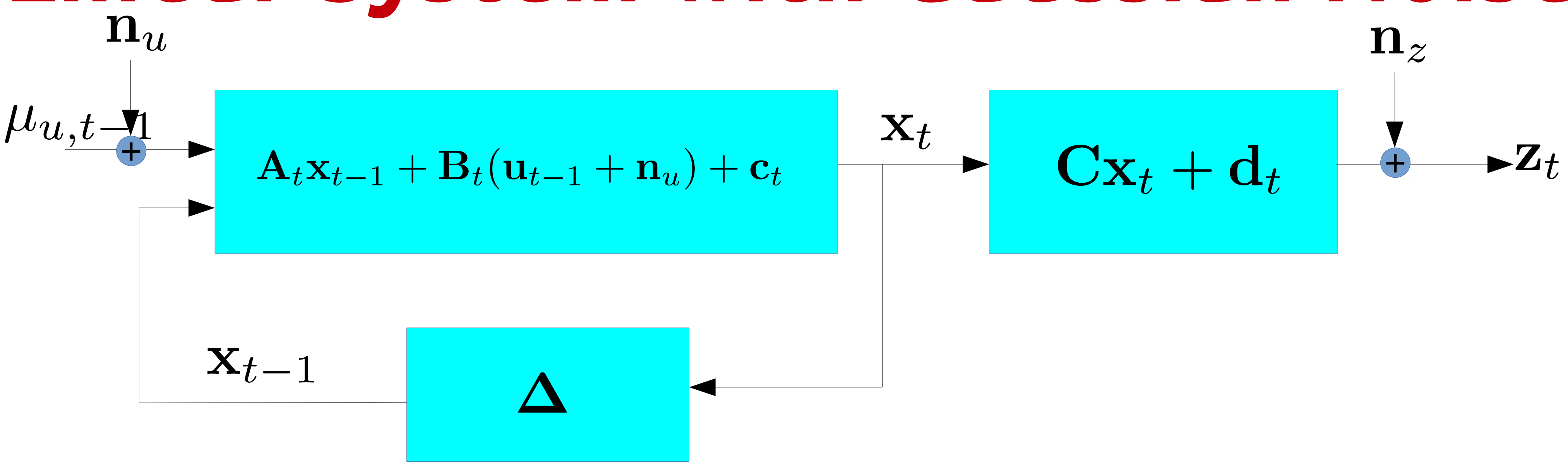
Linear System with Gaussian Noise



Transition and observation functions are replaced by conditional distributions

- $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$
- $p(\mathbf{z}_t \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{z}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{z}_t; \mathbf{C}\mathbf{x}_t + \mathbf{d}_t, \boldsymbol{\Sigma}_z)$

Linear System with Gaussian Noise



Inputs and Observations are affected by zero mean Gaussian noise. Initial belief is Gaussian

$$\mathbf{n}_u \sim \mathcal{N}(\mathbf{n}_u | 0, \boldsymbol{\Sigma}_u) \qquad p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0 | \mu_0, \boldsymbol{\Sigma}_0)$$

$$\mathbf{n}_z \sim \mathcal{N}(\mathbf{n}_z | 0, \boldsymbol{\Sigma}_z) \qquad p(\mathbf{z}_t \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{z}_t | \mathbf{C}\mathbf{x}_t + \mathbf{d}_t, \boldsymbol{\Sigma}_z)$$

For compactness we embed the noise in the control

$$\mathbf{u}_{t-1} \sim \mathcal{N}(\mathbf{u}_{t-1} | \mu_{u,t-1}, \boldsymbol{\Sigma}_{u,t})$$

Filtering

Some considerations

- The initial belief is Gaussian
- The noise is Gaussian
- All functions are affine

Gaussian distributions are closed under

- Affine transformation
- Chain rule
- Marginalization
- Conditioning

The belief remains Gaussian

Predict

Incorporate the control by computing the Gaussian distribution of the next state given the input

The next state is an affine transform of the past state and the controls`

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{A}_t & \mathbf{B}_t \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{u}_{t-1} \end{pmatrix} + \mathbf{c}_t$$

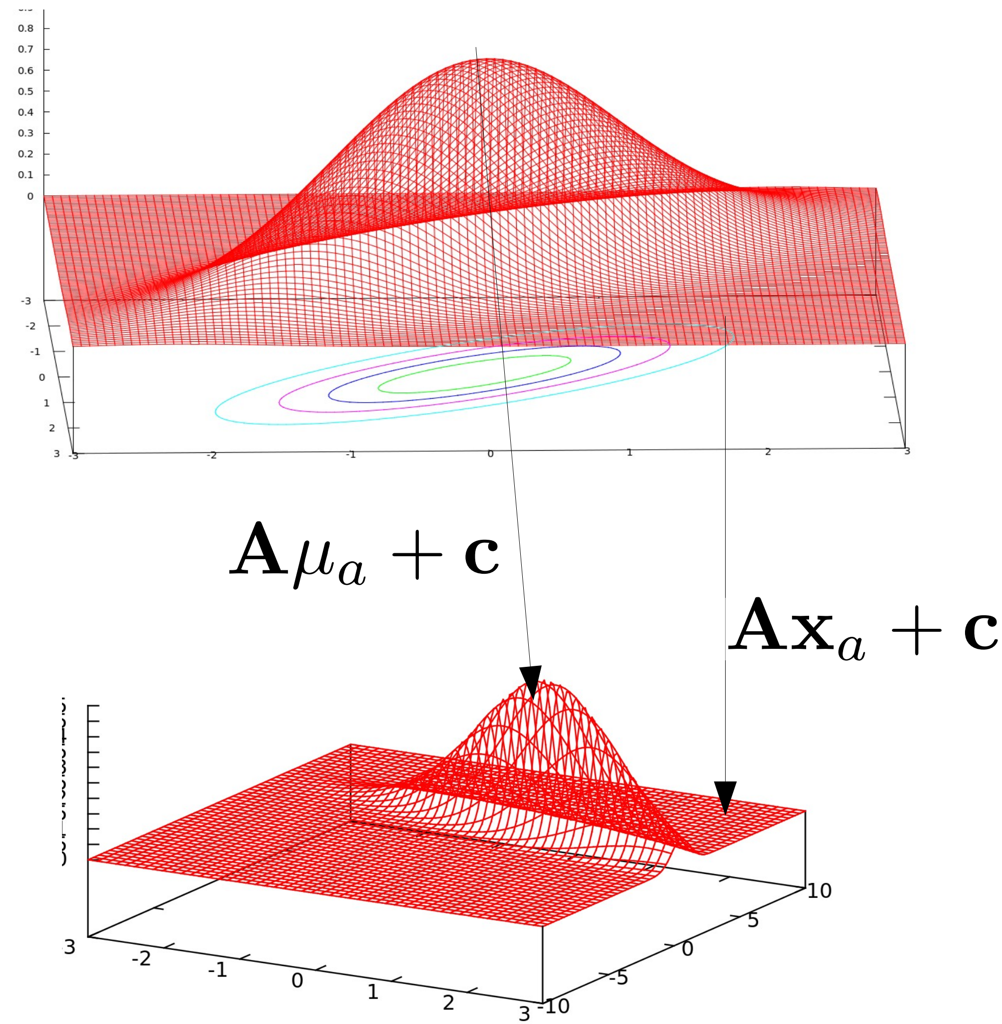
The previous states and controls are distributed according to

$$\begin{pmatrix} \mathbf{x}_{t-1|t-1} \\ \mathbf{u}_{t-1} \end{pmatrix} \sim \mathcal{N}\left[\left(\mu_{t-1|t-1} \mu_{u,t-1} \right), \begin{pmatrix} \mathbf{\Sigma}_{t-1|t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{u,t-1} \end{pmatrix} \right]$$

Affine Transformation

Let $\mathbf{x}_a$ be a Gaussian random variable such that

$$\mathbf{x}_a \sim \mathcal{N}(\mathbf{x}_a, \mu_a, \mathbf{\Sigma}_a)$$



Let $\mathbf{x}_b = \mathbf{f}(\mathbf{x}_a) = \mathbf{A}\mathbf{x}_a + \mathbf{c}$

an affine transformation of $\mathbf{x}_a$

$\mathbf{x}_b$ is Gaussian:

$$p(\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_b; \mu_b, \mathbf{\Sigma}_b)$$

The parameters are

$$\mu_b = \mathbf{A}\mu + \mathbf{c} \qquad \mathbf{\Sigma}_b = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$$

Predict

$$
\begin{aligned}
p(\mathbf{x}_{t|t-1}) &= \mathcal{N}(\mathbf{x}_{t|t-1}; \mu_{t|t-1}, \mathbf{\Sigma}_{t|t-1}) \\
\mu_{t|t-1} &= \mathbf{A}_t \mu_{t-1|t-1} + \mathbf{B}_t \mu_{u,t-1} + \mathbf{c}_t \\
\mathbf{\Sigma}_{t|t-1} &= \mathbf{A}_t \mathbf{\Sigma}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{B}_t \mathbf{\Sigma}_u \mathbf{B}_t^T
\end{aligned}
$$

Apply the rule on the affine transformation

No need to marginalize, since this step is included in the transformation

Update: Chain Rule

Given the following distributions

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{z}_t; \mathbf{C}_t\mathbf{x}_t + \mathbf{d}_t, \boldsymbol{\Sigma}_z) \\ p(\mathbf{x}_t) &= p(\mathbf{x}_{t|t-1}) \\ &= \mathcal{N}\left(\mathbf{x}_{t|t-1}; \mu_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right) \end{aligned}$$

We want to compute the joint p(x,z)

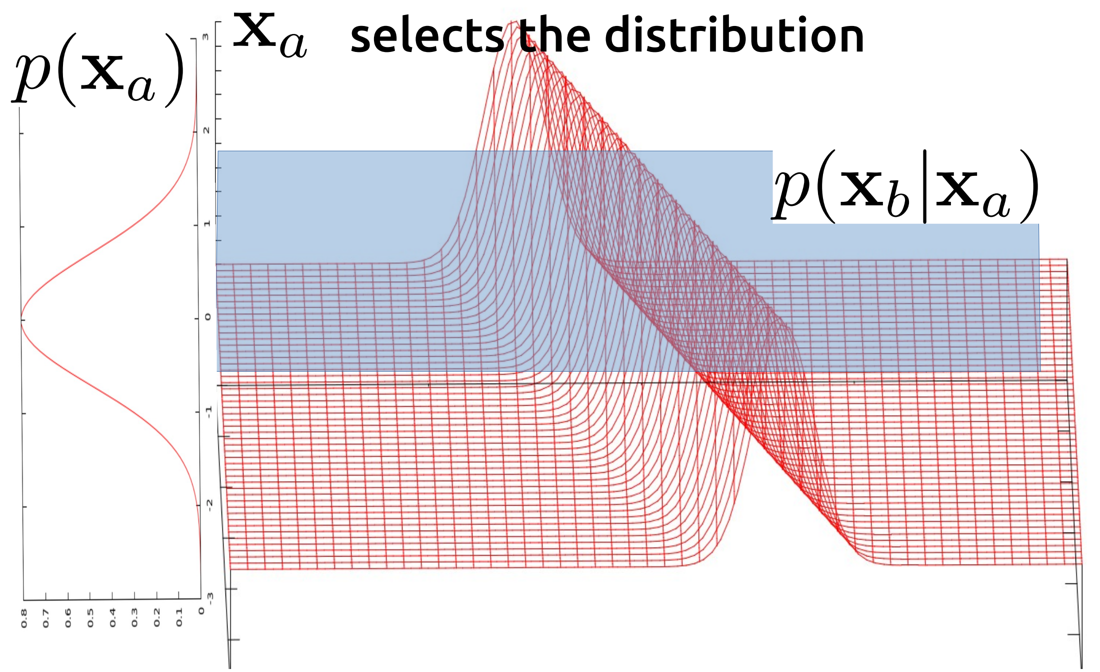
$$p(\mathbf{z}_t, \mathbf{x}_t) \quad =$$

Chain Rule

We know

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_a, \mathbf{\Sigma}_a).$$

$$p(\mathbf{x}_b|\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_b; \mathbf{A}\mathbf{x}_a + \mathbf{c}, \mathbf{\Sigma}_{b|a})$$



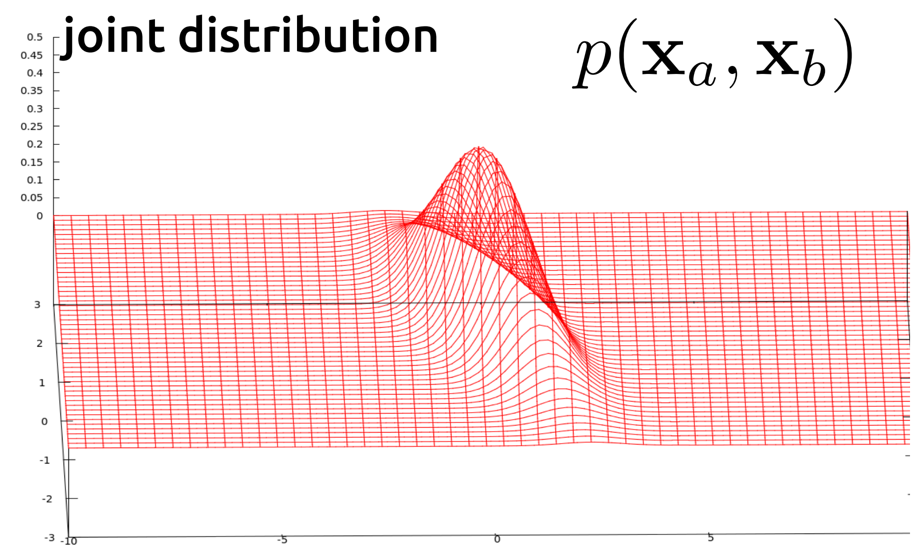
We want to compute

$$p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\mathbf{x}_{a,b}; \mu_{a,b}, \mathbf{\Sigma}_{a,b}\right)$$

The parameters are

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mu_a \\ \mathbf{A}\mu_a + \mathbf{c} \end{pmatrix}$$

$$\mathbf{\Sigma}_{a,b} = \begin{pmatrix} \mathbf{\Sigma}_a & \mathbf{\Sigma}_a \mathbf{A}^T \\ \mathbf{A}\mathbf{\Sigma}_a & \mathbf{\Sigma}_{b|a} + \mathbf{A}\mathbf{\Sigma}_a\mathbf{A}^T \end{pmatrix}$$



Update: Chain Rule

Given the following distributions

$$
\begin{aligned}
p(\mathbf{z}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{z}_t; \mathbf{C}_t\mathbf{x}_t + \mathbf{d}_t, \boldsymbol{\Sigma}_z) \\
p(\mathbf{x}_t) &= p(\mathbf{x}_{t|t-1}) \\
&= \mathcal{N}\left(\mathbf{x}_{t|t-1}; \mu_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)
\end{aligned}
$$

We want to compute the joint p(x,z)

$$
\begin{aligned}
p(\mathbf{z}_t, \mathbf{x}_t) &= \mathcal{N}\left(\begin{pmatrix} \mu_{t|t-1} \\ \mu_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T \\ \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_z + \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T \end{pmatrix}\right) \\
\mu_z &= \mathbf{C}_t\mu_{t|t-1} + \mathbf{d}_t
\end{aligned}
$$

Update: Conditioning

Conditioning

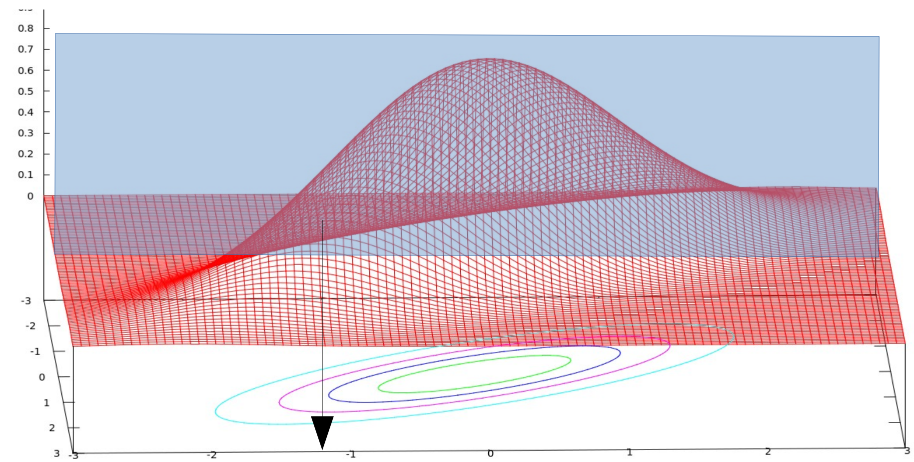
$$\begin{aligned} p(\mathbf{z}_t, \mathbf{x}_t) &= \mathcal{N}\left(\left(\begin{array}{c} \mu_{t|t-1} \\ \mu_z \end{array} \right), \left(\begin{array}{cc} \mathbf{\Sigma}_{t|t-1} & \mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T \\ \mathbf{C}_t\mathbf{\Sigma}_{t|t-1} & \mathbf{\Sigma}_z + \mathbf{C}_t\mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T \end{array} \right) \right) \\ \mu_z &= \mathbf{C}_t\mu_{t|t-1} + \mathbf{d}_t \end{aligned}$$

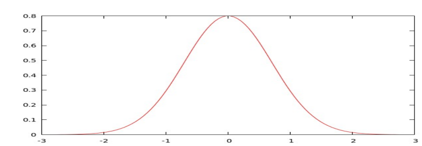
$$p(\mathbf{x}_{t|t}|\mathbf{z}_t) = \mathcal{N}(\mathbf{x}_{t|t}; \mu_{t|t}, \mathbf{\Sigma}_{t|t})$$

Conditioning

Let $\mathbf{x}^T = (\mathbf{x}_a^T \ \mathbf{x}_b^T)$ be a Gaussian random variable such that

$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mu, \mathbf{\Sigma})$$





The conditional

$$p(\mathbf{x}_a \mid \mathbf{x}_b) = \frac{p(\mathbf{x}_a, \mathbf{x}_b)}{\int_{\mathbf{x}_a} p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_a}$$

is Gaussian with parameters

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_{a|b}, \mathbf{\Sigma}_{a|b})$$

$$\mu_{a|b} = \mu_a + \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mu_b)$$

$$\mathbf{\Sigma}_{a|b} = \mathbf{\Sigma}_{aa} - \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}\mathbf{\Sigma}_{ba}$$

Update: Conditioning

Given the joint, we want to condition on the current measurement

$$
\begin{aligned}
p(\mathbf{z}_t, \mathbf{x}_t) &= \mathcal{N}\left(\begin{pmatrix} \mu_{t|t-1} \\ \mu_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T \\ \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1} & \boldsymbol{\Sigma}_z + \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T \end{pmatrix}\right) \\
\mu_z &= \mathbf{C}_t \mu_{t|t-1} + \mathbf{d}_t
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{x}_{t|t}|\mathbf{z}_t) &= \mathcal{N}(\mathbf{x}_{t|t}; \mu_{t|t}, \boldsymbol{\Sigma}_{t|t}) \\
\mu_{t|t} &= \mu_{t|t-1} + \underbrace{\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T\left(\boldsymbol{\Sigma}_z + \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T\right)^{-1}}_{\mathbf{K}_t}(\mathbf{z}_t - \mu_z) \\
\boldsymbol{\Sigma}_{t|t} &= \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T\left(\boldsymbol{\Sigma}_z + \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^T\right)^{-1}\mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}
\end{aligned}
$$

Wrapup (Kalman Filter)

Predict: incorporate new control

$$
\begin{aligned}
\mu_{t|t-1} &= \mathbf{A}_t \mu_{t-1|t-1} + \mathbf{B}_t \mathbf{u}_{t-1} + \mathbf{c}_{t-1} \\
\mathbf{\Sigma}_{t|t-1} &= \mathbf{A}_t \mathbf{\Sigma}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{B}_t \mathbf{\Sigma}_u \mathbf{B}_t^T
\end{aligned}
$$

Update: incorporate new measurement

$$
\begin{aligned}
\mu_z &= \mathbf{C}_t \mu_{t|t-1} + \mathbf{d}_t \\
\mathbf{K}_t &= \mathbf{\Sigma}_{t|t-1} \mathbf{C}_t^T \left(\mathbf{\Sigma}_z + \mathbf{C}_t \mathbf{\Sigma}_{t|t-1} \mathbf{C}_t^T \right)^{-1} \\
\mu_{t|t} &= \mu_{t|t-1} + \mathbf{K}_t \left(\mathbf{z}_t - \mu_z \right) \\
\mathbf{\Sigma}_{t|t} &= \left(\mathbf{I} - \mathbf{K}_t \mathbf{C}_t \right) \mathbf{\Sigma}_{t|t-1}
\end{aligned}
$$

Non Linear Systems

If the system is non-linear, we can still use a variant of the Kalman Filter by dynamically linearizing the system at each time

This is known as the Extended Kalman Filter (EKF)

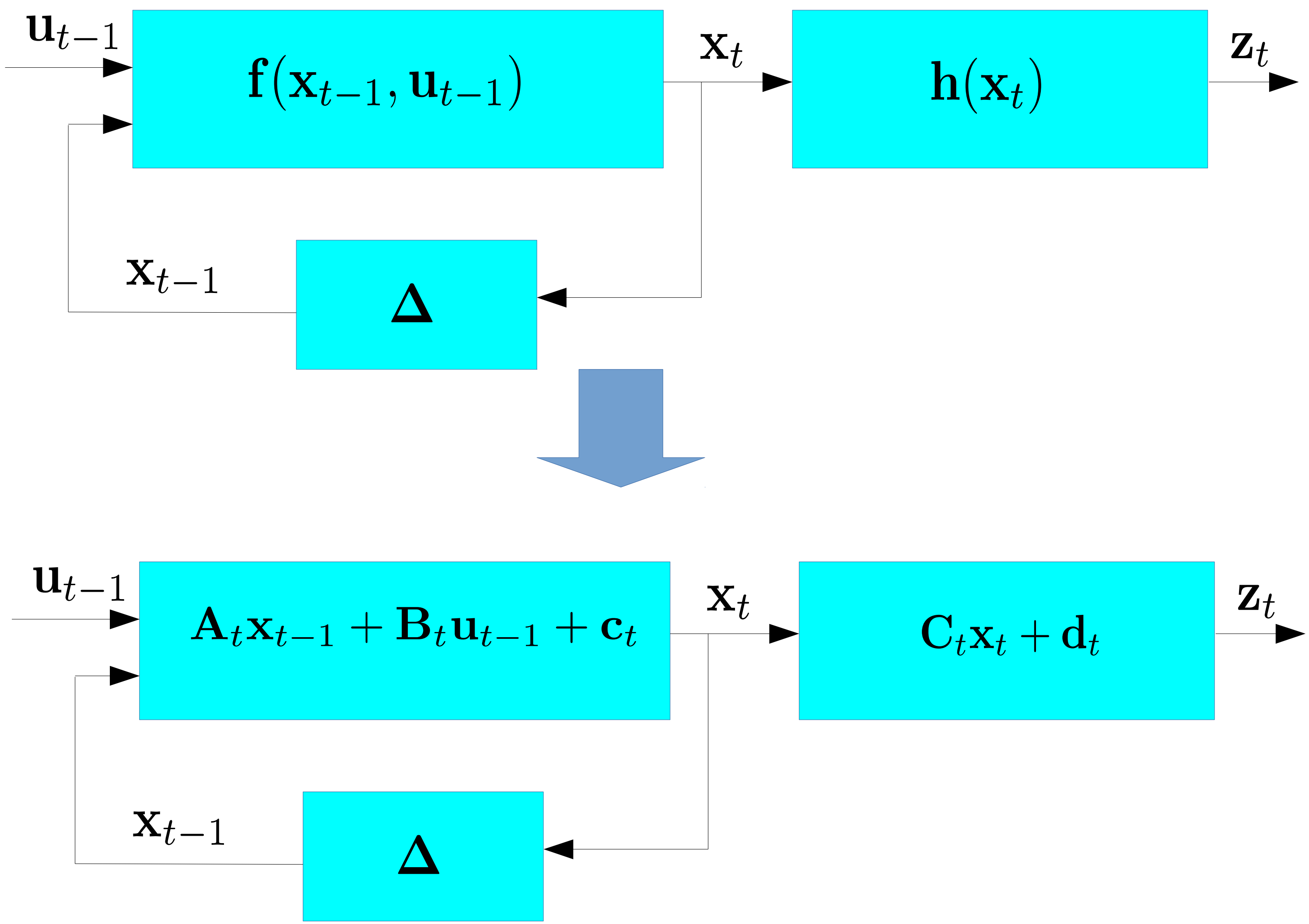
In contrast to the Kalman Filter, the EKF is not optimal, however

- for “smooth“ transition and observation models
- small uncertainties

it is shown to provide good results.

Linearizing a System

We want to construct a locally linear approximation of this model



Taylor Expansion

Linearize the functions around the current state and measurement, at each iteration

Derivatives are computed around the current mean of the estimate

$$
\begin{aligned}
\mathbf{f}(\mathbf{x},\mathbf{u}) &\simeq \mathbf{f}(\mathbf{x}_0,\mathbf{u}_0) + \underbrace{\frac{\partial \mathbf{f}(\mathbf{x},\mathbf{u})}{\partial \mathbf{x}}}_{\mathbf{A}}(\mathbf{x}-\mathbf{x}_0) + \underbrace{\frac{\partial \mathbf{f}(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}}}_{\mathbf{B}}(\mathbf{u}-\mathbf{u}_0) \\
&= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \underbrace{\mathbf{f}(\mathbf{x}_0,\mathbf{u}_0) - \mathbf{A}\mathbf{x}_0 - \mathbf{B}\mathbf{u}_0}_{\mathbf{c}} \\
\mathbf{h}(\mathbf{x}) &\simeq \mathbf{h}(\mathbf{x}_0) + \underbrace{\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}}_{\mathbf{C}}(\mathbf{x}-\mathbf{x}_0) \\
&= \mathbf{C}\mathbf{x} + \underbrace{\mathbf{h}(\mathbf{x}_0) - \mathbf{C}\mathbf{x}_0}_{\mathbf{d}}
\end{aligned}
$$

Wrapup (EKF)

Predict: incorporate new control

$$
\begin{aligned}
\mu_{t|t-1} &= \mathbf{f}(\mu_{t-1|t-1}, \mathbf{u}_{t-1}) \\
\mathbf{A}_t &= \left.\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mu_{t-1|t-1}} \\
\mathbf{B}_t &= \left.\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{u}}\right|_{\mathbf{u}=\mathbf{u}_{t-1}} \\
\mathbf{\Sigma}_{t|t-1} &= \mathbf{A}_t \mathbf{\Sigma}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{B}_t \mathbf{\Sigma}_u \mathbf{B}_t^T
\end{aligned}
$$

Update: incorporate new measurement

$$
\begin{aligned}
\mu_z &= \mathbf{h}(\mu_{t|t-1}) \\
\mathbf{C}_t &= \left.\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mu_{t|t-1}} \\
\mathbf{K}_t &= \mathbf{\Sigma}_{t|t-1} \mathbf{C}_t^T \left(\mathbf{\Sigma}_z + \mathbf{C}_t \mathbf{\Sigma}_{t|t-1} \mathbf{C}_t^T\right)^{-1} \\
\mu_{t|t} &= \mu_{t|t-1} + \mathbf{K}_t \left(\mathbf{z}_t - \mu_z\right) \\
\mathbf{\Sigma}_{t|t} &= \left(\mathbf{I} - \mathbf{K}_t \mathbf{C}_t\right) \mathbf{\Sigma}_{t|t-1}
\end{aligned}
$$

Summary

- The Kalman Filter is an optimal observer of the state for linear systems under Gaussian noise
- It is a Bayesian Filter, that operates in a closed Gaussian world by estimating the parameters of the state distribution as new measurements and controls become available
- For non-linear systems we can still use the Extended Kalman Filter, but we loose optimality
- Widely used