COMP6247 - **Reinforcement & Online Learning**

Online Learning

Week 4: Bayesian Inference

Christine Evers

School of Electronics & Computer Science
University of Southampton

Spring Semester 2020/2021

Lecture Overview

Week 4: Bayesian Inference

Part 1: Statistical Inference
Part 2: Frequentist Inference
Part 3: Bayesian Inference
Part 4: Bayesian Inference in the Wild...

Week 5: Monte Carlo method & MCMC

Week 6: Sequential Monte Carlo

Learning Outcomes

University of Southampton

Following this week's lecture, you should be able to:

1) **Distinguish between the frequentist and Bayesian paradigms**
2) **Understand the mathematical framework for Bayesian inference**
3) **Apply techniques for local linearisation to approximate integrals in non-linear state-spaces**

Notation

$t \in \{1,\ldots,T\}$:	Time index
$\mathbf{x}$:	$P \times 1$ data vector
$\mathbf{z}$:	$Q \times 1$ latent vector
$\boldsymbol{\theta}$:	Model parameters, e.g., $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ for Gaussian
$f(\mathbf{z})$:	Dynamical model. For linear models: $f(\mathbf{z}) = \mathbf{F}\,\mathbf{z}$
$h(\mathbf{z})$:	Observation model. For linear models: $h(\mathbf{z}) = \mathbf{H}\,\mathbf{z}$
$p(\,\cdot\,)$:	Probability density function (pdf)
$\mathbf{v}$:	Process noise
$\mathbf{Q}$:	Process noise covariance
$\mathbf{w}$:	Measurement noise
$\mathbf{R}$:	Measurement noise covariance
$\boldsymbol{\mu}$:	Mean of a Gaussian distribution
$\boldsymbol{\Sigma}$:	Covariance of a Gaussian distribution. Often: $\boldsymbol{\Sigma} = \sigma^2\,\mathbf{I}$
w:	Weight (e.g., of Gaussian mixture components or particles)
$\boldsymbol{I}$:	Identity matrix
$\mathbf{J}$:	Jacobian
$\mathcal{N}(\,\cdot\,)$:	Gaussian pdf

Notation

$t \in \{1, \ldots, T\}$:	Time index
$\mathbf{x}$:	$P \times 1$ data vector
$\mathbf{z}$:	$Q \times 1$ latent vector
$\boldsymbol{\theta}$:	Model parameters, e.g., $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ for Gaussian
$f(\mathbf{z})$:	Dynamical model. For linear models: $f(\mathbf{z}) = \mathbf{F}\,\mathbf{z}$
$h(\mathbf{z})$:	Observation model. For linear models: $h(\mathbf{z}) = \mathbf{H}\,\mathbf{z}$
$p(\,\cdot\,)$:	Probability density function (pdf)
$\mathbf{v}$:	Process noise
$\mathbf{Q}$:	Process noise covariance
$\mathbf{w}$:	Measurement noise
$\mathbf{R}$:	Measurement noise covariance
$\boldsymbol{\mu}$:	Mean of a Gaussian distribution
$\boldsymbol{\Sigma}$:	Covariance of a Gaussian distribution. Often: $\boldsymbol{\Sigma} = \sigma^2\,\mathbf{I}$
w:	Weight (e.g., of Gaussian mixture components or particles)
$\boldsymbol{I}$:	Identity matrix
$\mathbf{J}$:	Jacobian
$\mathcal{N}(\,\cdot\,)$:	Gaussian pdf

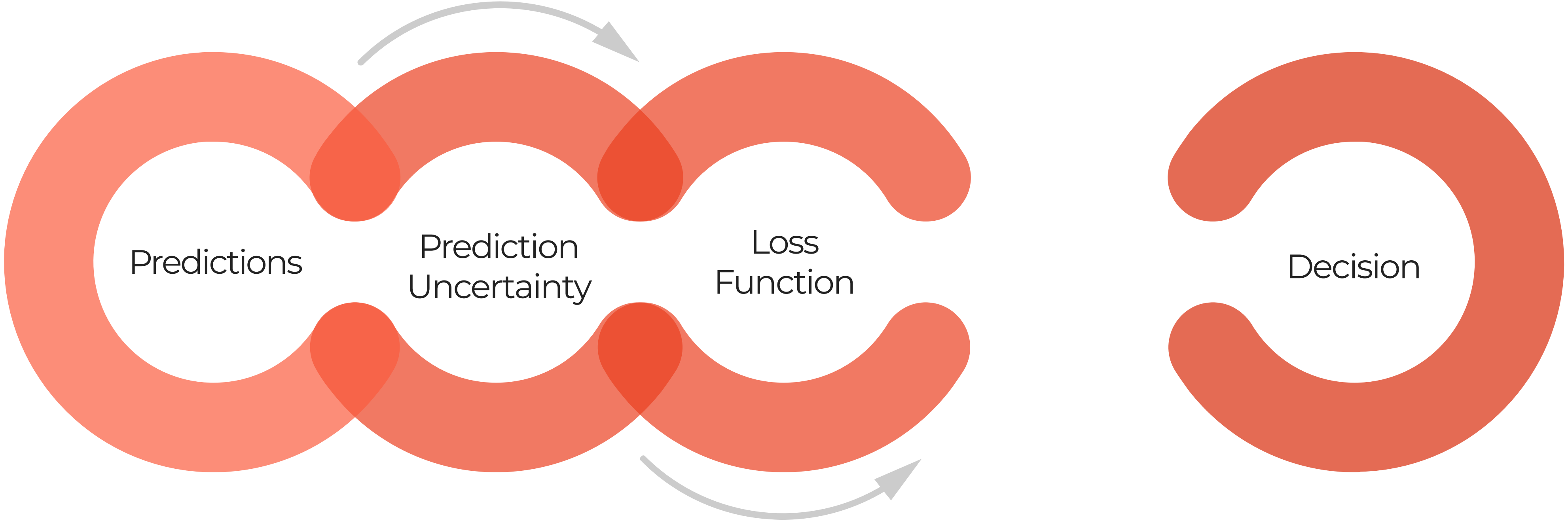
Statistical Inference

Decision Making

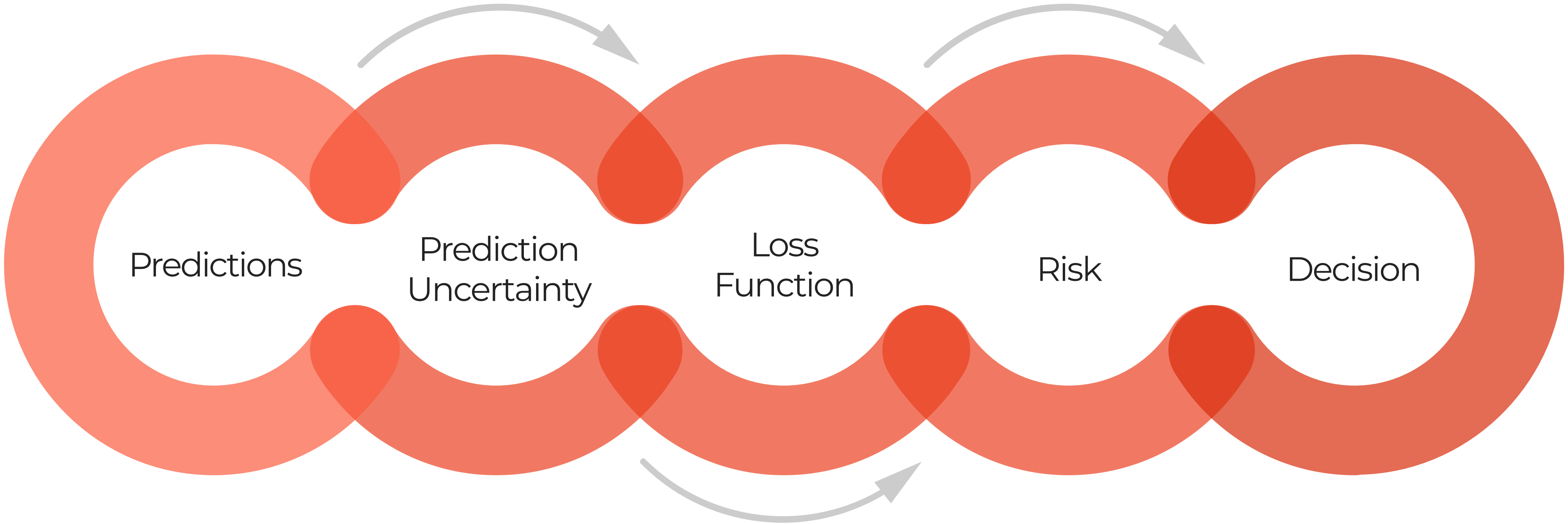
Decision Making

Decision Making

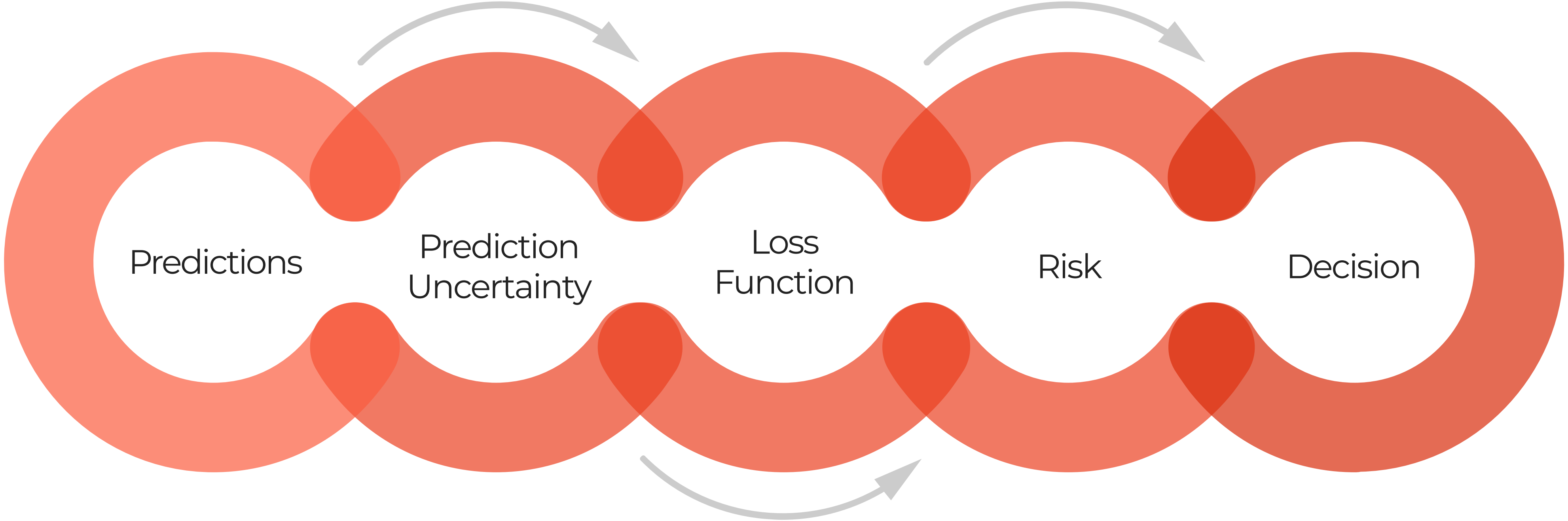
Decision Making



Decision Making

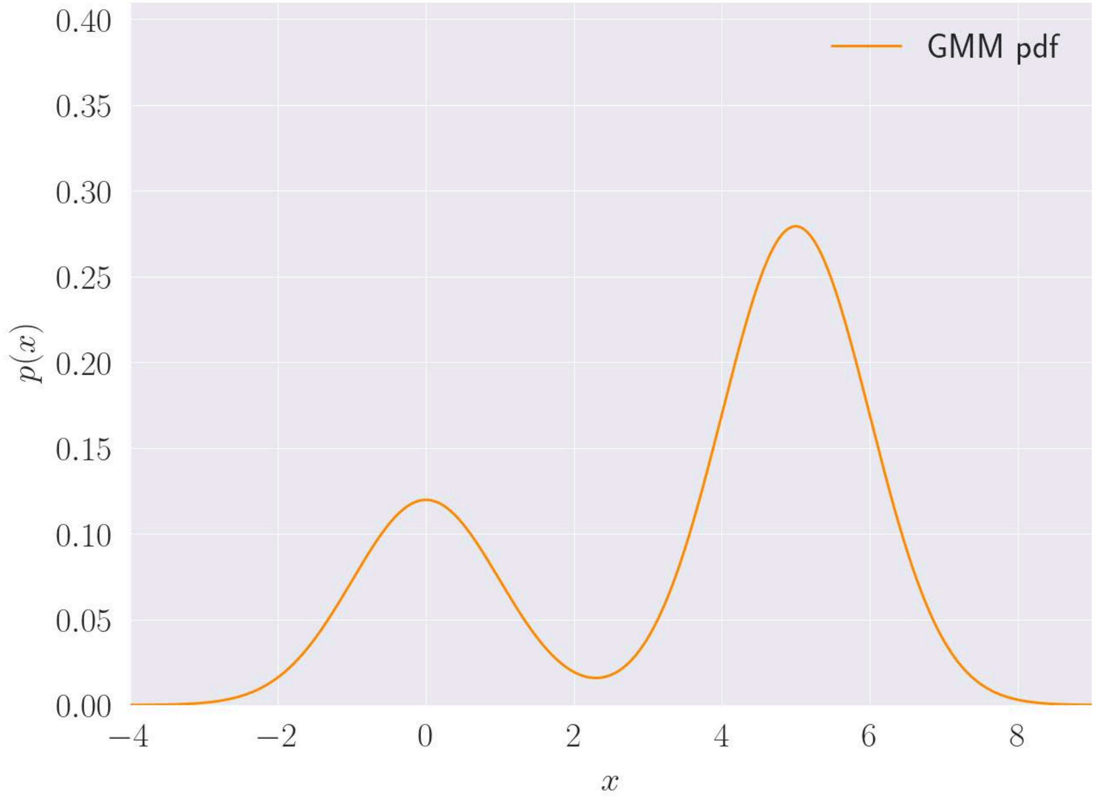


Decision Making



Rational decisions require quantification of prediction uncertainty!
Inference is statistical if associated with a measure of uncertainty

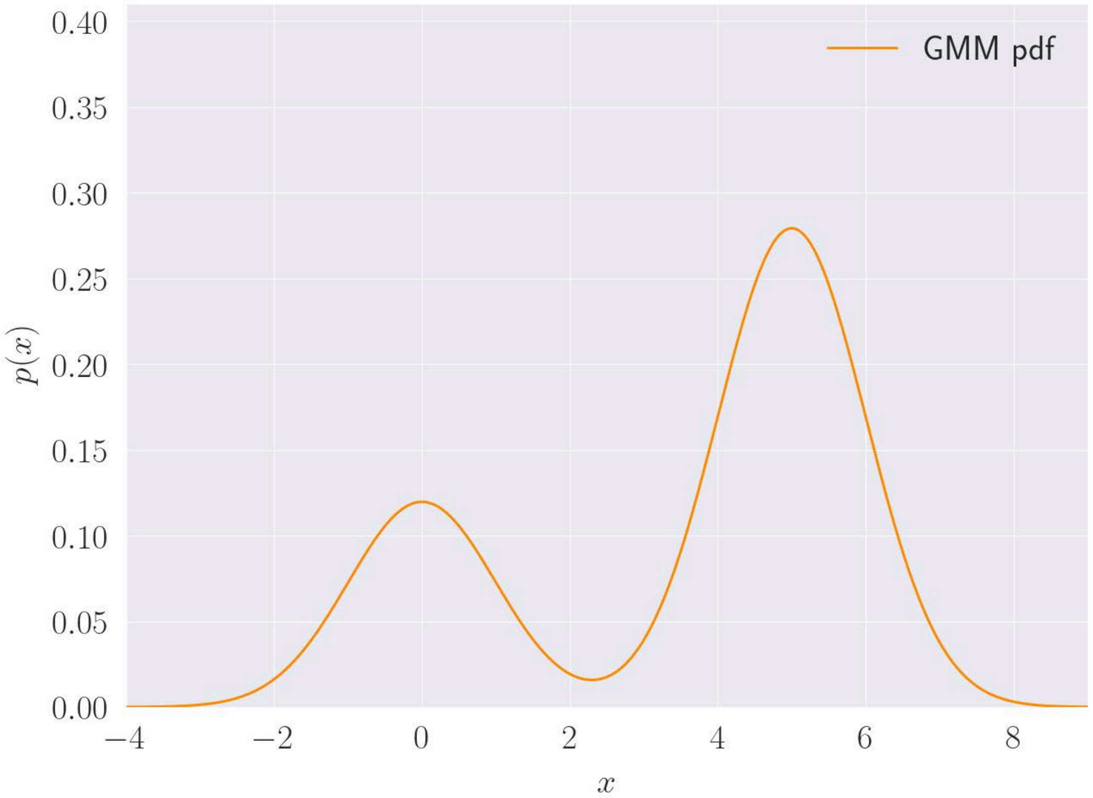
Probability Density Function



Example: Gaussian Mixture Model (GMM)

Probability Density Function

Probability density function, $p(\mathbf{y}; \boldsymbol{\theta})$, facilitates:

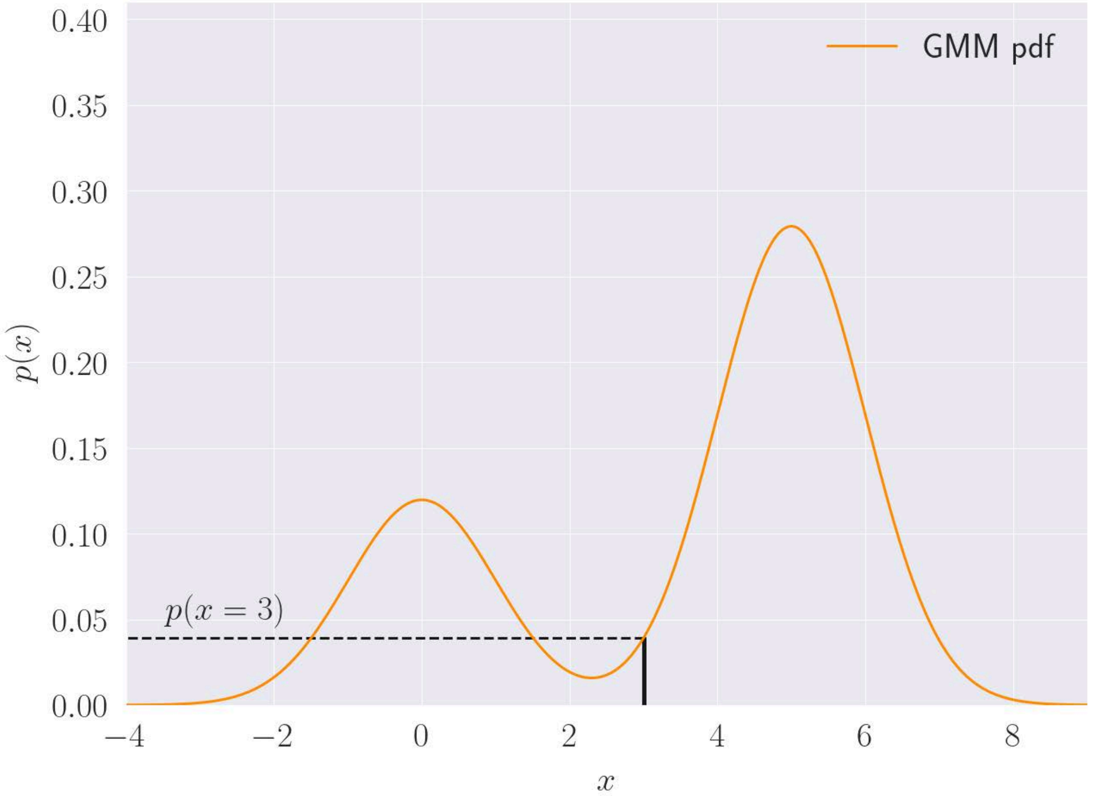


Example: Gaussian Mixture Model (GMM)

Probability Density Function

Probability density function, $p(\mathbf{y};\boldsymbol{\theta})$, facilitates:

- Predict probability of outcomes



Example: Gaussian Mixture Model (GMM)

Probability Density Function

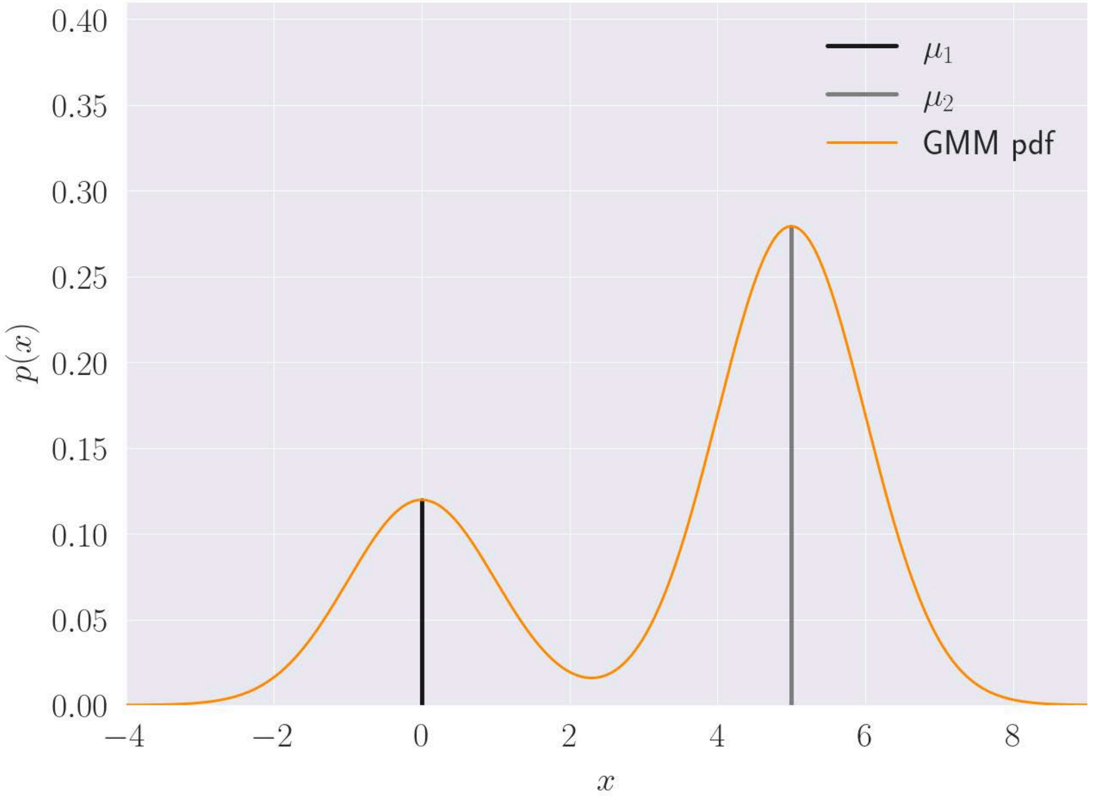
Probability density function, $p(\mathbf{y};\boldsymbol{\theta})$, facilitates:

- Predict probability of outcomes
- Evaluation of expectations:

$$\boldsymbol{\mu} = \mathbb{E}_p\left[\mathbf{y}\right] = \int_{\mathcal{Y}} \mathbf{y}\, p(\mathbf{y};\boldsymbol{\theta})\, d\mathbf{y}$$

$\mathbf{y}$: Realisation of random vector, Y with support $\mathcal{Y} \subset \mathbb{R}^d$,

$\tilde{\mathbf{y}}$: Random variate



Example: Gaussian Mixture Model (GMM)

Probability Density Function

Probability density function, $p(\mathbf{y};\boldsymbol{\theta})$, facilitates:

- Predict probability of outcomes
- Evaluation of expectations:

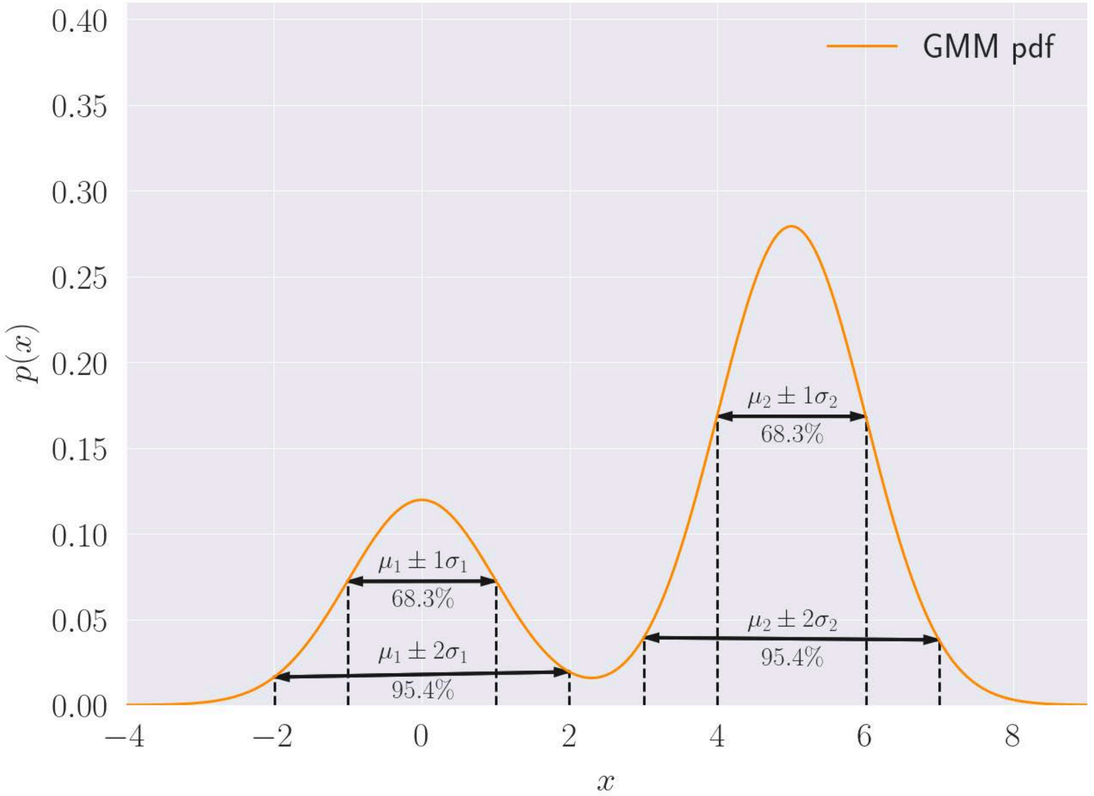
$$\boldsymbol{\mu} = \mathbb{E}_p\left[\mathbf{y}\right] = \int_{\mathcal{Y}} \mathbf{y}\, p(\mathbf{y};\boldsymbol{\theta})\, d\mathbf{y}$$

- Quantification of uncertainty:

$$\sigma^2 = \mathrm{var}_p\left[\mathbf{y}\right] = \mathbb{E}_p\left[\mathbf{y}^2\right] - \mathbb{E}_p\left[\mathbf{y}\right]^2 = \int_{\mathcal{Y}} \mathbf{y}^2\, p(\mathbf{y};\boldsymbol{\theta})\, d\mathbf{y} - \boldsymbol{\mu}$$

$\mathbf{y}$: Realisation of random vector, Y with support $\mathcal{Y} \subset \mathbb{R}^d$,

$\tilde{\mathbf{y}}$: Random variate



Example: Gaussian Mixture Model (GMM)

Probability Density Function

Probability density function, $p(\mathbf{y};\boldsymbol{\theta})$, facilitates:

- Predict probability of outcomes
- Evaluation of expectations:

$$\boldsymbol{\mu} = \mathbb{E}_p\left[\mathbf{y}\right] = \int_{\mathcal{Y}} \mathbf{y}\, p(\mathbf{y};\boldsymbol{\theta})\, d\mathbf{y}$$

- Quantification of uncertainty:

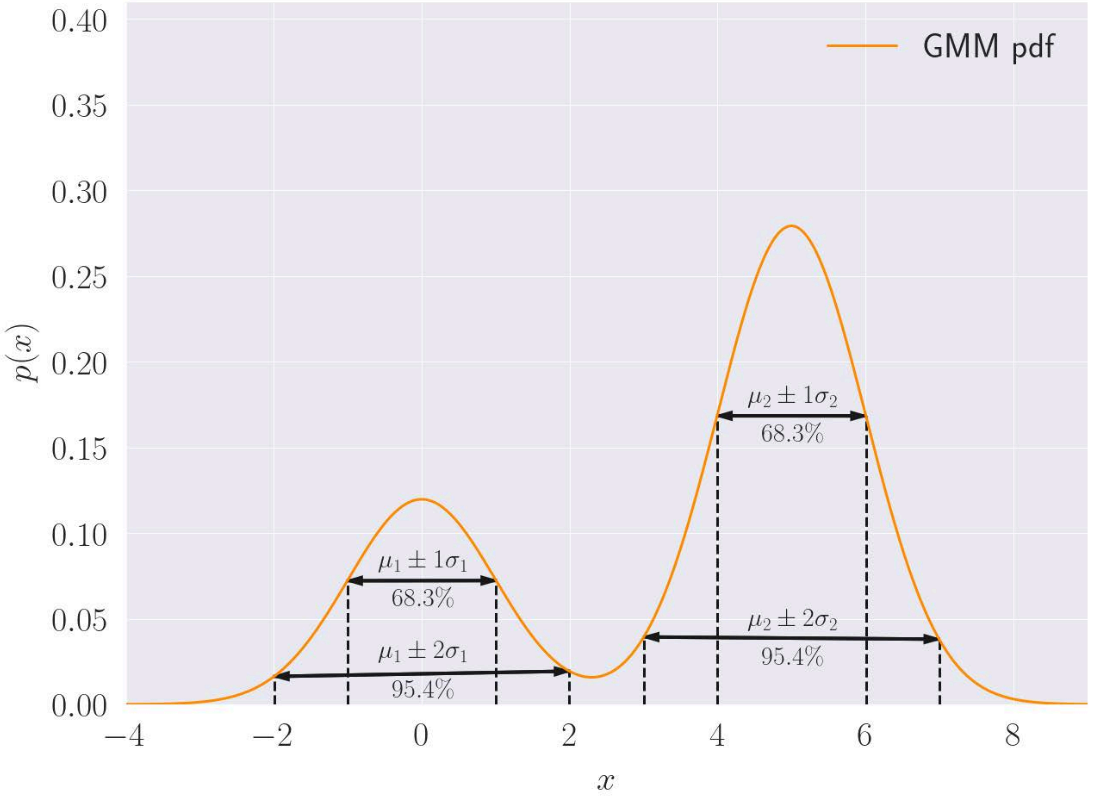
$$\sigma^2 = \mathrm{var}_p\left[\mathbf{y}\right] = \mathbb{E}_p\left[\mathbf{y}^2\right] - \mathbb{E}_p\left[\mathbf{y}\right]^2 = \int_{\mathcal{Y}} \mathbf{y}^2\, p(\mathbf{y};\boldsymbol{\theta})\, d\mathbf{y} - \boldsymbol{\mu}$$

- Generation of new data points:

$$\tilde{\mathbf{y}} \sim p(\mathbf{y};\boldsymbol{\theta})$$

$\mathbf{y}$: Realisation of random vector, Y with support $\mathcal{Y} \subset \mathbb{R}^d$,

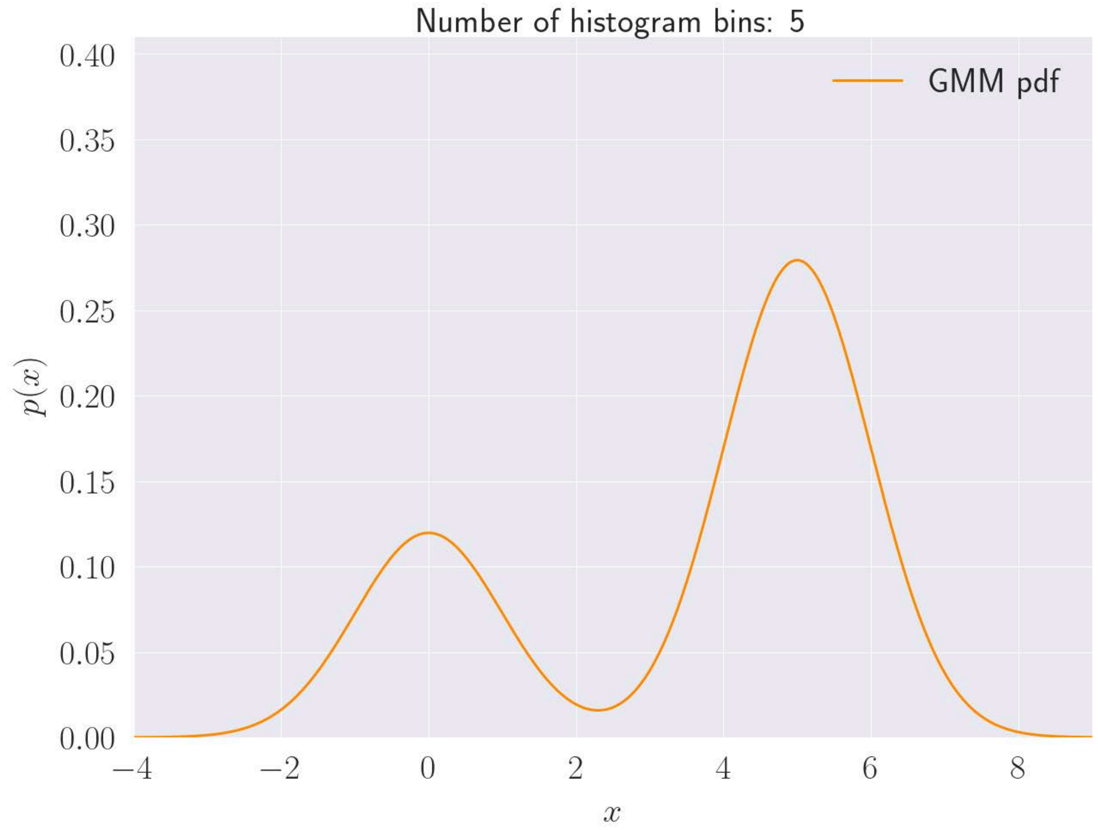
$\tilde{\mathbf{y}}$: Random variate



Example: Gaussian Mixture Model (GMM)

Frequentist Inference

Frequentist Density Estimation



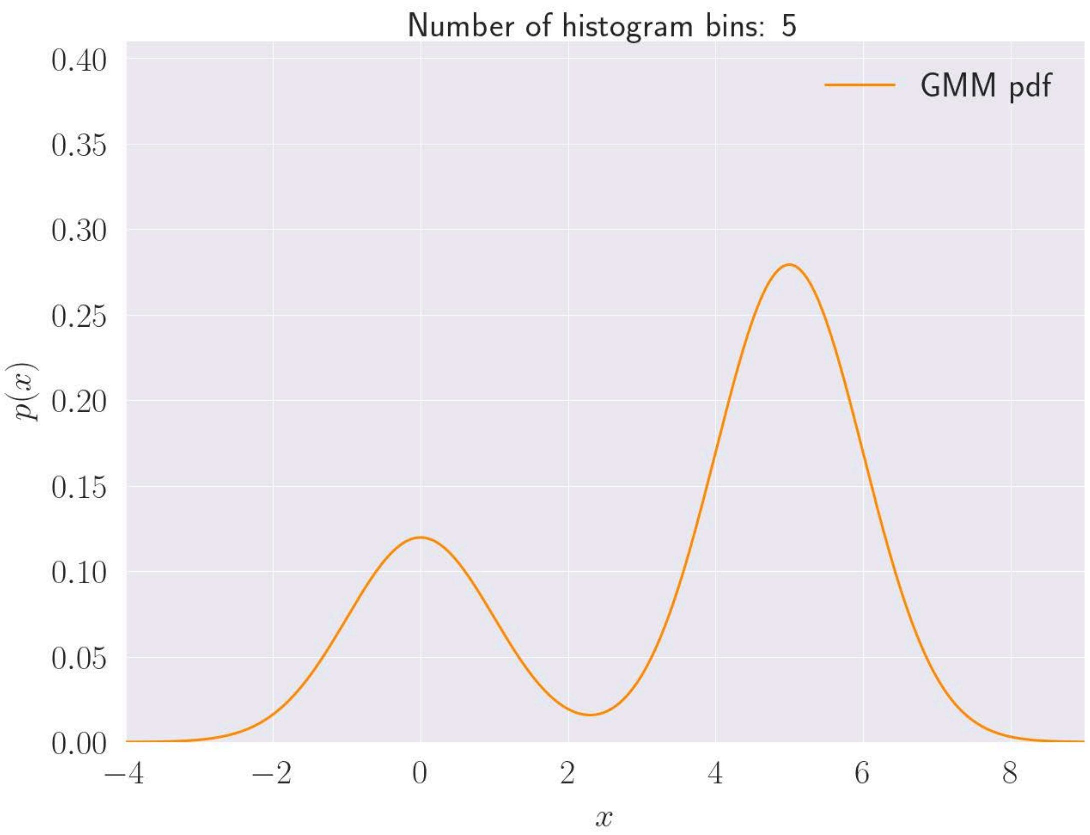
Example: Gaussian Mixture Model (GMM)

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

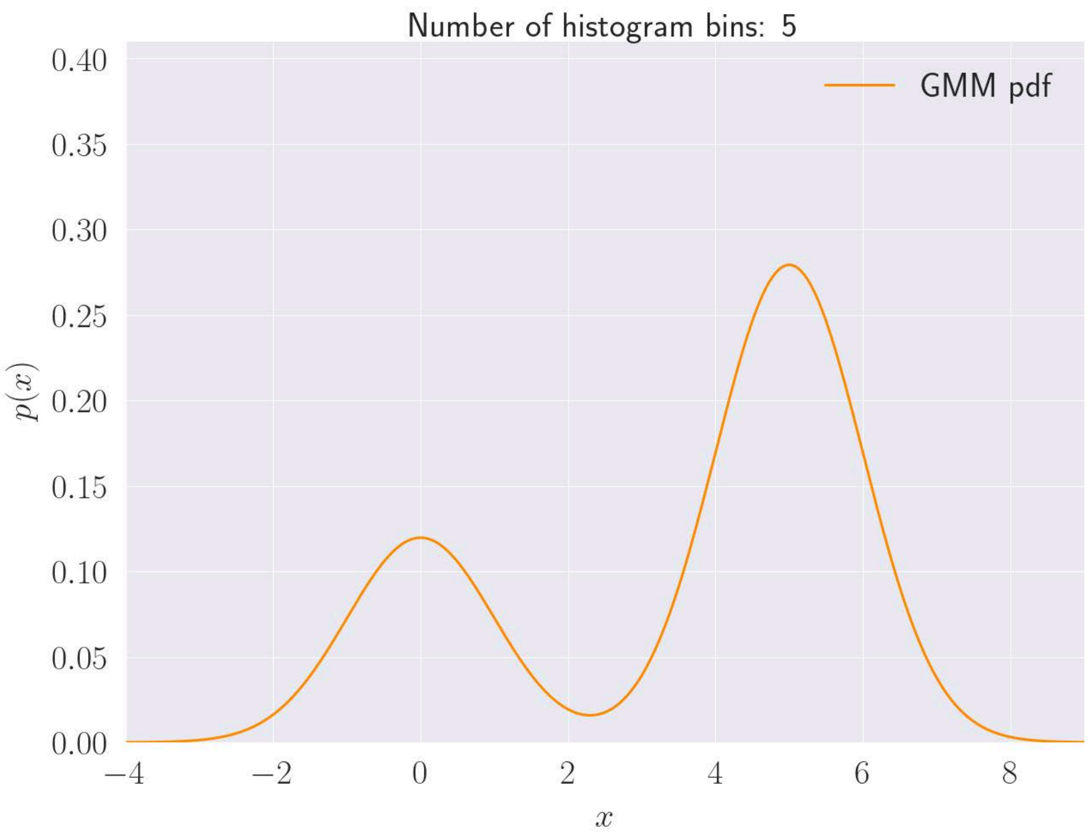
Frequentist Density Estimation

University of Southampton

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



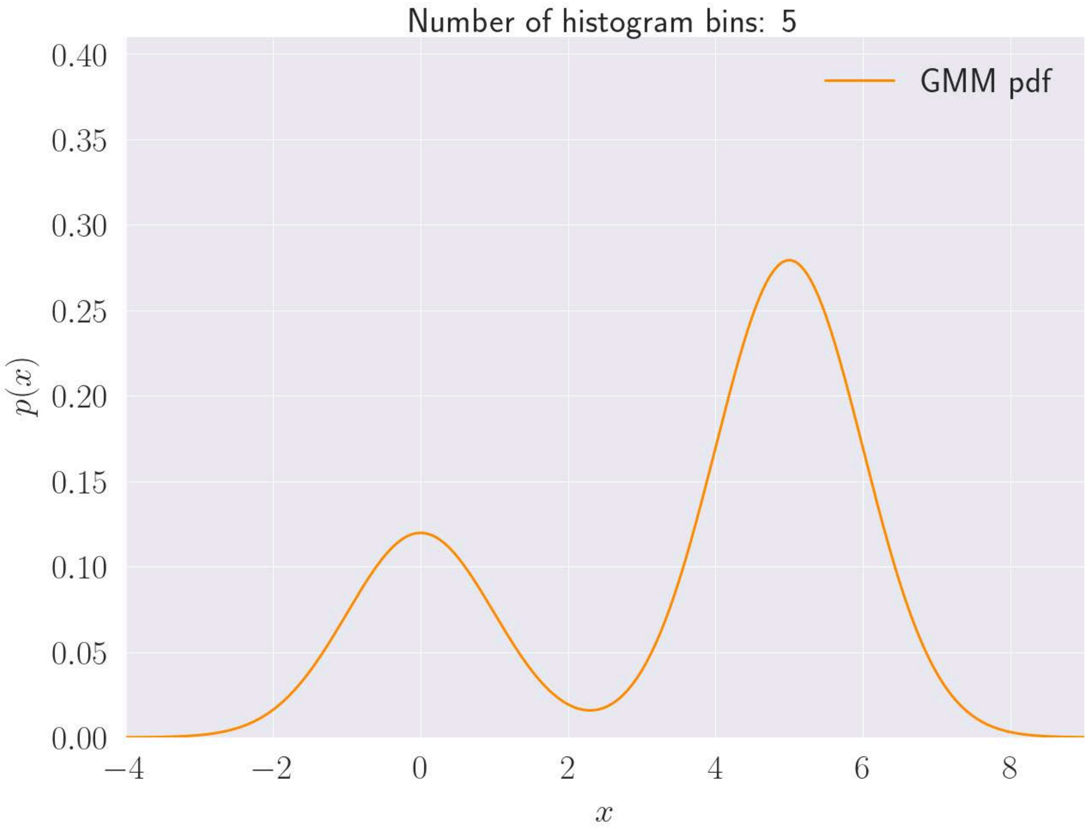
Example: Gaussian Mixture Model (GMM)

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

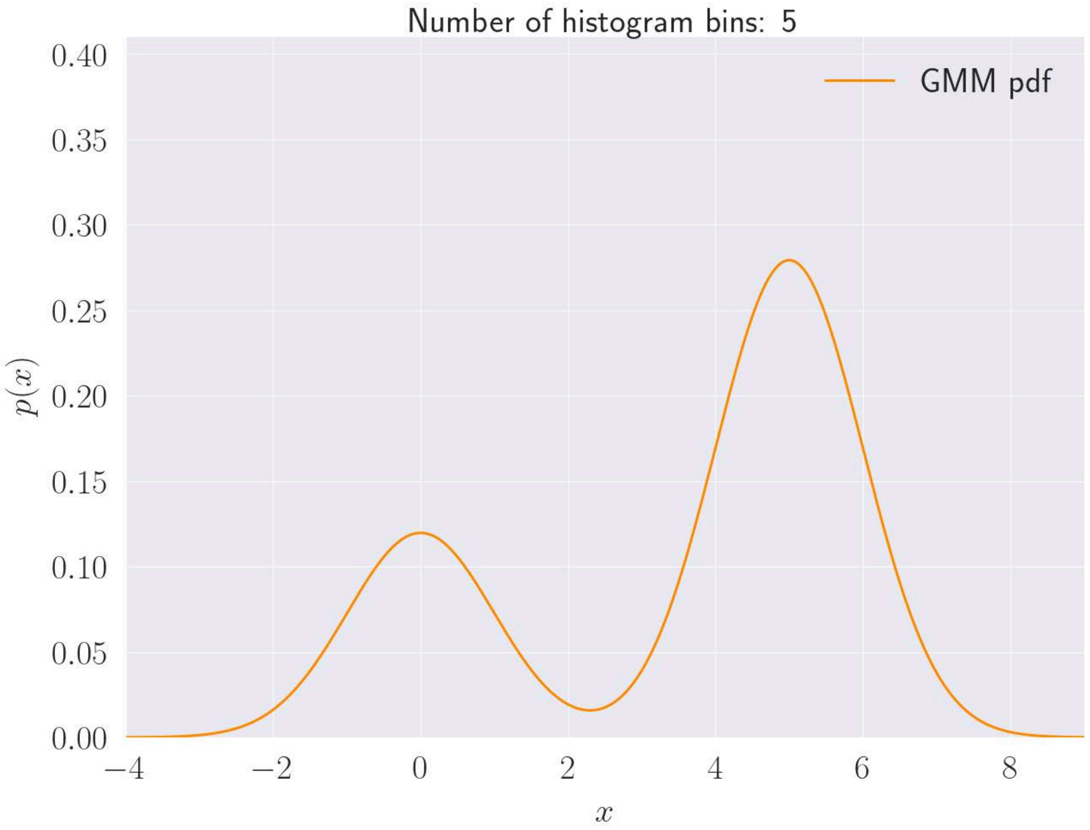
$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

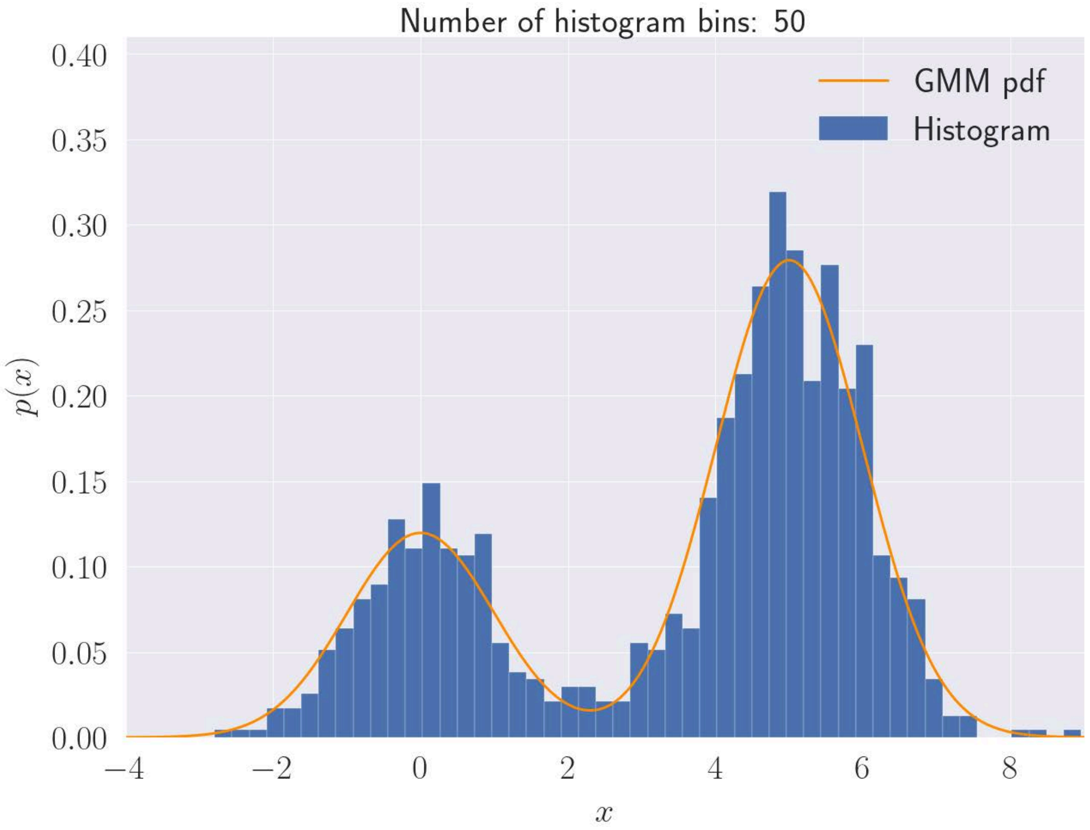
$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

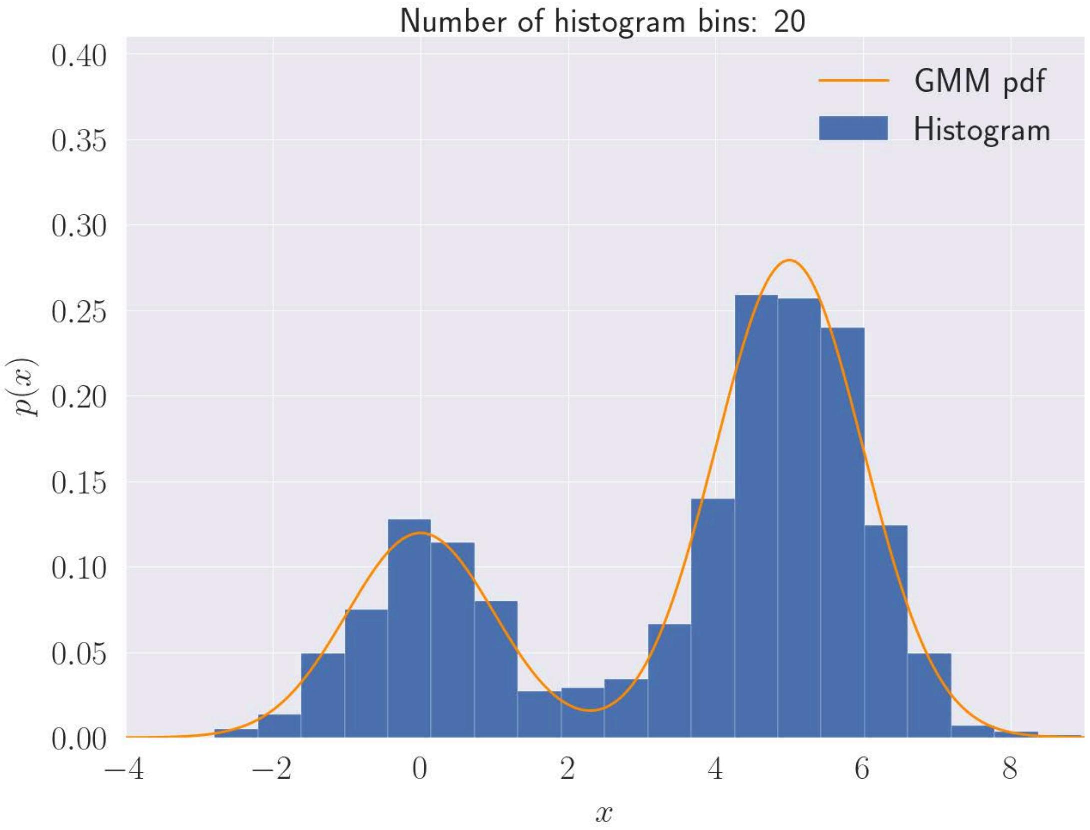
$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

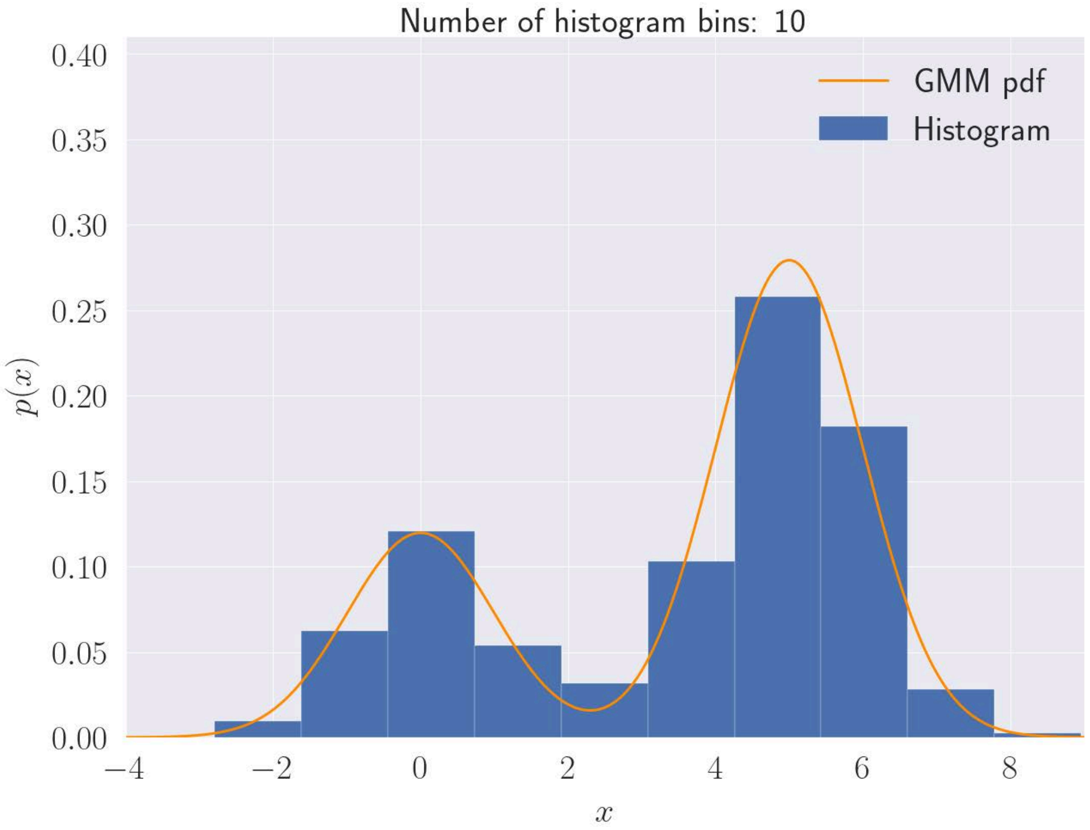
$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

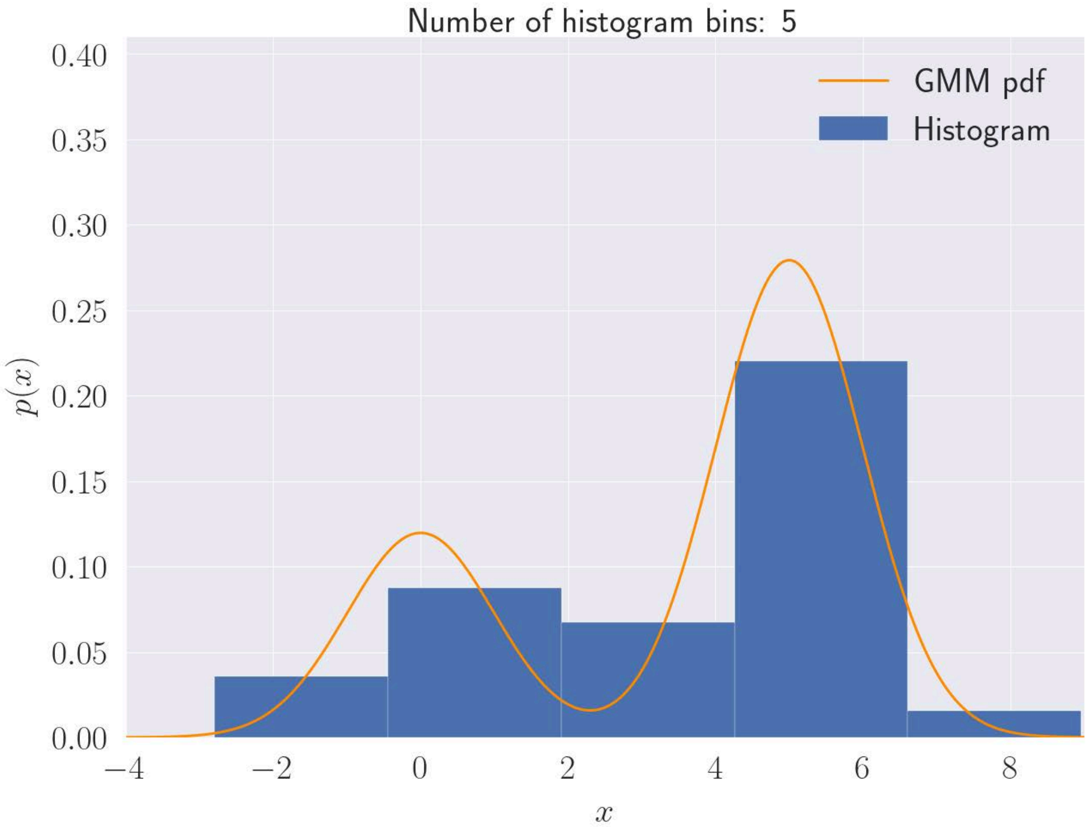
Frequentist Density Estimation

University of Southampton

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

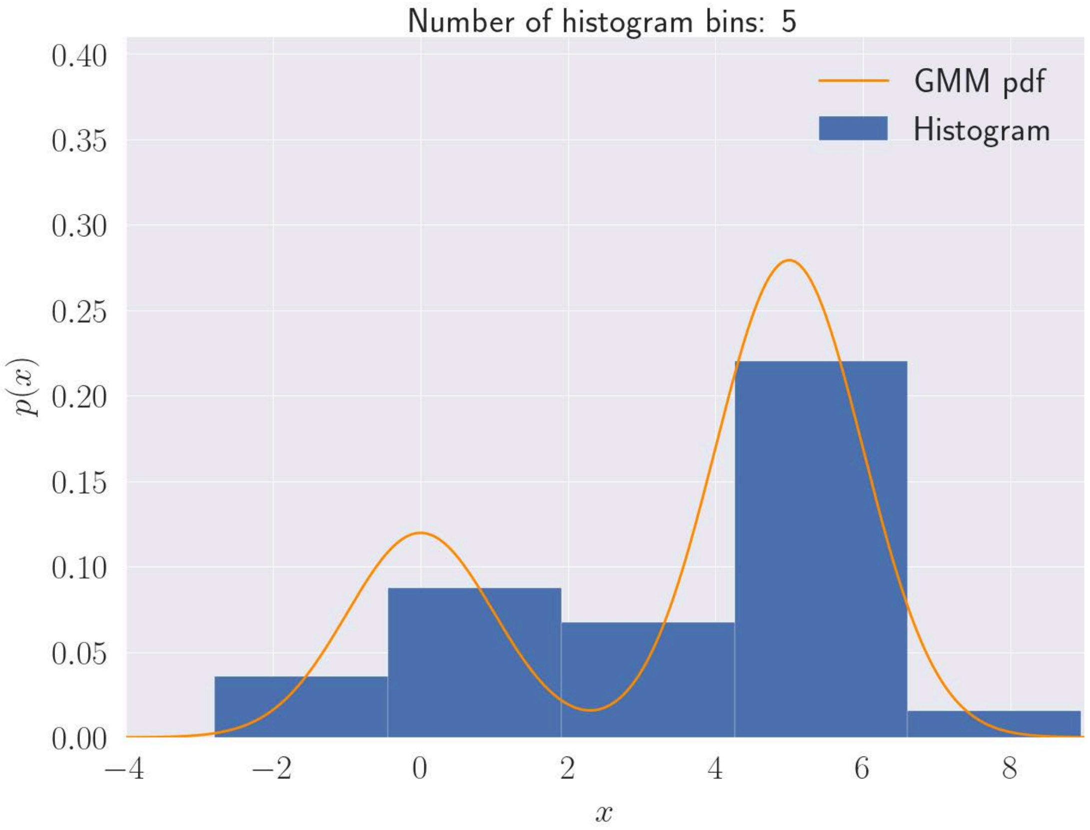
Frequentist Density Estimation

UNIVERSITY OF Southampton

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \,|\, \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function
- Model fitting:

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x;\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \,|\, \mu, \sigma^2\right)$$

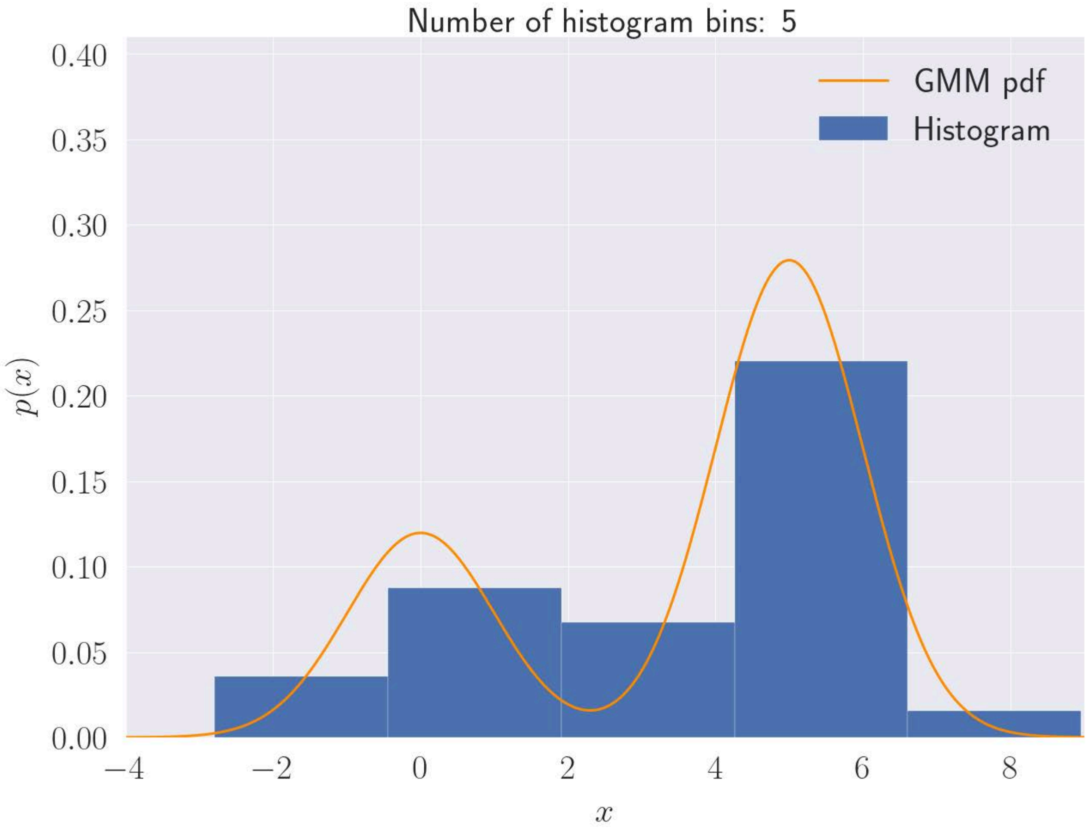
$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
    - Histogram: Probability mass function
- Model fitting:
    - Expectation Maximisation

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector



Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

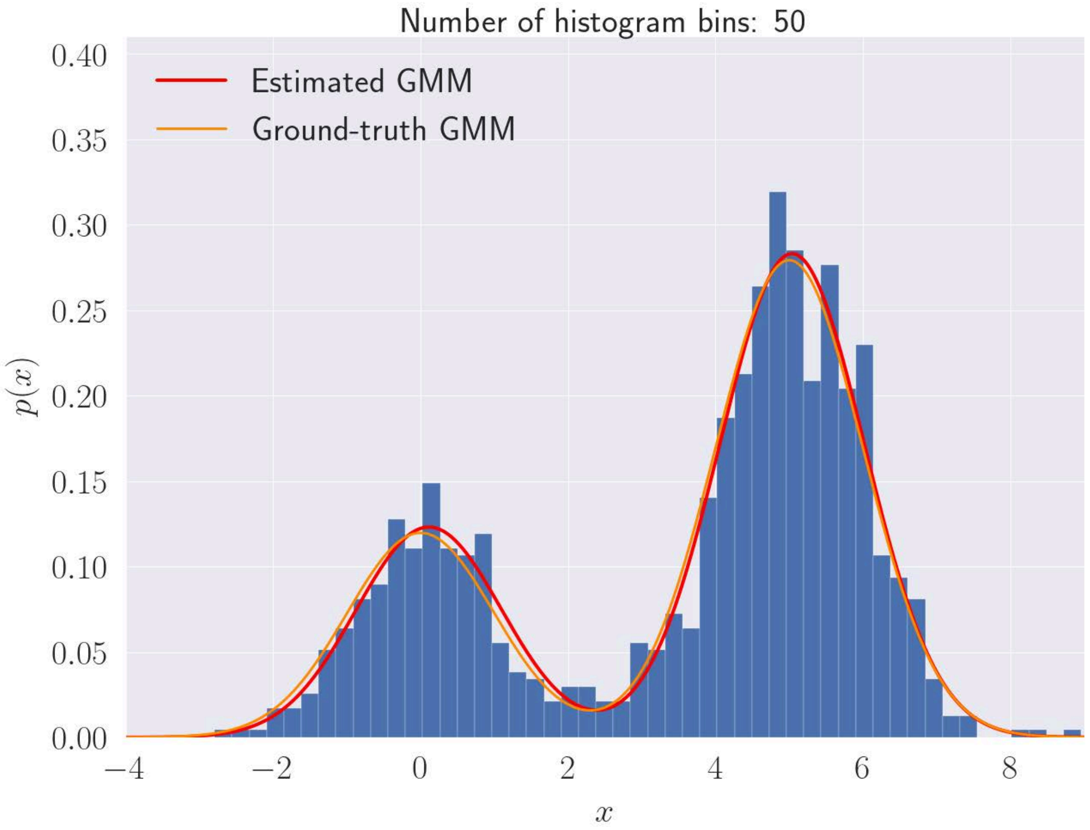
Frequentist Density Estimation

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$
- Density Estimation, e.g.,
  - Histogram: Probability mass function
- Model fitting:
  - Expectation Maximisation
  - Point estimation: Maximum likelihood

$$\hat{\boldsymbol{\theta}}^{\text{ML}} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$$

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

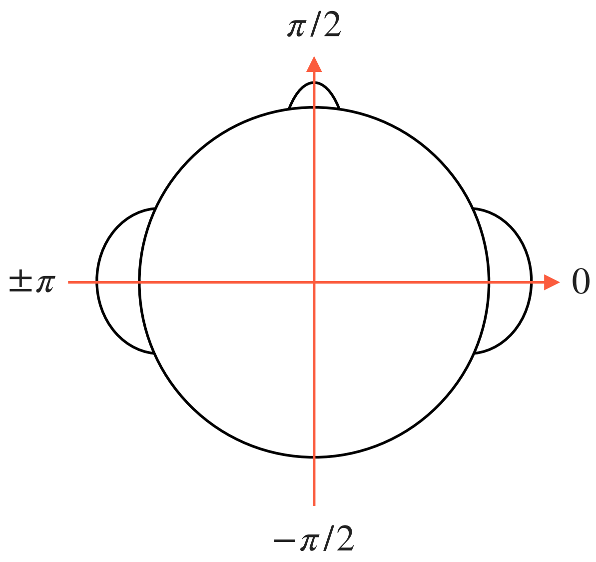


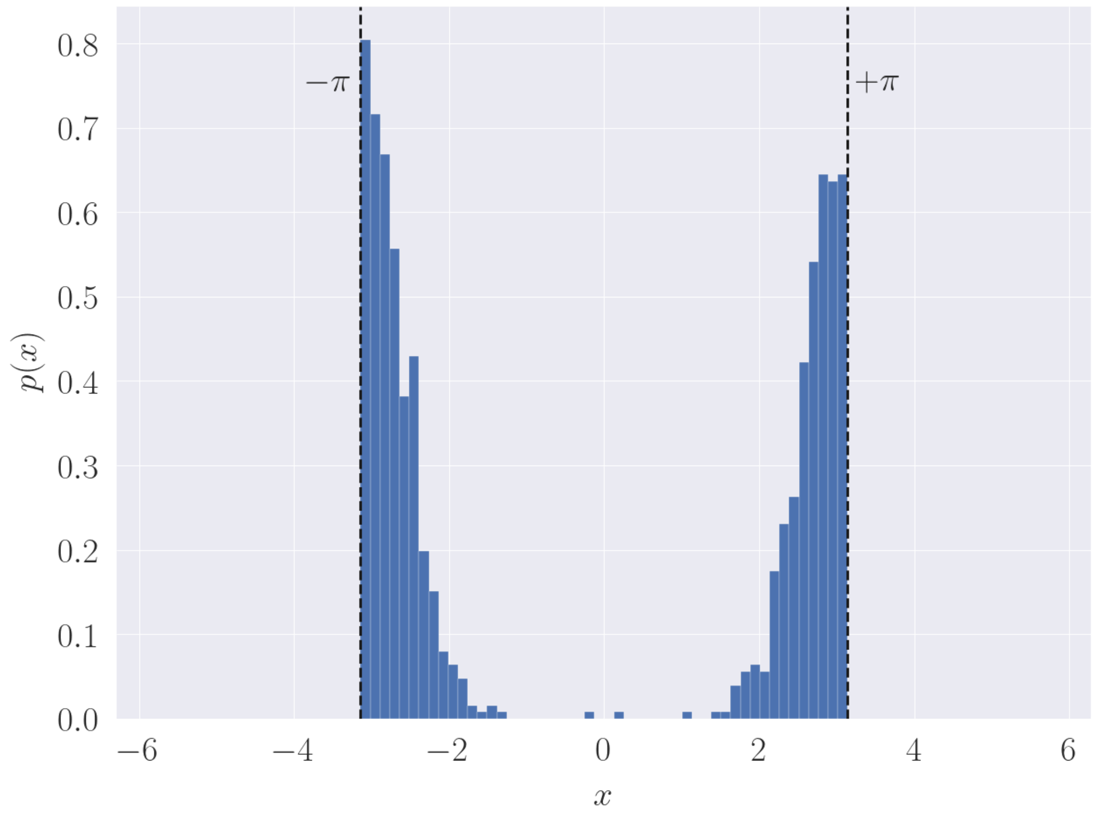
Example: Gaussian Mixture Model (GMM)

$$p(x; \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \, \mathcal{N}\left(x \mid \mu, \sigma^2\right)$$

$$\boldsymbol{\theta} = \left\{(w_1, \mu_1, \sigma_1^2), (w_2, \mu_2, \sigma_2^2)\right\}$$

Frequentist Density Estimation



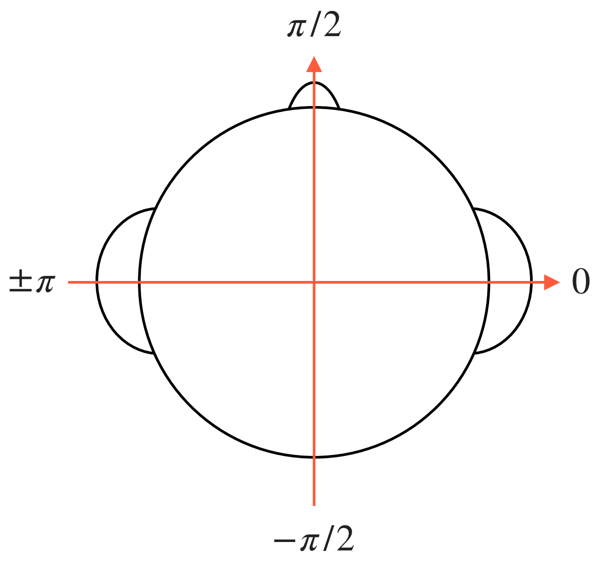


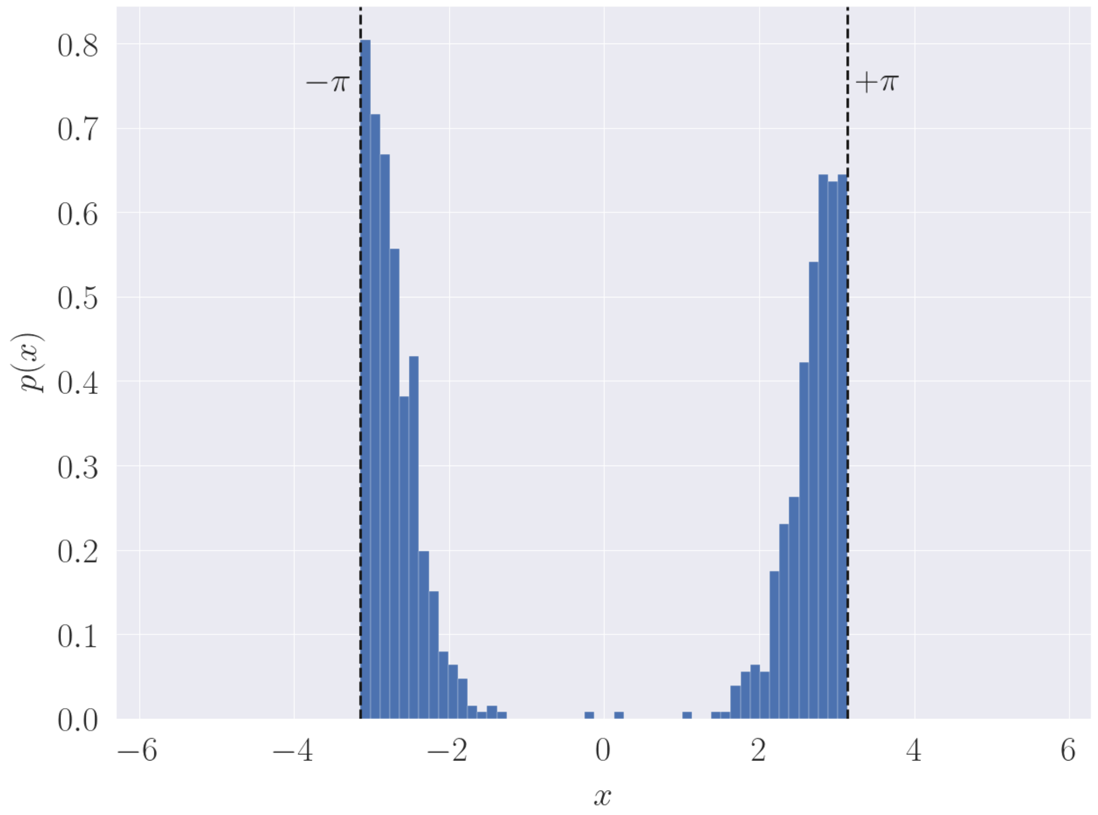
Frequentist Density Estimation

Von Mises Distribution:

$$p(x \mid \mathbf{z}; \boldsymbol{\theta}) = \frac{\exp\left\{\kappa \, \cos(x - \mu)\right\}}{2\pi \, I_0(\kappa)}$$

$\boldsymbol{\theta} = \{\kappa, \mu\}$, $\quad -\pi \leq x \leq \pi$: Angle, $\quad \kappa > 0$: Concentration,
$I_0(\cdot)$: Modified Bessel function of order 0



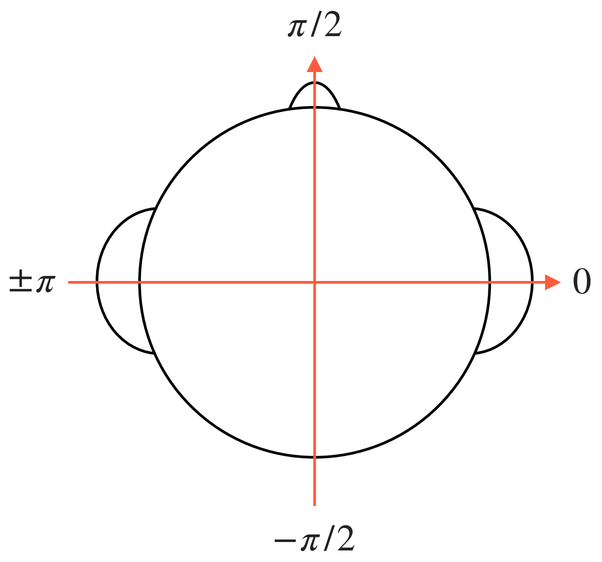


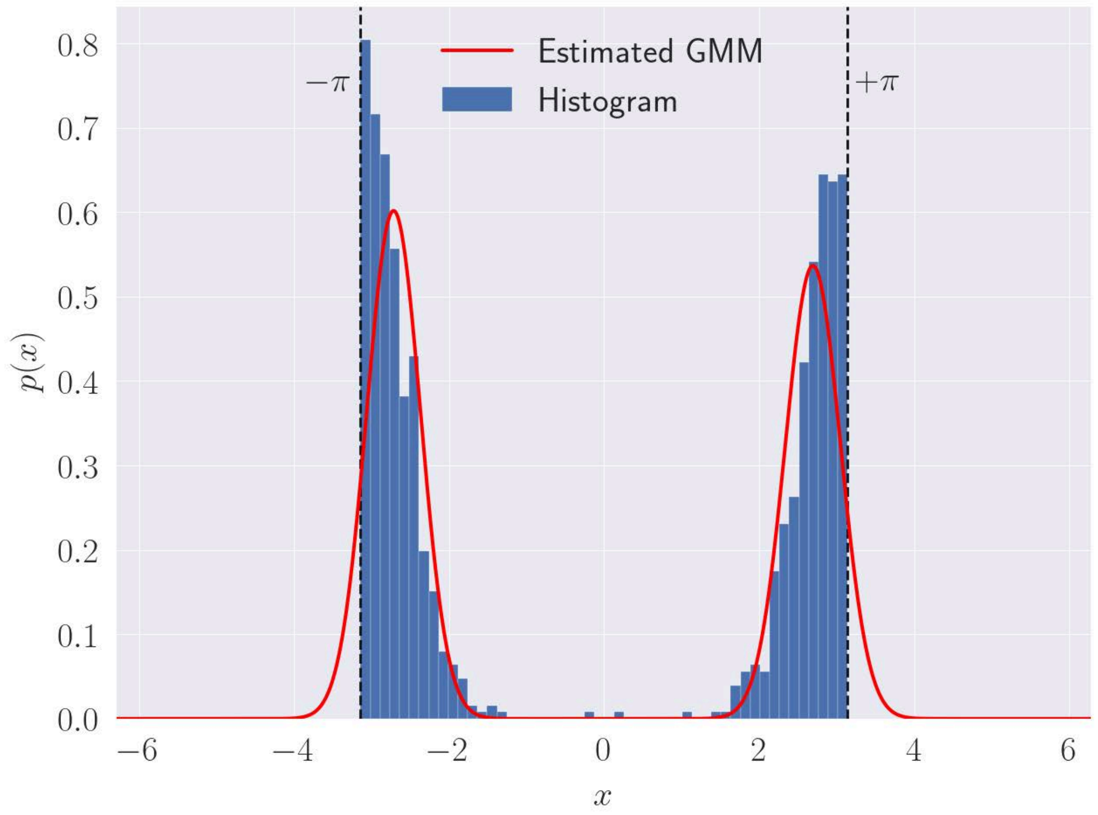
Frequentist Density Estimation

Von Mises Distribution:

$$p(x \mid \mathbf{z}; \boldsymbol{\theta}) = \frac{\exp\left\{\kappa \, \cos(x - \mu)\right\}}{2\pi \, I_0(\kappa)}$$

$\boldsymbol{\theta} = \{\kappa, \mu\}$, $\quad -\pi \leq x \leq \pi :$ Angle, $\quad \kappa > 0 :$ Concentration,
$I_0(\cdot) :$ Modified Bessel function of order 0





Bayesian Inference

Bayesian Inference

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x};\, \boldsymbol{\theta}) = \frac{p(\mathbf{x} \mid \mathbf{z};\, \boldsymbol{\theta})\, p\left(\mathbf{z};\, \boldsymbol{\theta}\right)}{p(\mathbf{x};\, \boldsymbol{\theta})}$$

Posterior pdf quantifies belief in the model of the underlying process in light of the data

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Bayesian Inference

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}) = \frac{\overbrace{p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})}^{\text{Likelihood}} \, p\left(\mathbf{z}; \boldsymbol{\theta}\right)}{p(\mathbf{x}; \boldsymbol{\theta})}$$

Posterior pdf quantifies belief in the model of the underlying process in light of the data

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Likelihood function: Data generation model

Bayesian Inference

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta}) = \frac{\overbrace{p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})}^{\text{Likelihood}} \, \overbrace{p\left(\mathbf{z}; \boldsymbol{\theta}\right)}^{\text{Prior pdf}}}{p(\mathbf{x}; \boldsymbol{\theta})}$$

Posterior pdf quantifies belief in the model of the underlying process in light of the data

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Likelihood function: Data generation model
Prior pdf: Subjective belief

Bayesian Inference

Bayes's Theorem:

$$p(\mathbf{z} \,|\, \mathbf{x};\, \boldsymbol{\theta}) = \frac{\overbrace{p(\mathbf{x} \,|\, \mathbf{z};\, \boldsymbol{\theta})}^{\text{Likelihood}}\, \overbrace{p\left(\mathbf{z};\, \boldsymbol{\theta}\right)}^{\text{Prior pdf}}}{\underbrace{p(\mathbf{x};\, \boldsymbol{\theta})}_{\text{Evidence}}}$$

Posterior pdf quantifies belief in the model of the underlying process in light of the data

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Likelihood function: Data generation model
Prior pdf: Subjective belief
Evidence: Ensures that the posterior is a valid pdf. Marginal likelihood:

$$p(\mathbf{x};\boldsymbol{\theta}) = \int_{\mathcal{Z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})\, d\mathbf{z}$$

Bayesian Inference

Bayes's Theorem:

$$p(\mathbf{z} \mid \mathbf{x};\, \boldsymbol{\theta}) = \frac{\overbrace{p(\mathbf{x} \mid \mathbf{z};\, \boldsymbol{\theta})}^{\text{Likelihood}}\, \overbrace{p\left(\mathbf{z};\, \boldsymbol{\theta}\right)}^{\text{Prior pdf}}}{\underbrace{p(\mathbf{x};\, \boldsymbol{\theta})}_{\text{Evidence}}}$$

Posterior pdf quantifies belief in the model of the underlying process in light of the data

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Likelihood function: Data generation model
Prior pdf: Subjective belief
Evidence: Ensures that the posterior is a valid pdf. Marginal likelihood:

$$p(\mathbf{x};\boldsymbol{\theta}) = \int_{\mathcal{Z}} p(\mathbf{x},\mathbf{z};\boldsymbol{\theta})\, d\mathbf{z} = \int_{\mathcal{Z}} p(\mathbf{x} \mid \mathbf{z};\boldsymbol{\theta})\, p\left(\mathbf{z};\boldsymbol{\theta}\right)\, d\mathbf{z}$$

Bayesian vs Frequentist Paradigm

Bayesian Paradigm

Frequentist Paradigm:

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Bayesian vs Frequentist Paradigm

Bayesian Paradigm

- Incorporates subjective belief via prior pdf, $p(\mathbf{z}; \boldsymbol{\theta})$

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Bayesian vs Frequentist Paradigm

Bayesian Paradigm

- Incorporates subjective belief via prior pdf, $p(\mathbf{z}; \boldsymbol{\theta})$
- Relies on the posterior pdf, $p(\mathbf{z} \mid \mathbf{x}; \boldsymbol{\theta})$

Frequentist Paradigm:

- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \mid \mathbf{z}; \boldsymbol{\theta})$

$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Bayesian vs Frequentist Paradigm

Bayesian Paradigm

- Incorporates subjective belief via prior pdf, $p(\mathbf{z};\, \boldsymbol{\theta})$
- Relies on the posterior pdf, $p(\mathbf{z} \,|\, \mathbf{x};\, \boldsymbol{\theta})$
- Maximum *a posteriori* (MAP) estimates:

$$\hat{\mathbf{z}}^{\text{MAP}} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{z} \,|\, \mathbf{x};\, \boldsymbol{\theta})$$

- *Principled framework for inference of hidden variables*

Frequentist Paradigm:

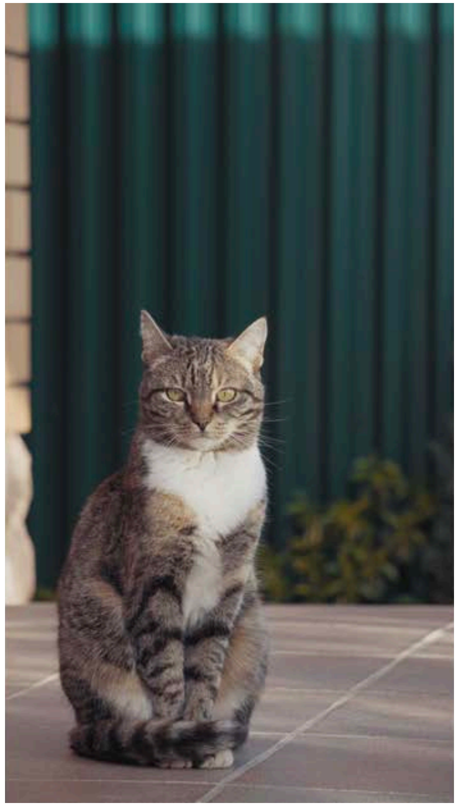
- Long-term expected frequency of the occurrence of a random event
- Relies on likelihood function, $p(\mathbf{x} \,|\, \mathbf{z};\, \boldsymbol{\theta})$
- Maximum likelihood (ML) estimates:

$$\hat{\boldsymbol{\theta}}^{\text{ML}} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{x} \,|\, \mathbf{z};\, \boldsymbol{\theta})$$

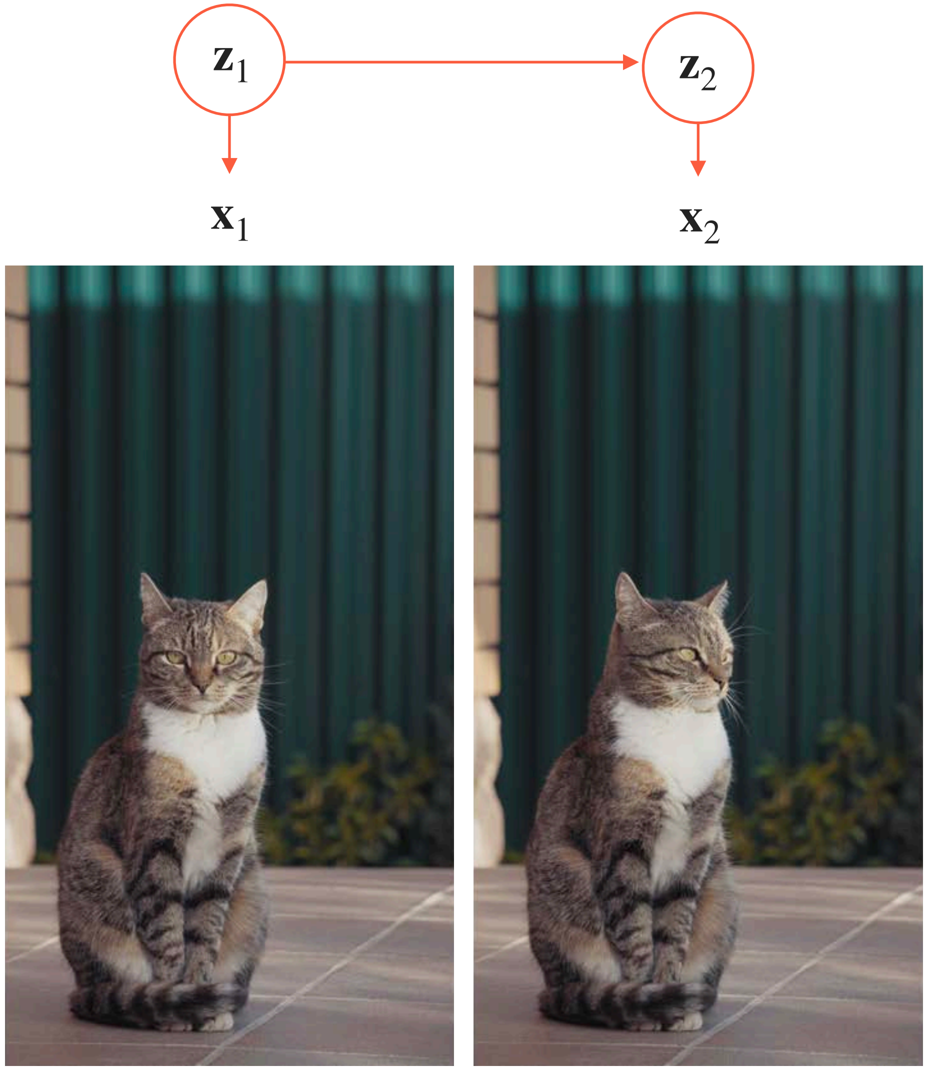
$\mathbf{x}$: Data vector, $\boldsymbol{\theta}$: Model parameters, $\mathbf{z}$: Latent vector

Predictions

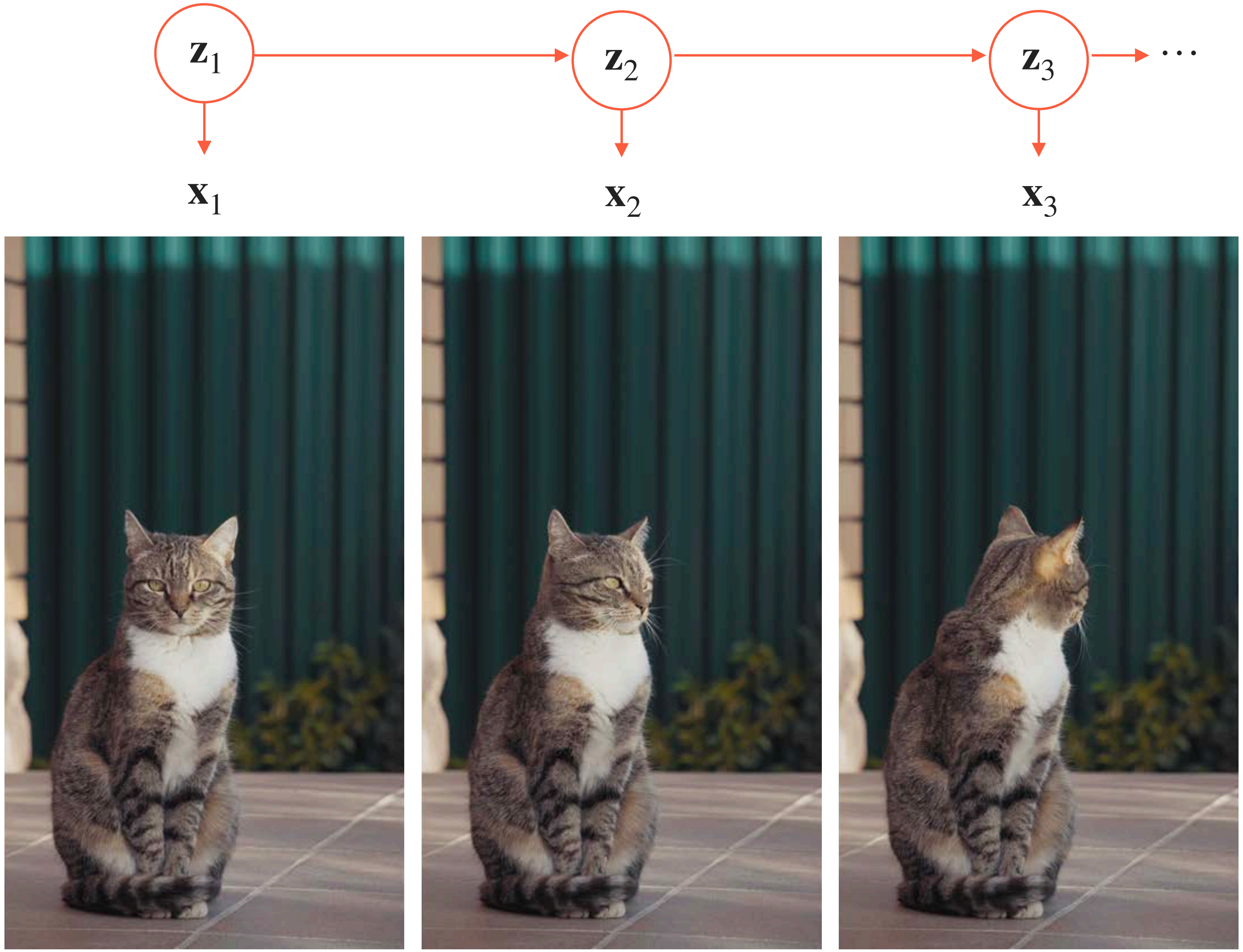
$$\mathbf{z}_1 \rightarrow \mathbf{x}_1$$



Predictions



Predictions

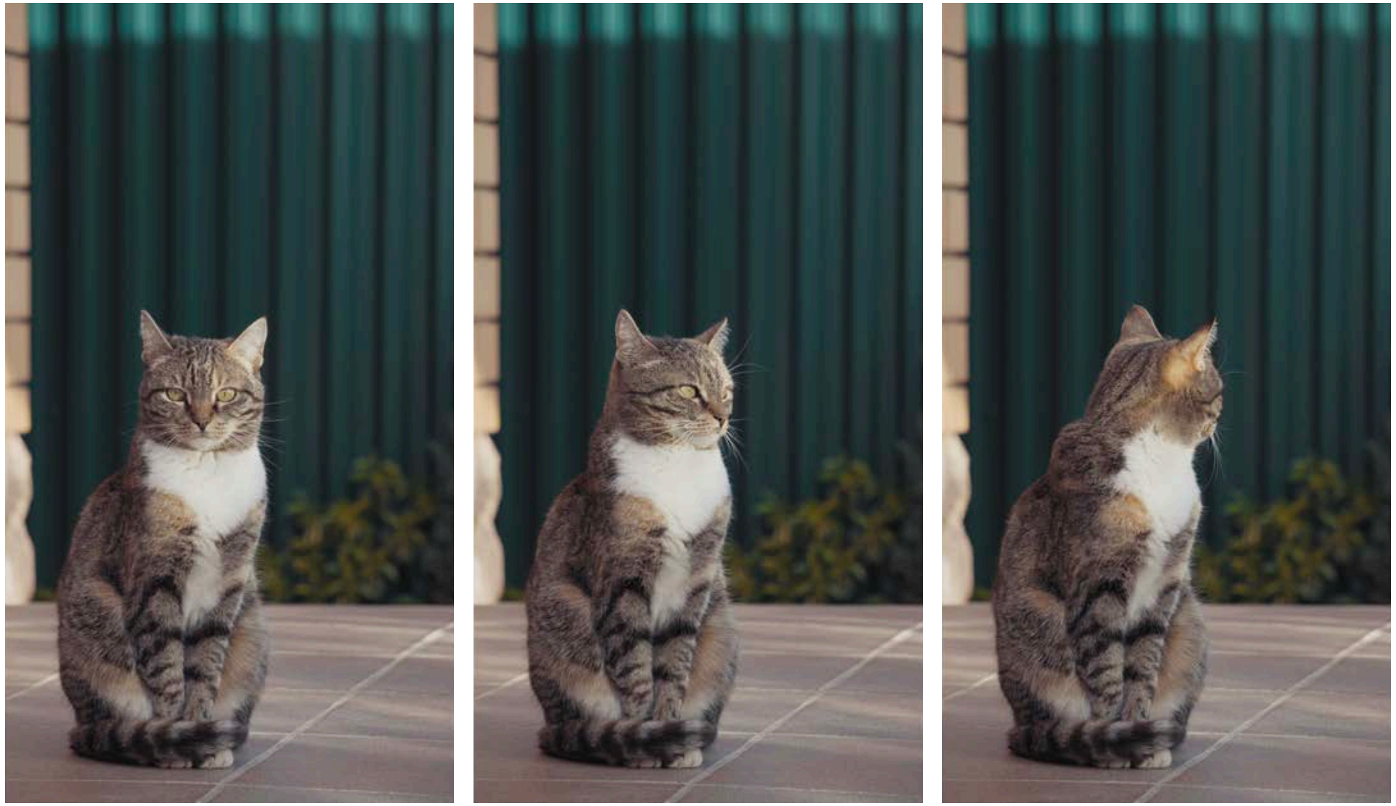


Predictions

UNIVERSITY OF Southampton

$$
\begin{tikzcd}
\mathbf{z}_1 \arrow[r] \arrow[d] & \mathbf{z}_2 \arrow[r] \arrow[d] & \mathbf{z}_3 \arrow[r] \arrow[d] & \cdots \\
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 &
\end{tikzcd}
$$

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

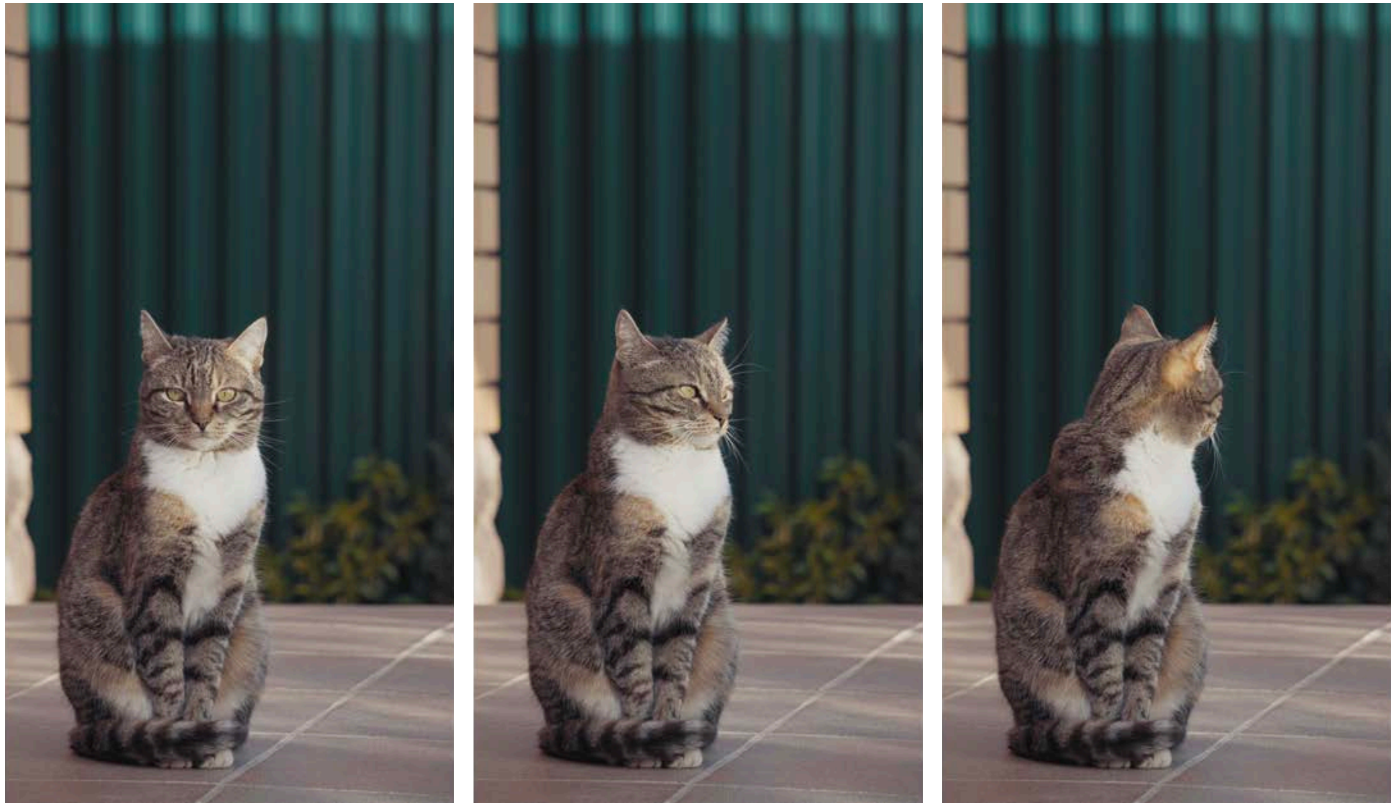


Predictions

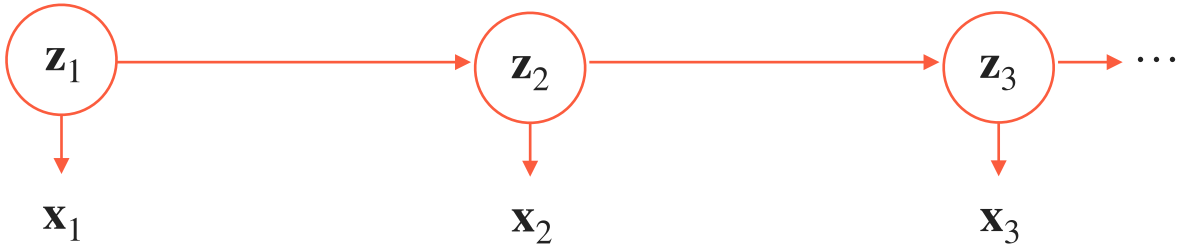
$$\begin{tikzcd} \mathbf{z}_1 \arrow[r] \arrow[d] & \mathbf{z}_2 \arrow[r] \arrow[d] & \mathbf{z}_3 \arrow[r] \arrow[d] & \cdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \end{tikzcd}$$

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$



Predictions



$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

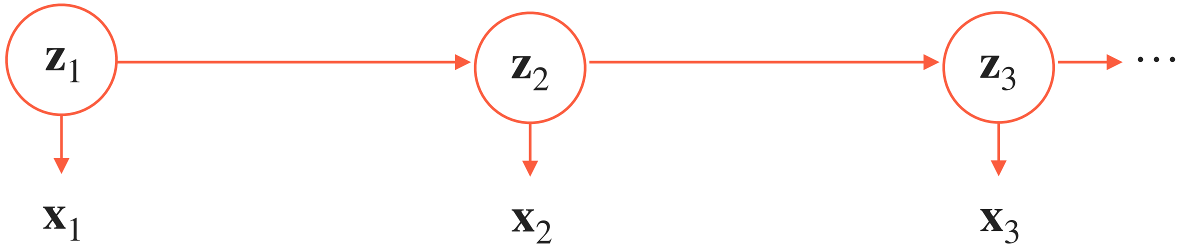
$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$

Predictive pdf:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p\left(\mathbf{z}_t, \mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1}$$

where $\mathbf{z}_{1:t} = [\mathbf{z}_1, \ldots, \mathbf{z}_t]$ and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

Predictions



$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

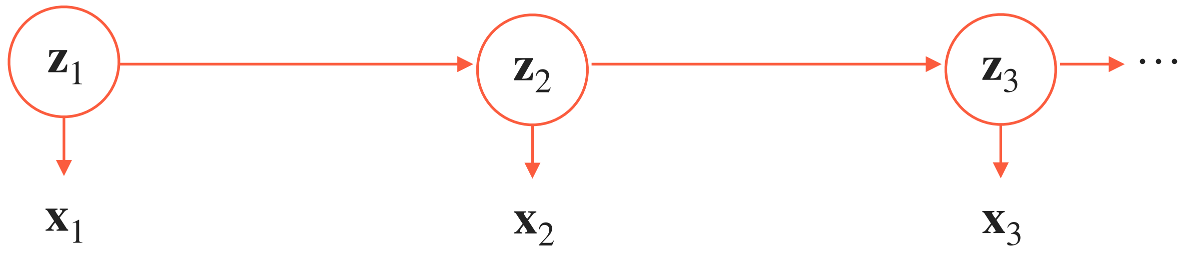
$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$

Predictive pdf:

$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p\left(\mathbf{z}_t, \mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

where $\mathbf{z}_{1:t} = [\mathbf{z}_1, \ldots, \mathbf{z}_t]$ and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

Predictions



$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

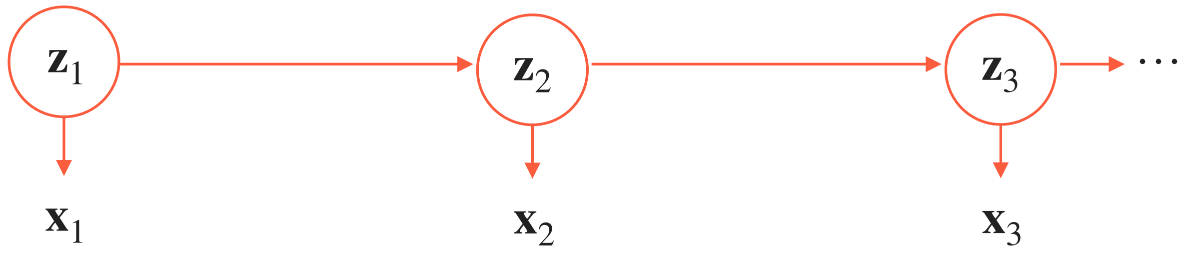
$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$

Predictive pdf:

$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p\left(\mathbf{z}_t, \mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} = \int_{\mathcal{Z}} \underbrace{p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})}_{\text{Posterior pdf at } t-1} p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

where $\mathbf{z}_{1:t} = [\mathbf{z}_1, \ldots, \mathbf{z}_t]$ and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

Predictions



$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

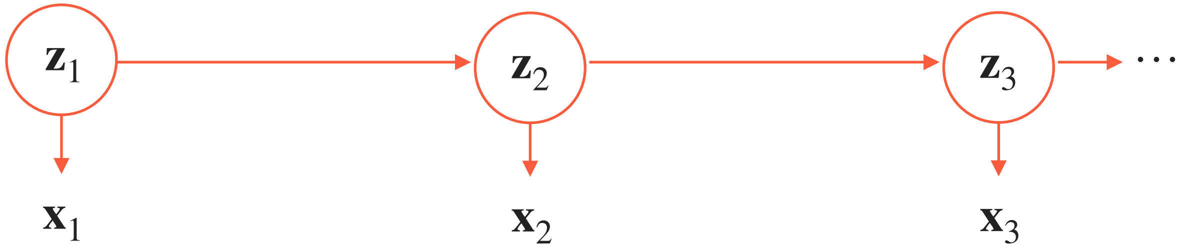
$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$

Predictive pdf:

$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p\left(\mathbf{z}_t, \mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} = \int_{\mathcal{Z}} \underbrace{p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})}_{\text{Posterior pdf at } t-1} \underbrace{p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})}_{\text{Transition pdf}} d\mathbf{z}_{t-1}$$

where $\mathbf{z}_{1:t} = [\mathbf{z}_1, \ldots, \mathbf{z}_t]$ and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

Predictions



$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}, \mathbf{v}_t\right)$$

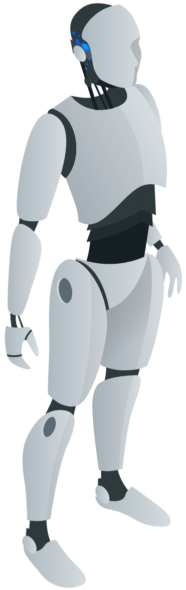
$$\mathbf{x}_t = h\left(\mathbf{z}_t, \mathbf{w}_t\right)$$

Predictive pdf:

$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p\left(\mathbf{z}_t, \mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} = \int_{\mathcal{Z}} \underbrace{p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})}_{\text{Posterior pdf at } t-1} \underbrace{p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})}_{\text{Transition pdf}} d\mathbf{z}_{t-1}$$

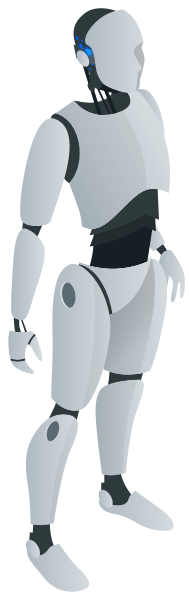
where $\mathbf{z}_{1:t} = [\mathbf{z}_1, \ldots, \mathbf{z}_t]$ and $\mathbf{x}_{1:t} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$

Sequential Inference Illustrated



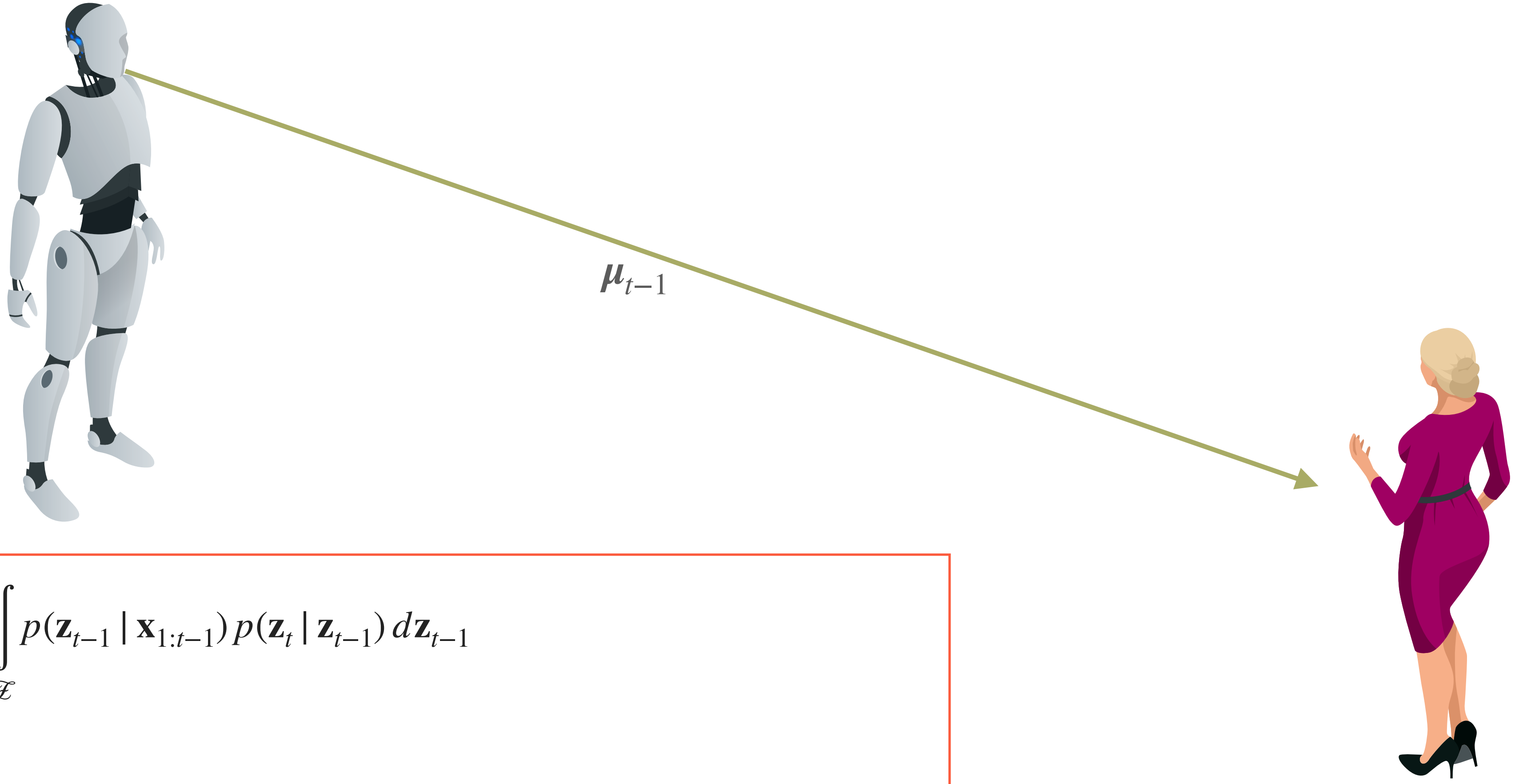
Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Sequential Inference Illustrated



Predictive pdf: $$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

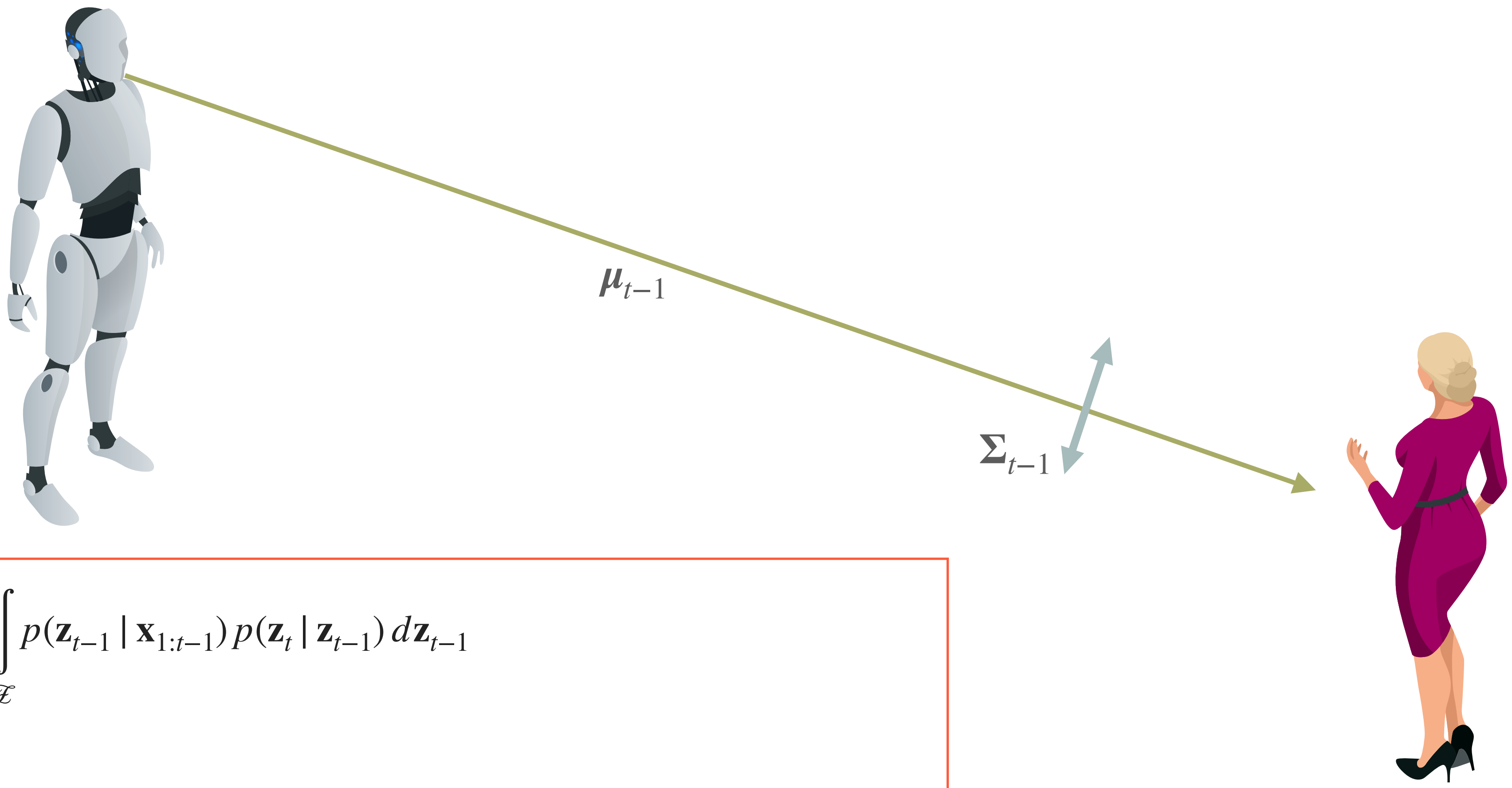
Sequential Inference Illustrated



Predictive pdf:
$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

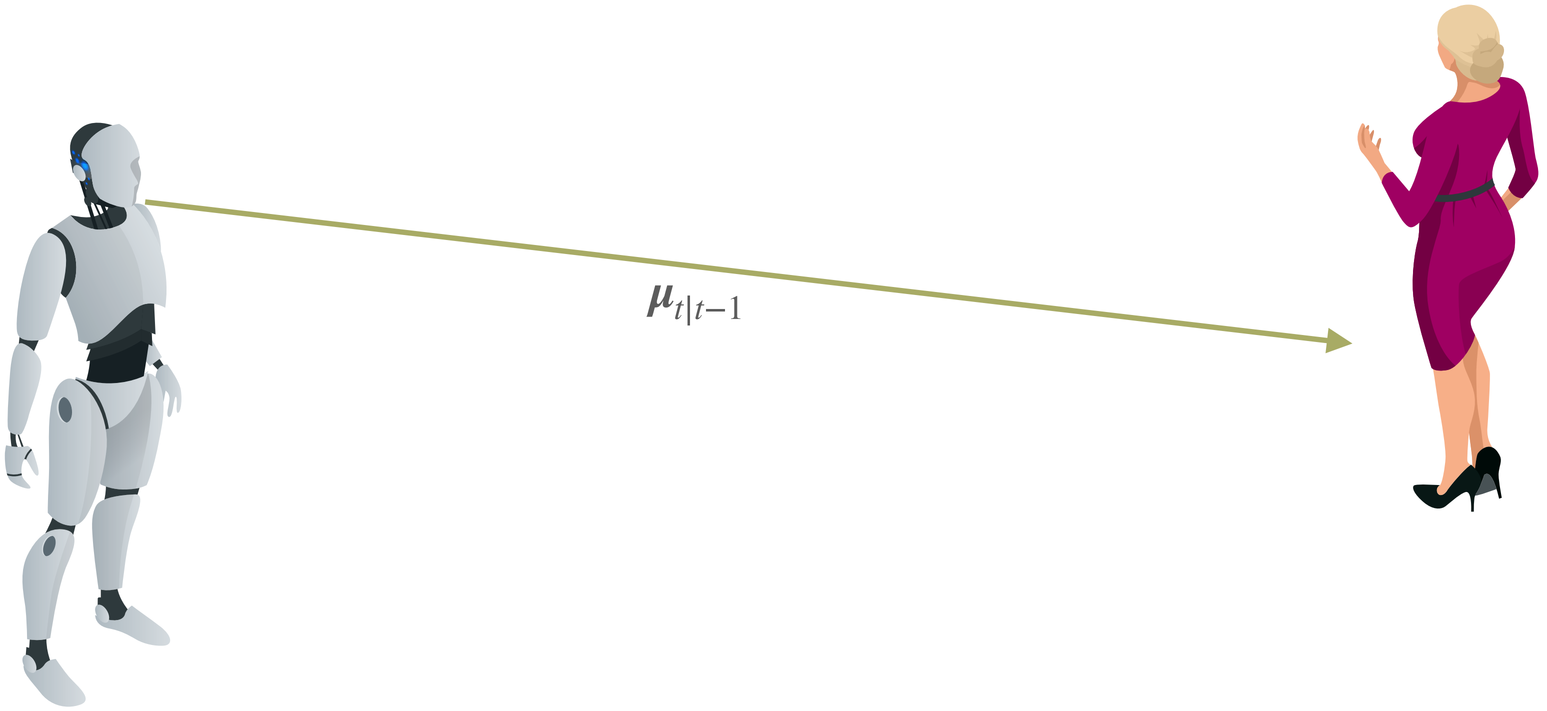
Sequential Inference Illustrated

UNIVERSITY OF Southampton



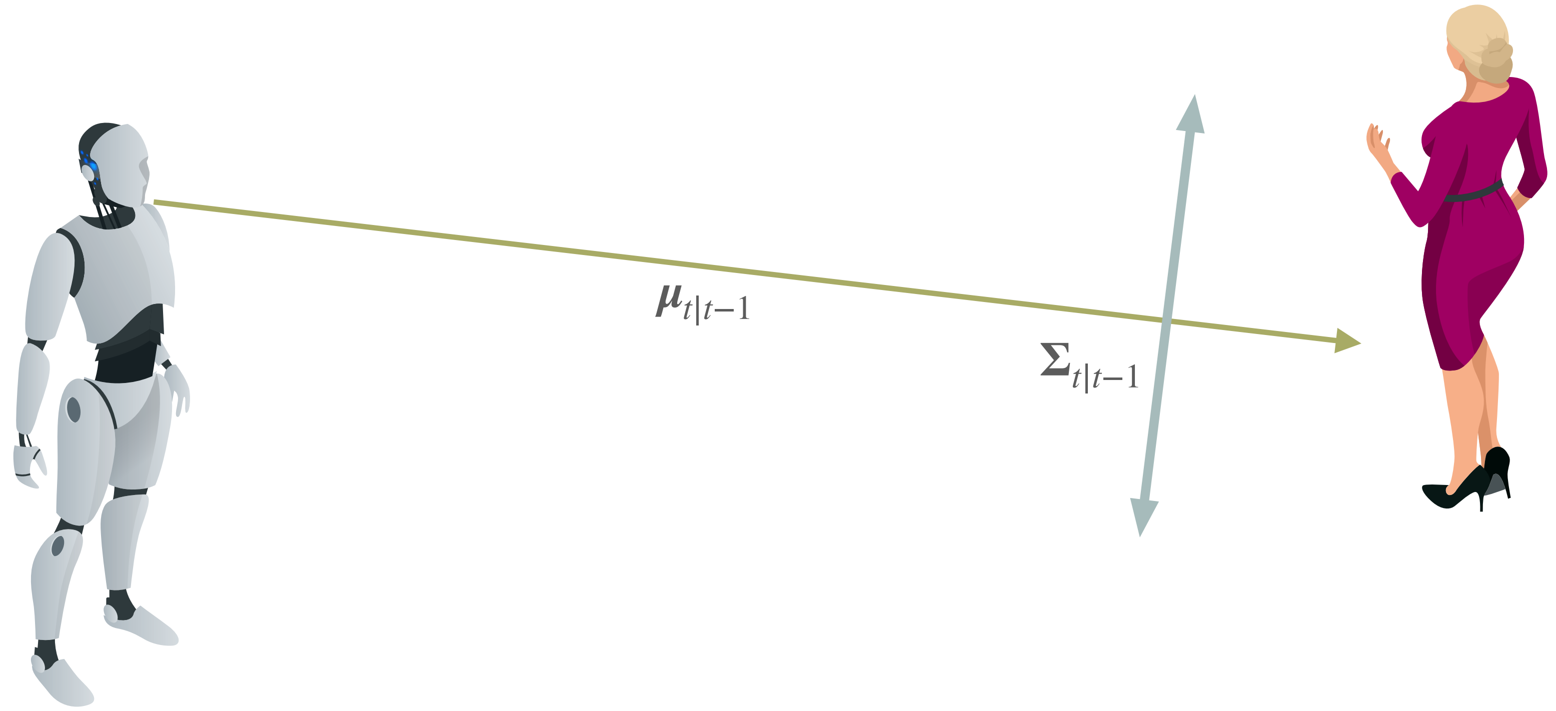
Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Sequential Inference Illustrated



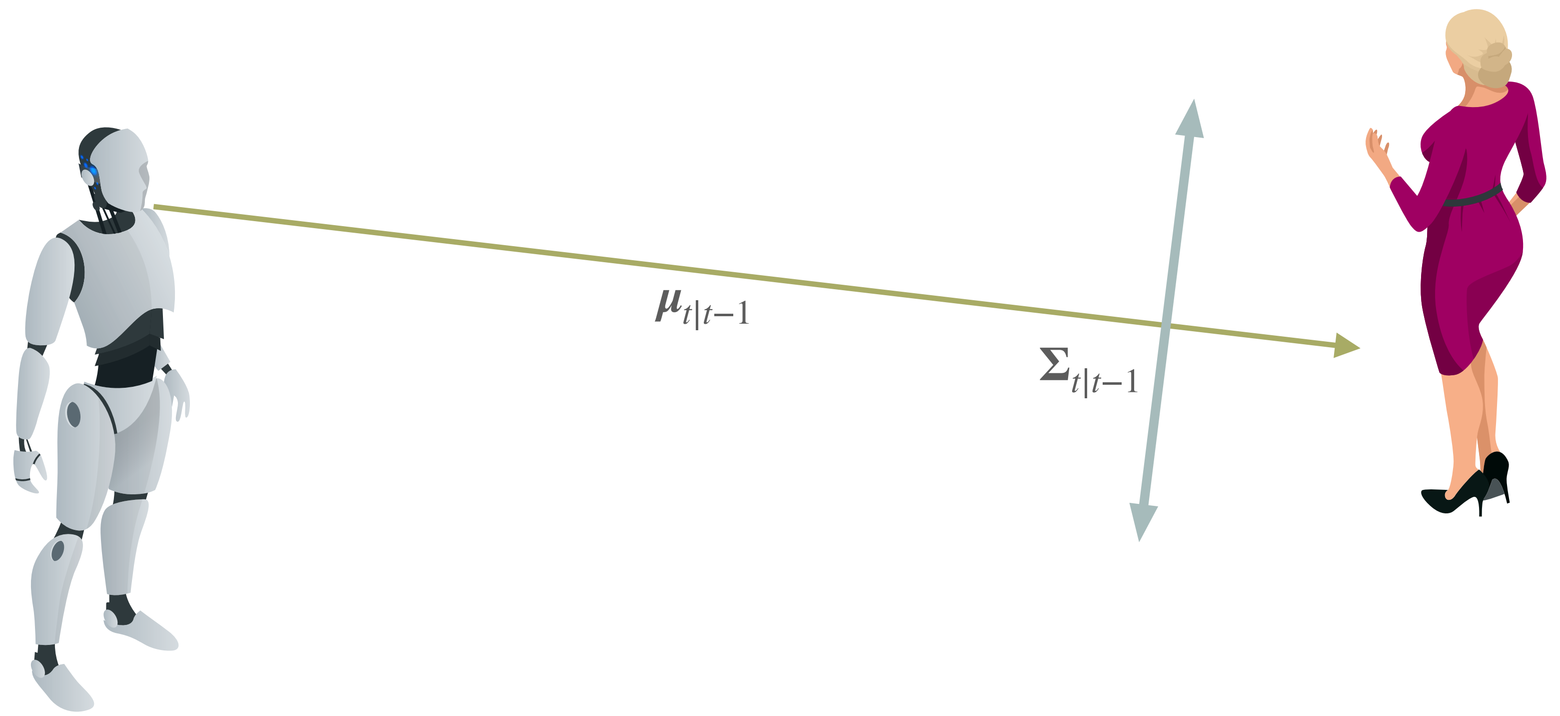
Predictive pdf: $p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$

Sequential Inference Illustrated



Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

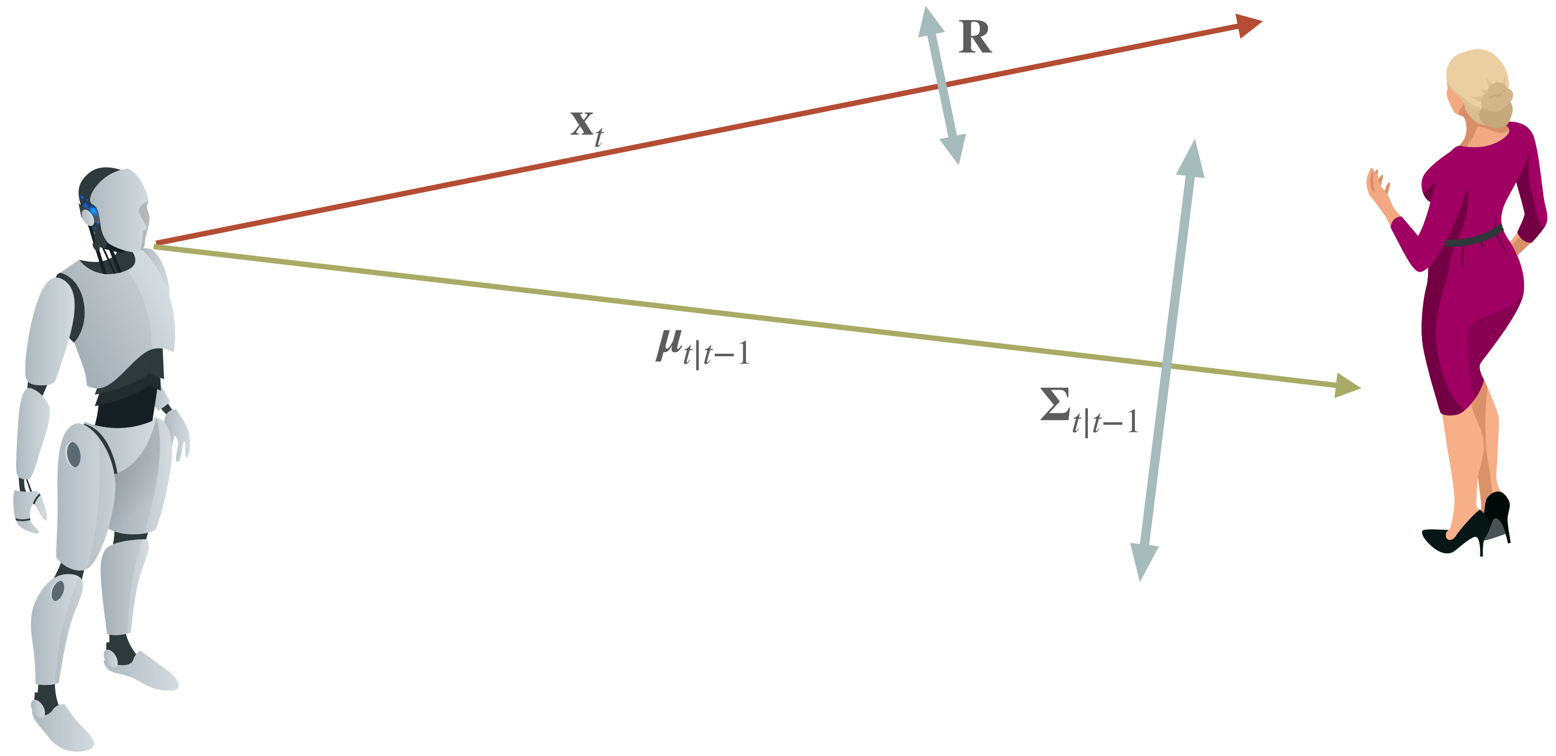
Sequential Inference Illustrated



Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t}$$

Sequential Inference Illustrated

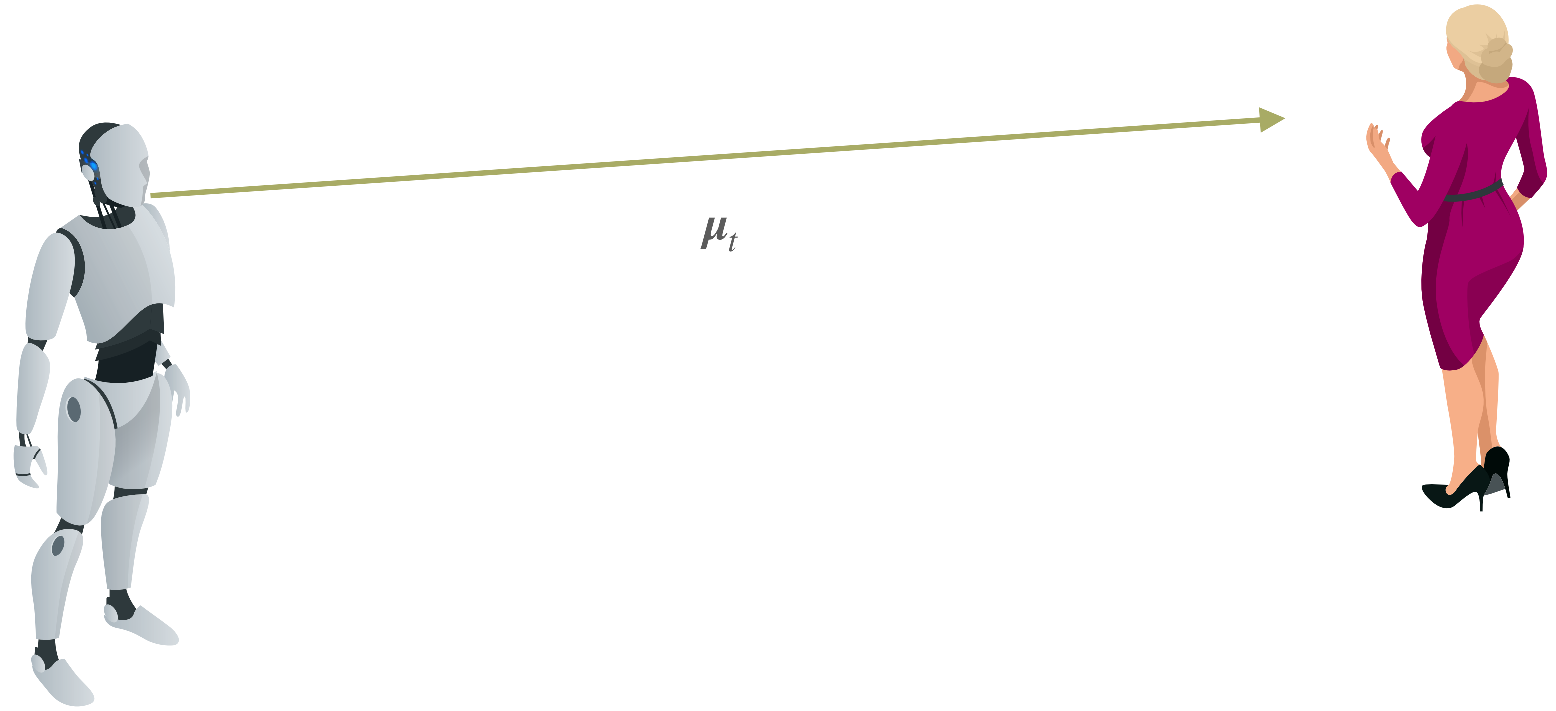


Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t}$$

Sequential Inference Illustrated

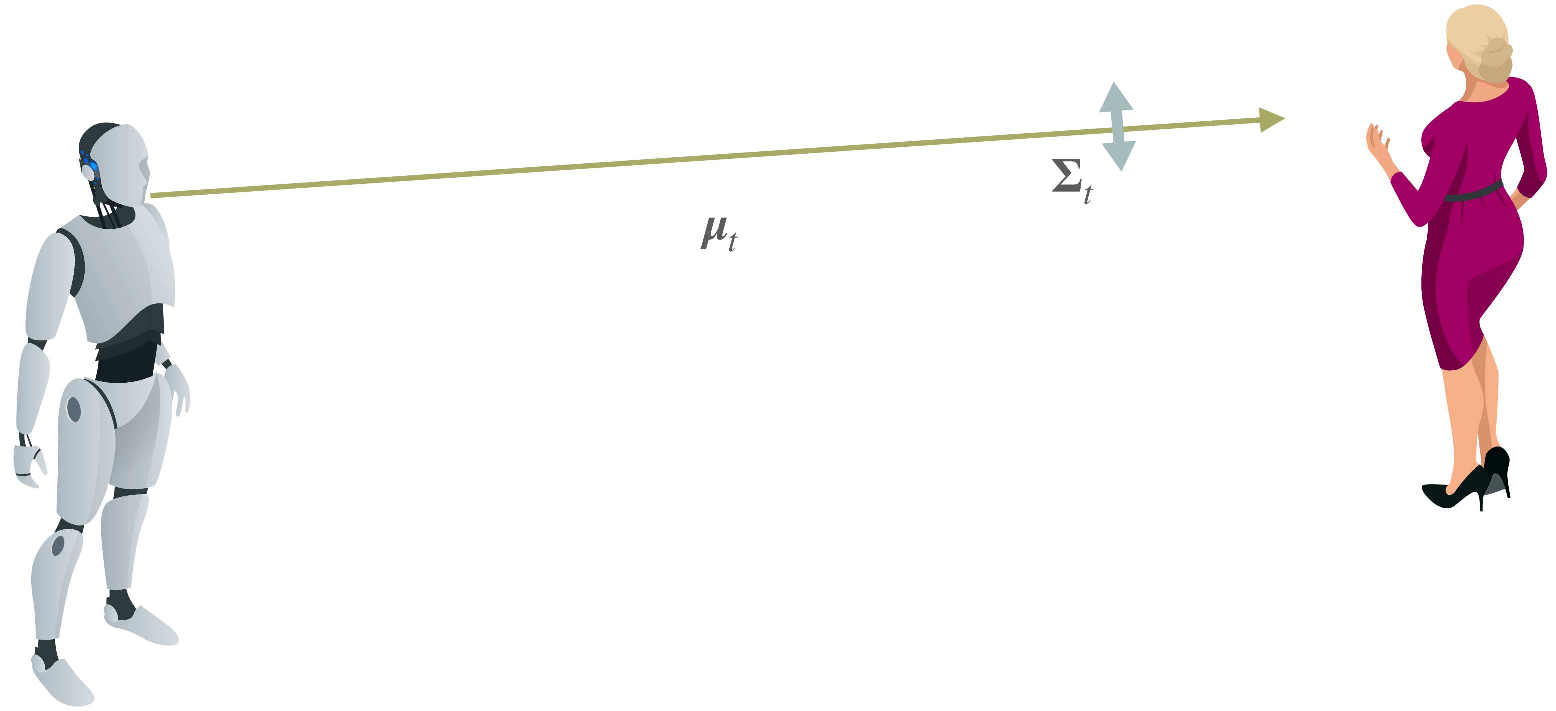
University of Southampton



Predictive pdf: $$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density: $$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \,|\, \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \,|\, \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t}$$

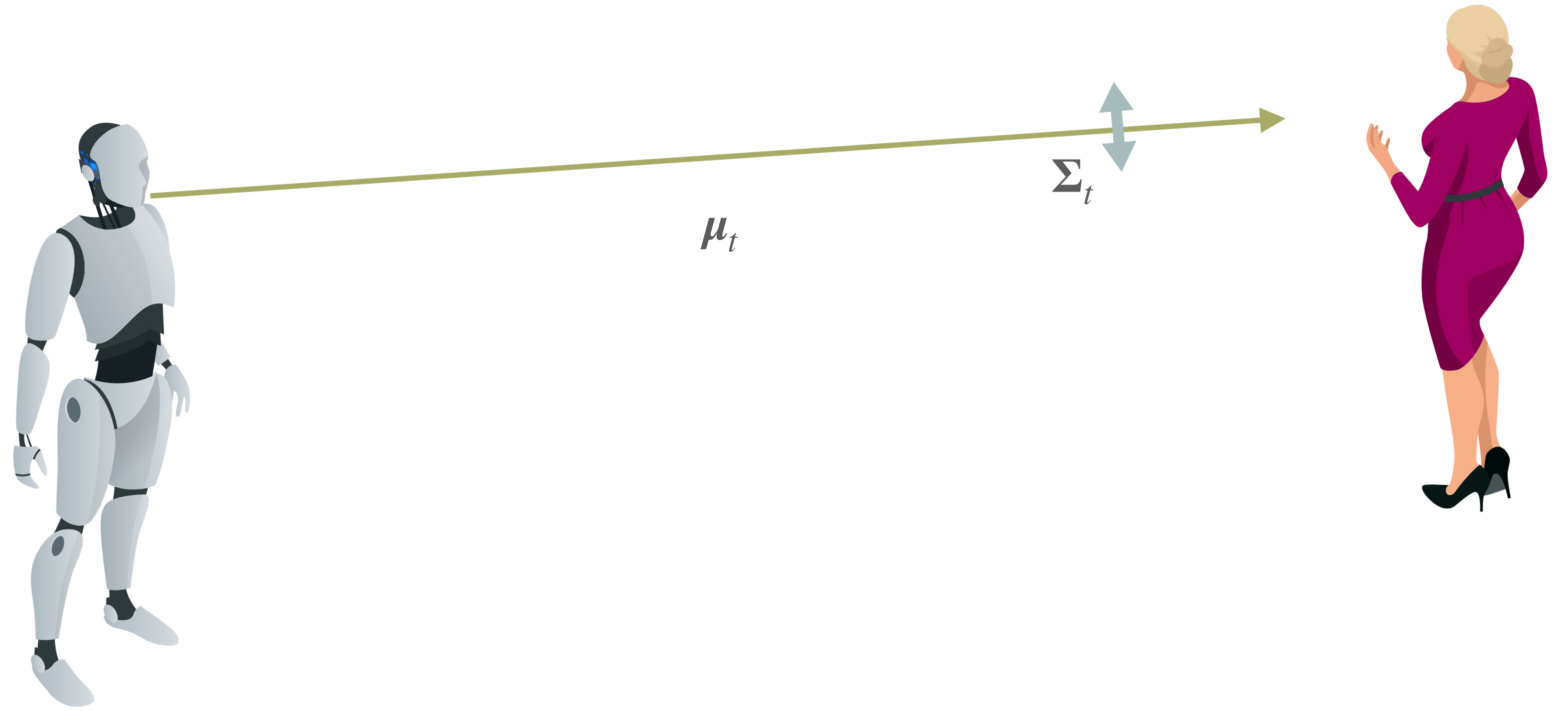
Sequential Inference Illustrated



Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t}$$

Sequential Inference Illustrated



Predictive pdf:
$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \,|\, \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$$

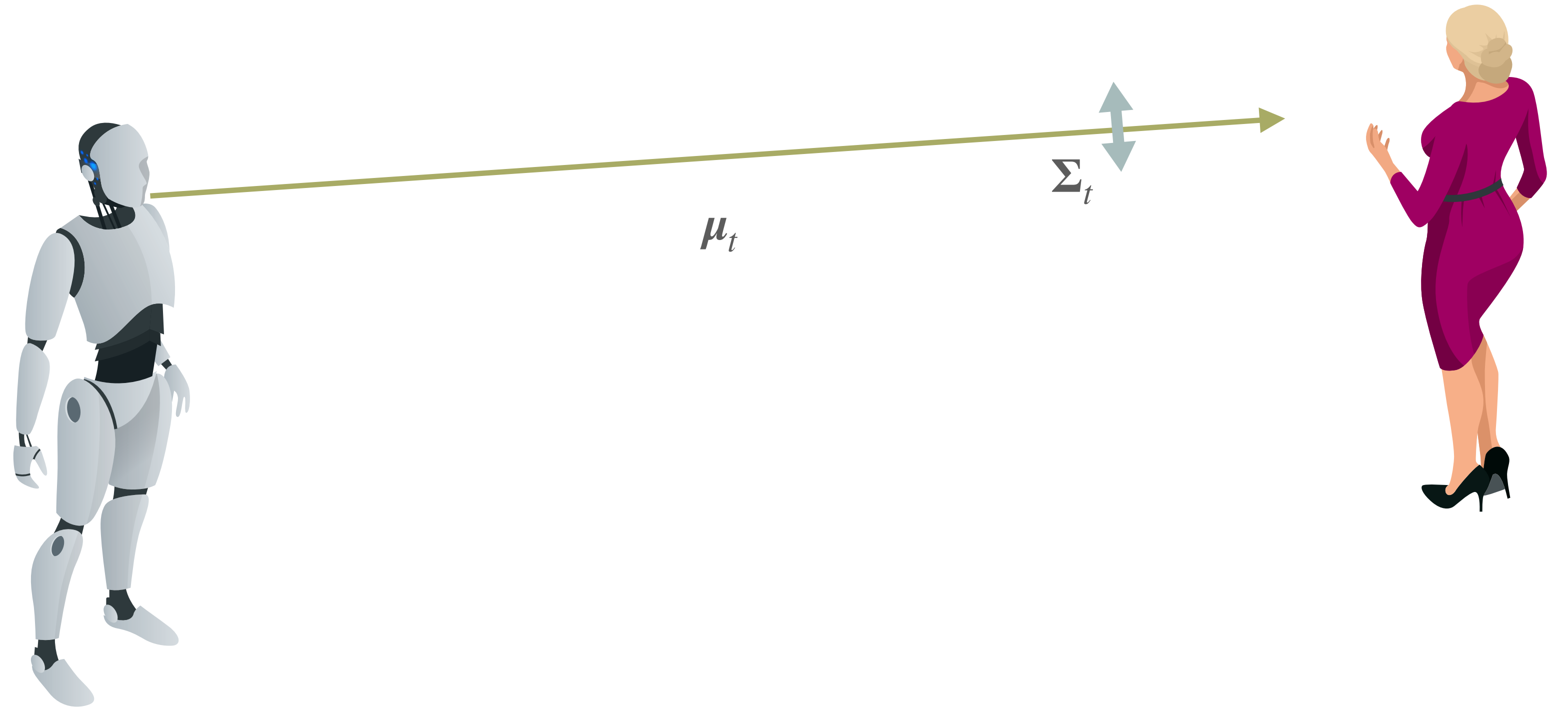
Filtering density:
$$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \,|\, \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \,|\, \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$$

Sequential Inference Illustrated

Kalman prediction:

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q}_t + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$$



Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$$

Filtering density: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$$

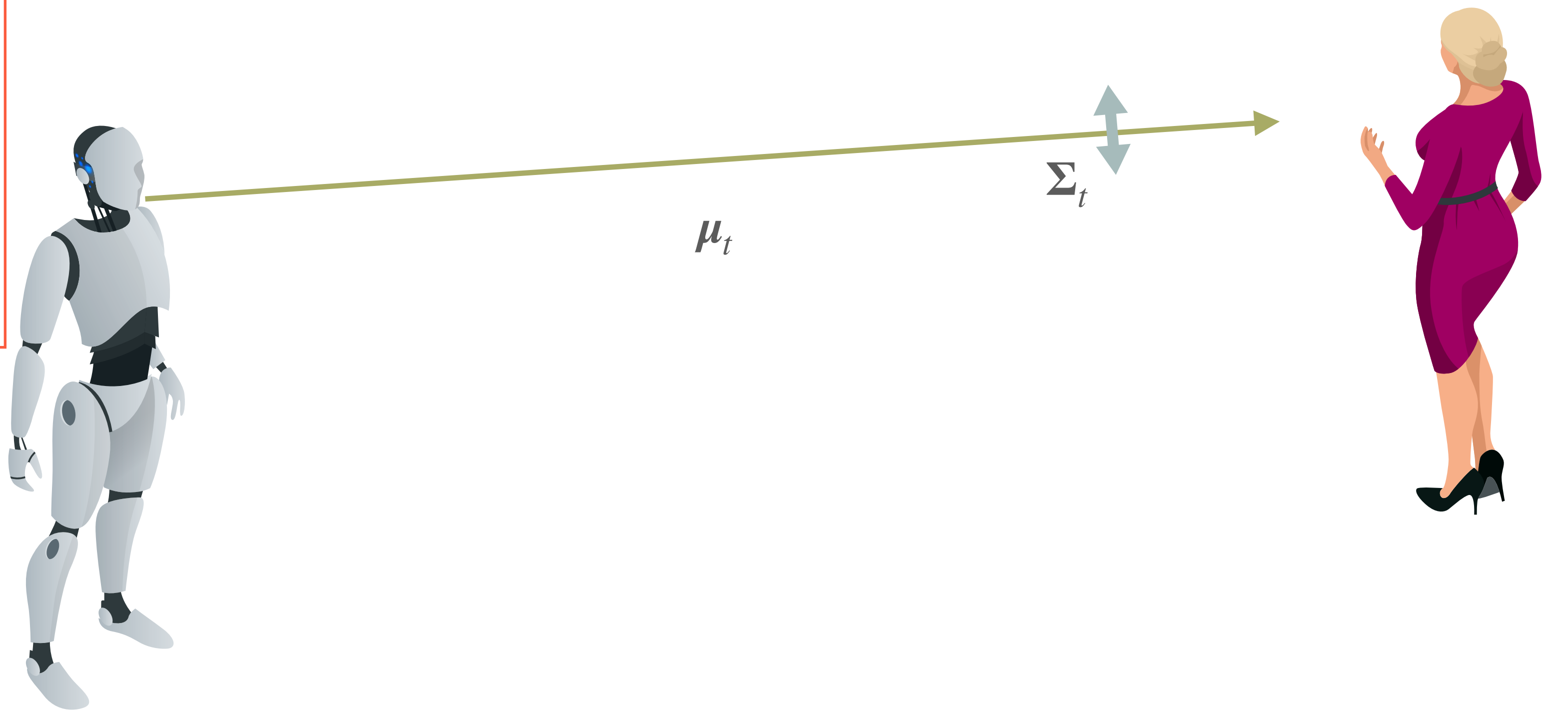
Sequential Inference Illustrated

Kalman update:

$$\mathbf{K}_t = \mathbf{\Sigma}_{t|t-1}\mathbf{H}^T\left(\mathbf{H}\,\mathbf{\Sigma}_{t|t-1}\,\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\left(\mathbf{x}_t - \mathbf{H}\,\boldsymbol{\mu}_{t|t-1}\right)$$

$$\mathbf{\Sigma}_t = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t\,\mathbf{H}\,\mathbf{\Sigma}_{t|t-1}$$



Predictive pdf: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \mathbf{\Sigma}_{t|t-1}\right)$$

Filtering density: $$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)\, d\mathbf{z}_t} \overset{\text{LGSS}}{=} \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \mathbf{\Sigma}_t\right)$$

Sequential Bayesian Framework

Bayesian paradigm provides a principled framework:

1. Modelling assumptions:
   A. Distribution of transition pdf, $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$
   B. Distribution of likelihood function, $p(\mathbf{x}_t \mid \mathbf{z}_t)$
   C. Distribution of previous posterior pdf, $p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})$

Sequential Bayesian Framework

Predictive pdf:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Bayesian paradigm provides a principled framework:

1. Modelling assumptions:
    A. Distribution of transition pdf, $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$
    B. Distribution of likelihood function, $p(\mathbf{x}_t \mid \mathbf{z}_t)$
    C. Distribution of previous posterior pdf, $p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})$
2. Derive predictive pdf, $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$

Sequential Bayesian Framework

Predictive pdf:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}$$

Bayesian paradigm provides a principled framework:

1. Modelling assumptions:
   A. Distribution of transition pdf, $p(\mathbf{z}_t \mid \mathbf{z}_{t-1})$
   B. Distribution of likelihood function, $p(\mathbf{x}_t \mid \mathbf{z}_t)$
   C. Distribution of previous posterior pdf, $p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})$
2. Derive predictive pdf, $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1})$
3. Derive posterior pdf, $p(\mathbf{z}_t \mid \mathbf{x}_{1:t})$

Sequential Bayesian Framework

Predictive pdf:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int_{\mathcal{Z}} p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, d\mathbf{z}_{t-1}$$

Filtering density:

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) = \frac{p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}{\int_{\mathcal{Z}} p\left(\mathbf{x}_t \mid \mathbf{z}_t, \mathbf{x}_{1:t-1}\right)\, p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right)}$$

Integrals are typically intractable for non-linear and/or non-Gaussian state spaces!

- Local linearisation (today)
- Monte Carlo methods (Week 5/6):
  - Numerical approximation by random sampling

Bayesian Inference in the Wild...

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \,|\, f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \,|\, h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

$$p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) \, p\left(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) \, d\mathbf{z}_{t-1}$$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

$$p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \Rightarrow p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \,|\, f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \Rightarrow p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \,|\, h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\displaystyle\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \,|\, f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \,|\, h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1}\mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\displaystyle\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\,\mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} \mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $$\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \Rightarrow p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \,|\, f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \Rightarrow p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \,|\, h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

$$p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T\mathbf{Q}^{-1}f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T\boldsymbol{\Sigma}_{t-1}^{-1}\mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T\boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T\boldsymbol{\Sigma}_{t-1}^{-1}\boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \quad \Rightarrow \quad p(\mathbf{z}_t \mid \mathbf{z}_{t-1}) = \mathcal{N}\left(\mathbf{z}_t \mid f(\mathbf{z}_{t-1}), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h(\mathbf{z}_t) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad \Rightarrow \quad p(\mathbf{x}_t \mid \mathbf{z}_t) = \mathcal{N}\left(\mathbf{x}_t \mid h(\mathbf{z}_t), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) = \int p(\mathbf{z}_t \mid \mathbf{z}_{t-1})\, p(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1})\, d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \quad \Rightarrow \quad p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \quad \Rightarrow \quad p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\displaystyle \int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \Rightarrow p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \mid f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \Rightarrow p\left(\mathbf{x}_t \mid \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \mid h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \to \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \to \mathbb{R}^M$

$$p\left(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \mid \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1}\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1}\mathbf{z}_t + f(\mathbf{z}_{t-1})^T\mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1}\mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $$\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$$

Local Linearisation

Non-Linear State Space:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right) \Rightarrow p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) = \mathcal{N}\left(\mathbf{z}_t \,|\, f\left(\mathbf{z}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}\right) \Rightarrow p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) = \mathcal{N}\left(\mathbf{x}_t \,|\, h\left(\mathbf{z}_t\right), \mathbf{R}\right)$$

Transition function: $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$; Observation function: $h : \mathbb{R}^N \rightarrow \mathbb{R}^M$

$$p\left(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}\right) = \int p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) p\left(\mathbf{z}_{t-1} \,|\, \mathbf{x}_{1:t-1}\right) d\mathbf{z}_{t-1} \propto \int \exp\left\{-\frac{1}{2}\left[\left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right)^T \mathbf{Q}^{-1} \left(\mathbf{z}_t - f(\mathbf{z}_{t-1})\right) + \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right)^T \boldsymbol{\Sigma}_{t-1}^{-1} \left(\mathbf{z}_t - \boldsymbol{\mu}_{t-1}\right)\right]\right\} d\mathbf{z}_{t-1}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mathbf{z}_t^T \mathbf{Q}^{-1} \mathbf{z}_t - 2f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} \mathbf{z}_t + f(\mathbf{z}_{t-1})^T \mathbf{Q}^{-1} f(\mathbf{z}_{t-1}) + \mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1} + \boldsymbol{\mu}_{t-1}^T \boldsymbol{\Sigma}_{t-1}^{-1} \boldsymbol{\mu}_{t-1}\right]\right\} d\mathbf{z}_{t-1}$$

Wanted: $\int \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t-1}^T \mathbf{P}\, \mathbf{z}_{t-1} - 2\mathbf{z}_{t-1}^T \boldsymbol{\beta} + \gamma\right)\right\} d\mathbf{z}_{t-1}$ $\Rightarrow$ **Problem:** $f(\mathbf{z}_{t-1}), h(\mathbf{z}_t)$

Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth

Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2\,\mathbf{I}\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth

Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2\, \mathbf{I}\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth

Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2\,\mathbf{I}\right)$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth

Example: Bearing-Only Tracking

University of Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth

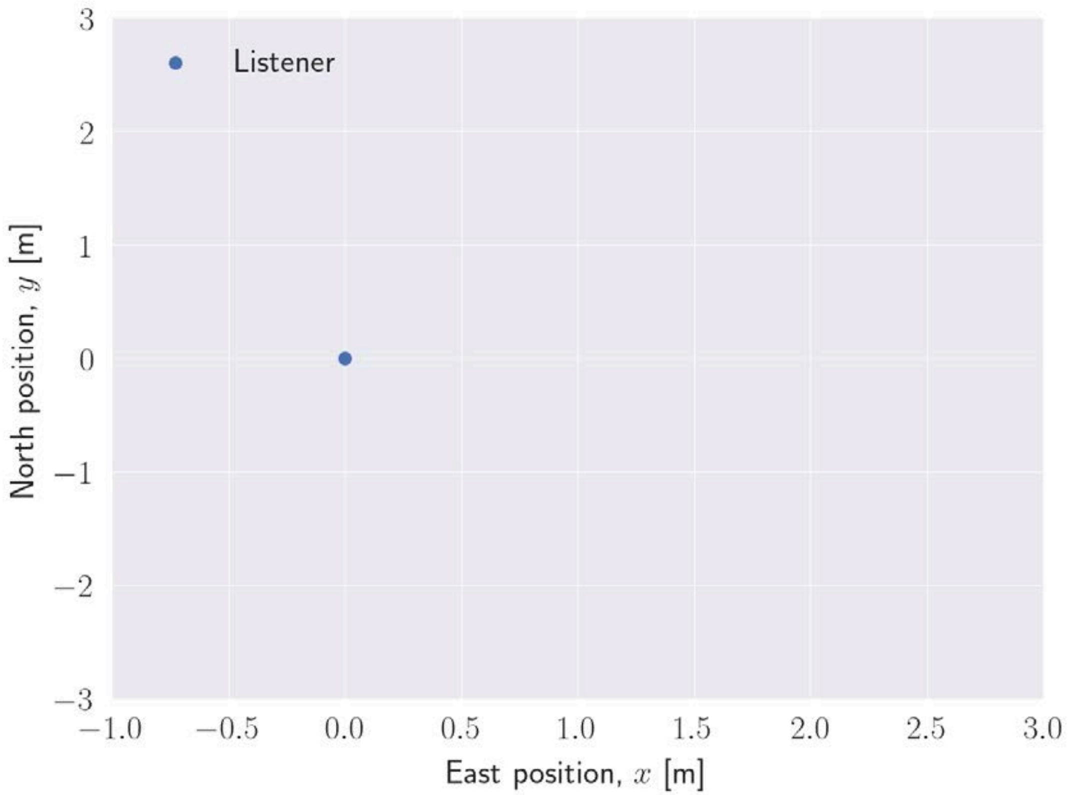
Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Example: Bearing-Only Tracking

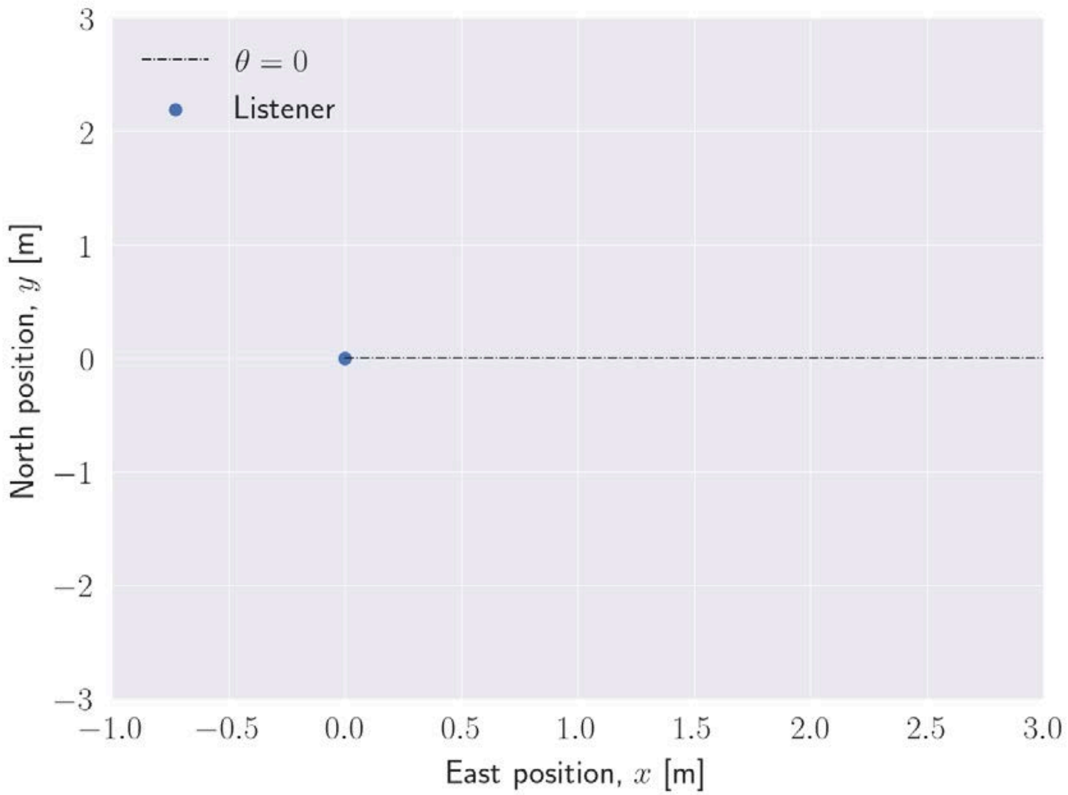
UNIVERSITY OF Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \, \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Example: Bearing-Only Tracking

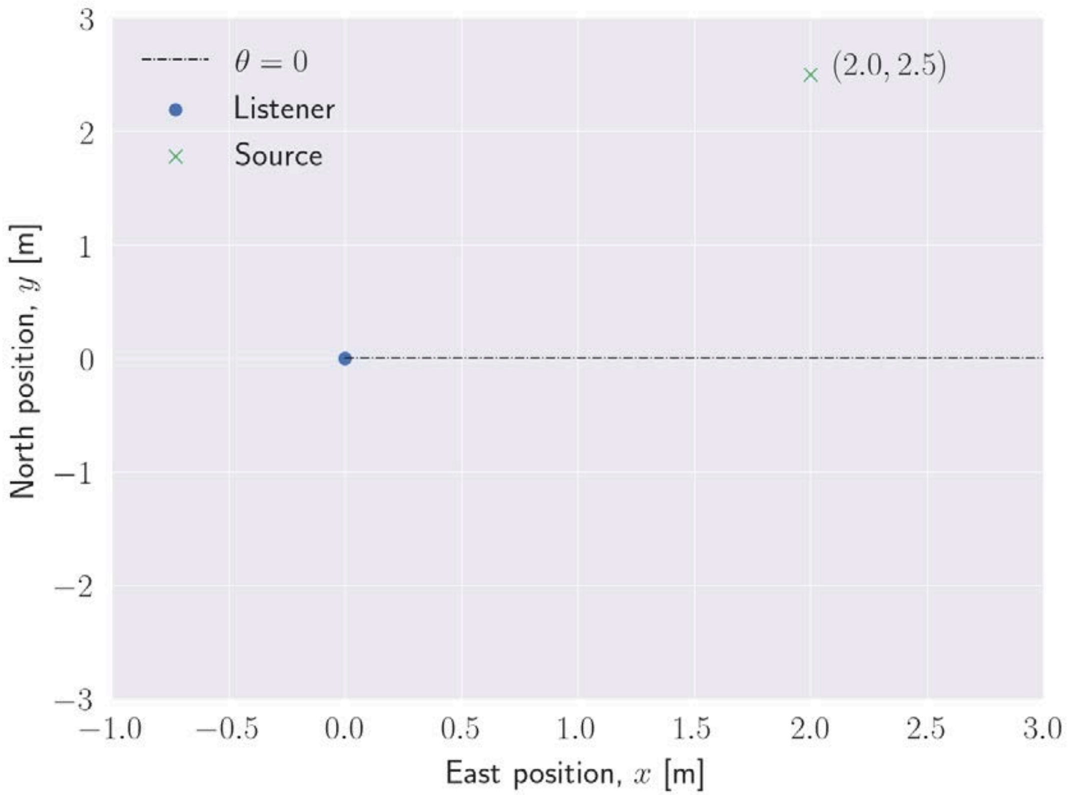
UNIVERSITY OF Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2\,\mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Example: Bearing-Only Tracking

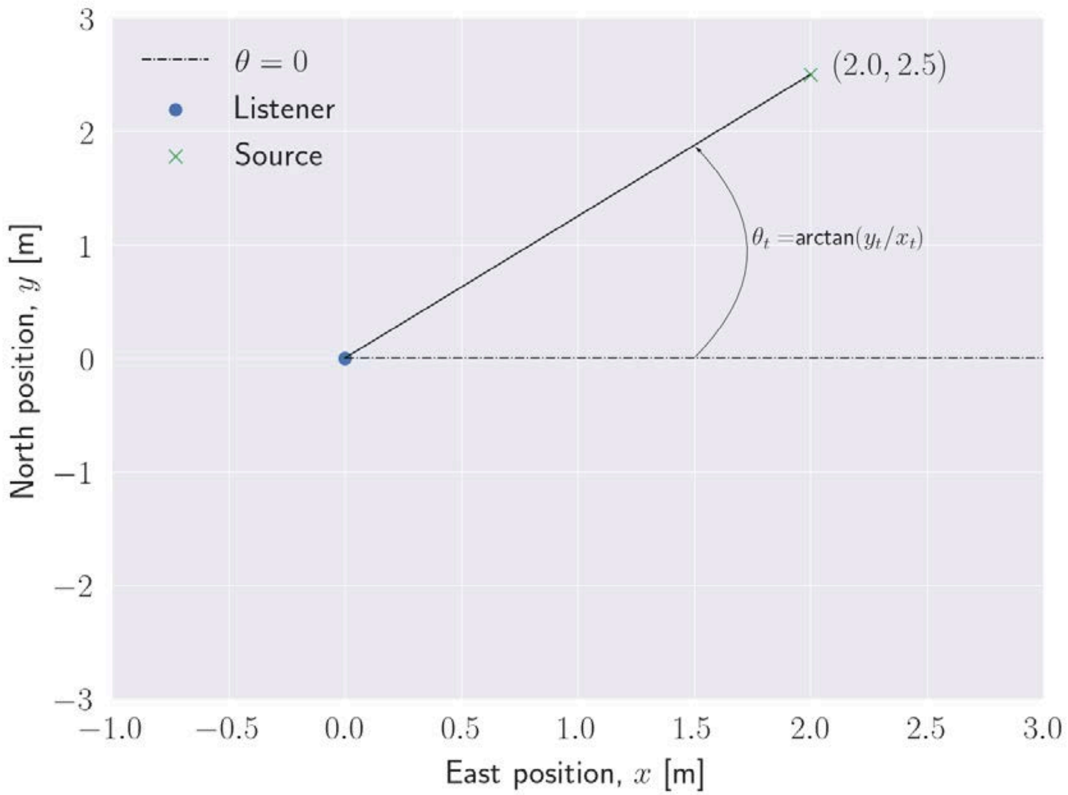
UNIVERSITY OF Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



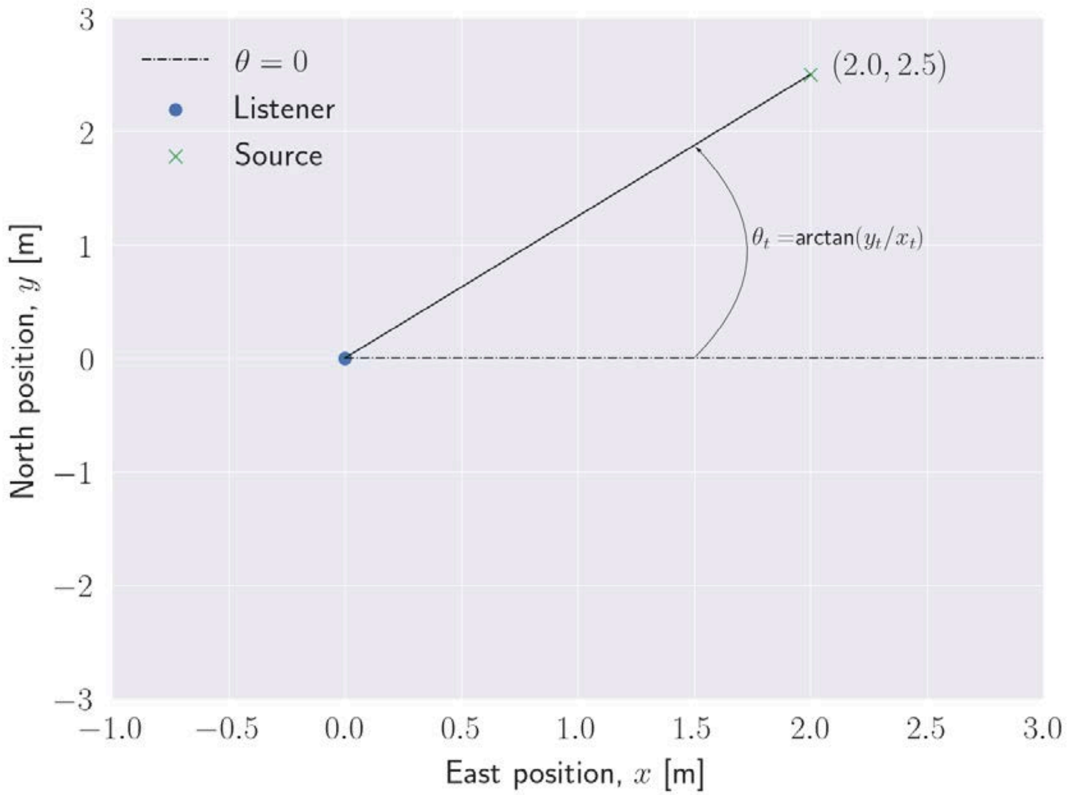
Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Example: Bearing-Only Tracking

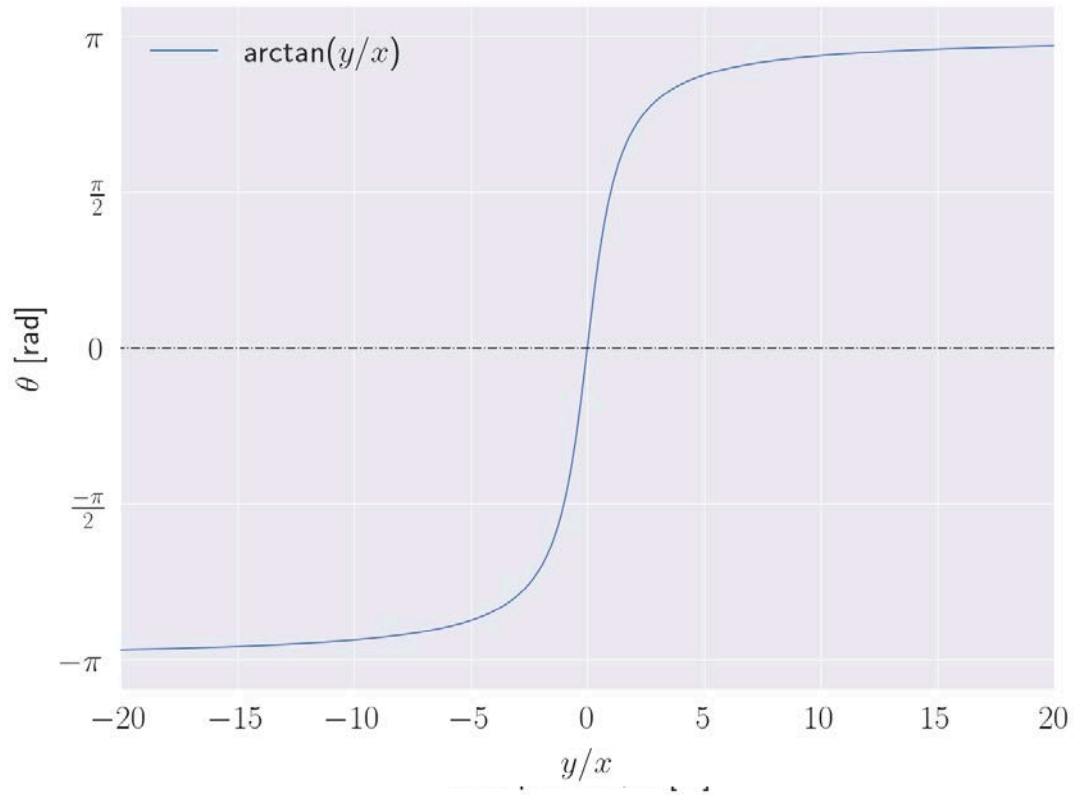
UNIVERSITY OF Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Example: Bearing-Only Tracking

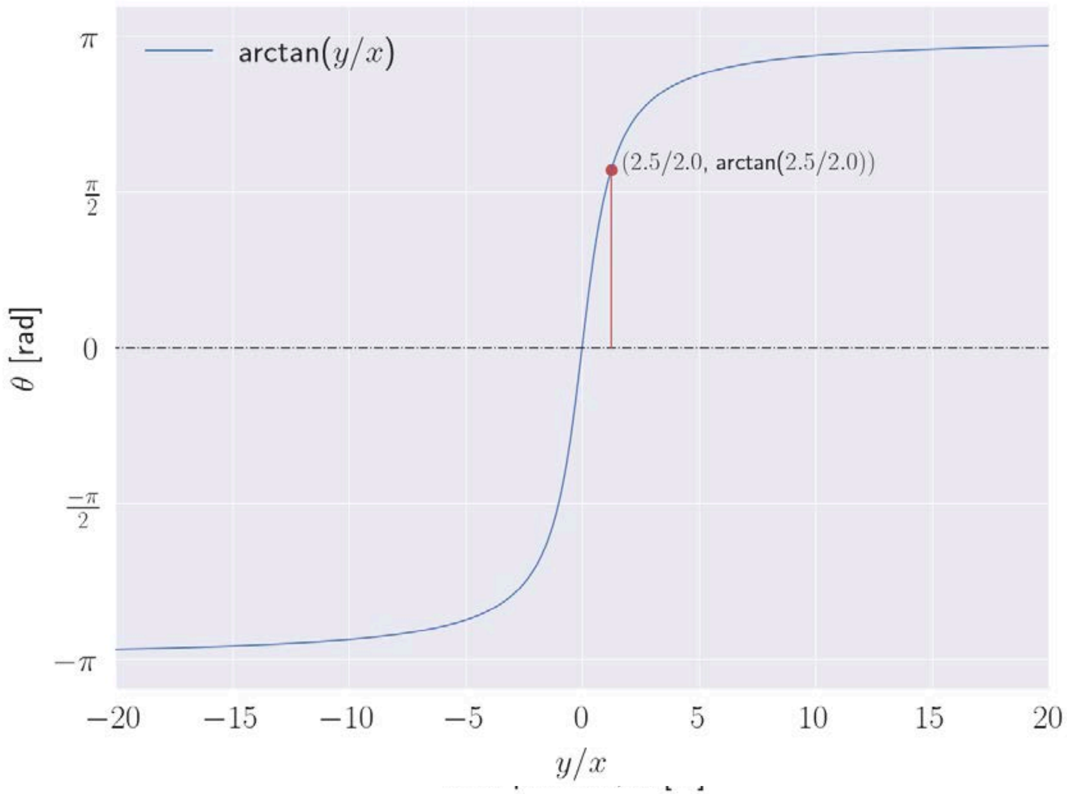
UNIVERSITY OF Southampton

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



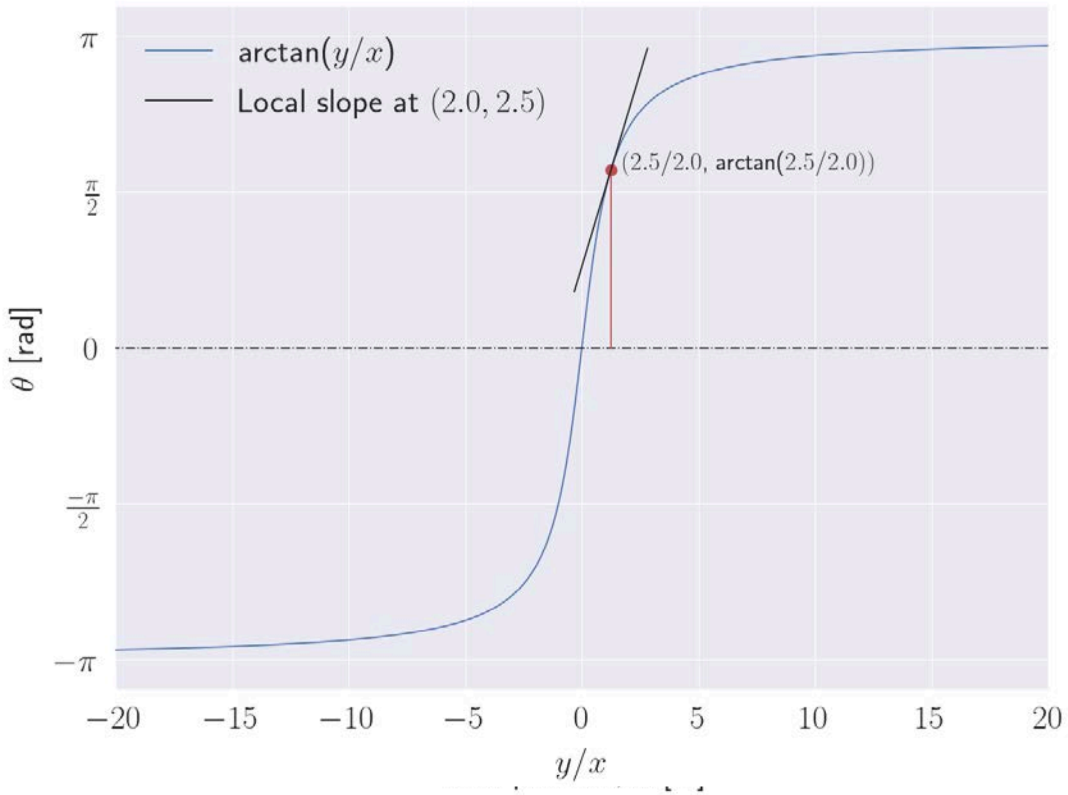
Example: Bearing-Only Tracking

Linear state transition, non-linear observation:

$$\mathbf{z}_t = \begin{bmatrix} x_t \\ y_t \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \dot{x}_{t-1} \\ \dot{y}_{t-1} \end{bmatrix}}_{\mathbf{z}_{t-1}} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}\right)$$

$$\theta_t = \underbrace{\text{atan2}\left(\frac{y_t}{x_t}\right)}_{h(\mathbf{z}_t)} + w_t, \quad w_t \sim \mathcal{N}\left(0, \sigma_\theta^2\right)$$

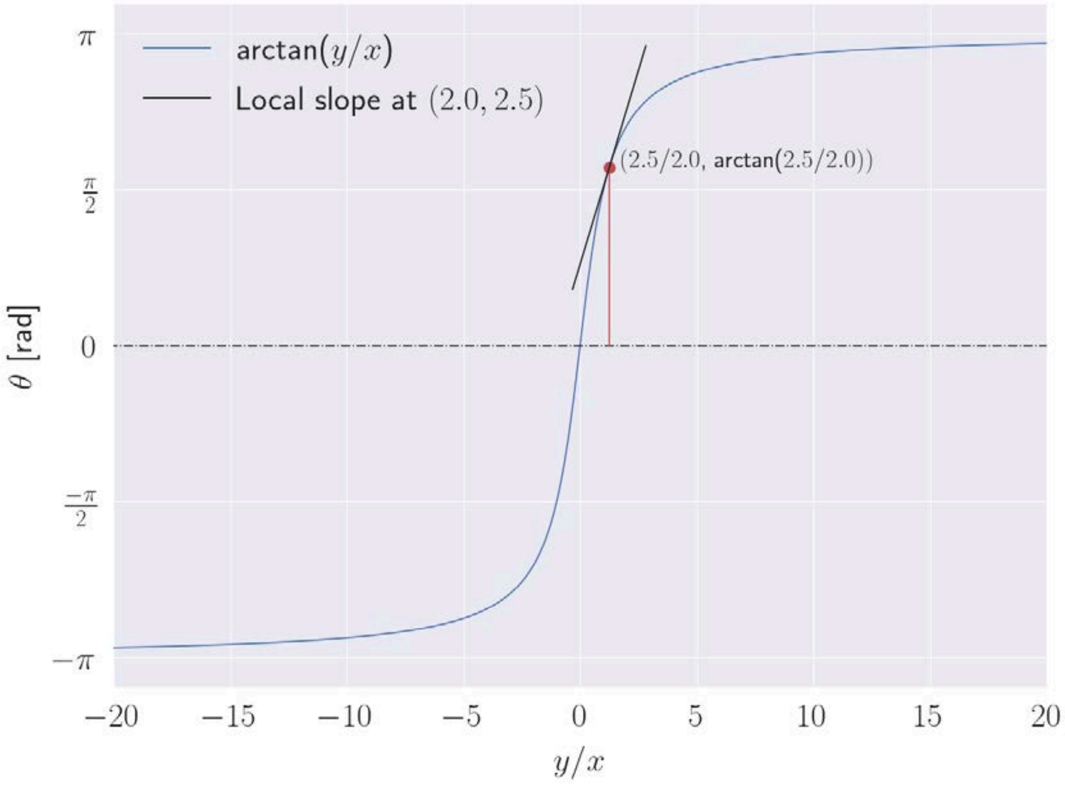
$\mathbf{z}_t$: Target state, ΔT : Sampling time, θ_t : Observed target azimuth



Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\Big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a} - \mathbf{s})$$

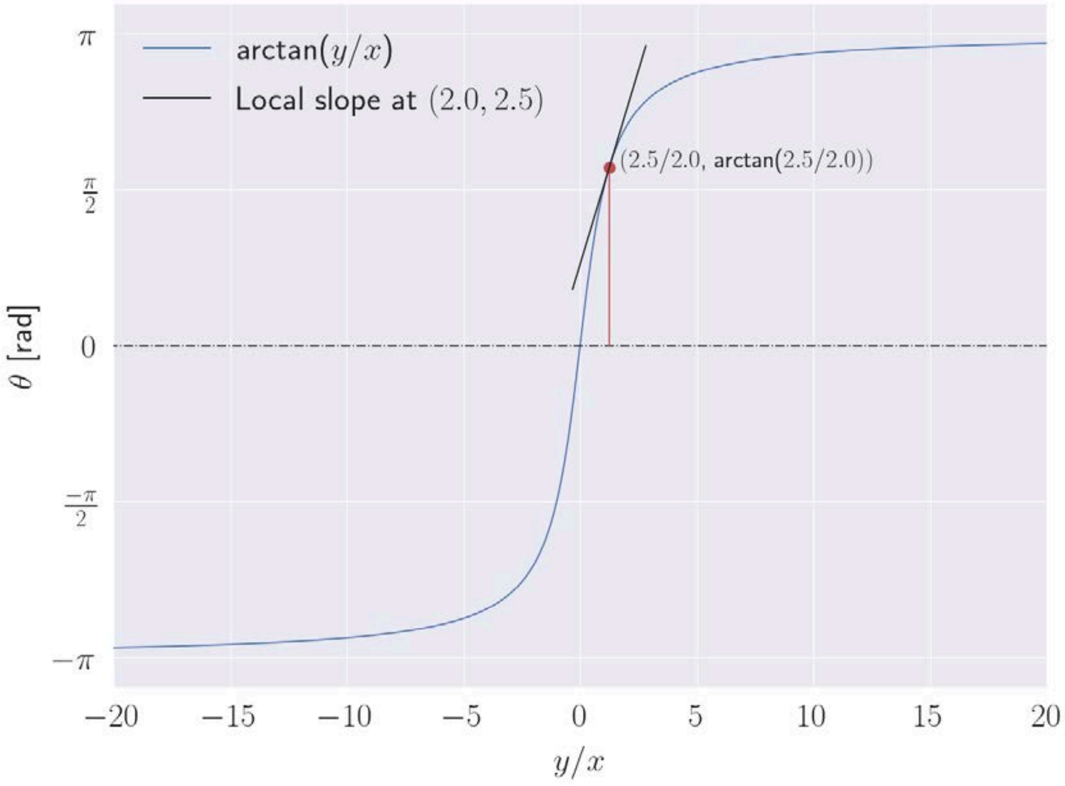


Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a} - \mathbf{s})$$

Jacobian ($M \times N$ matrix): $$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$$

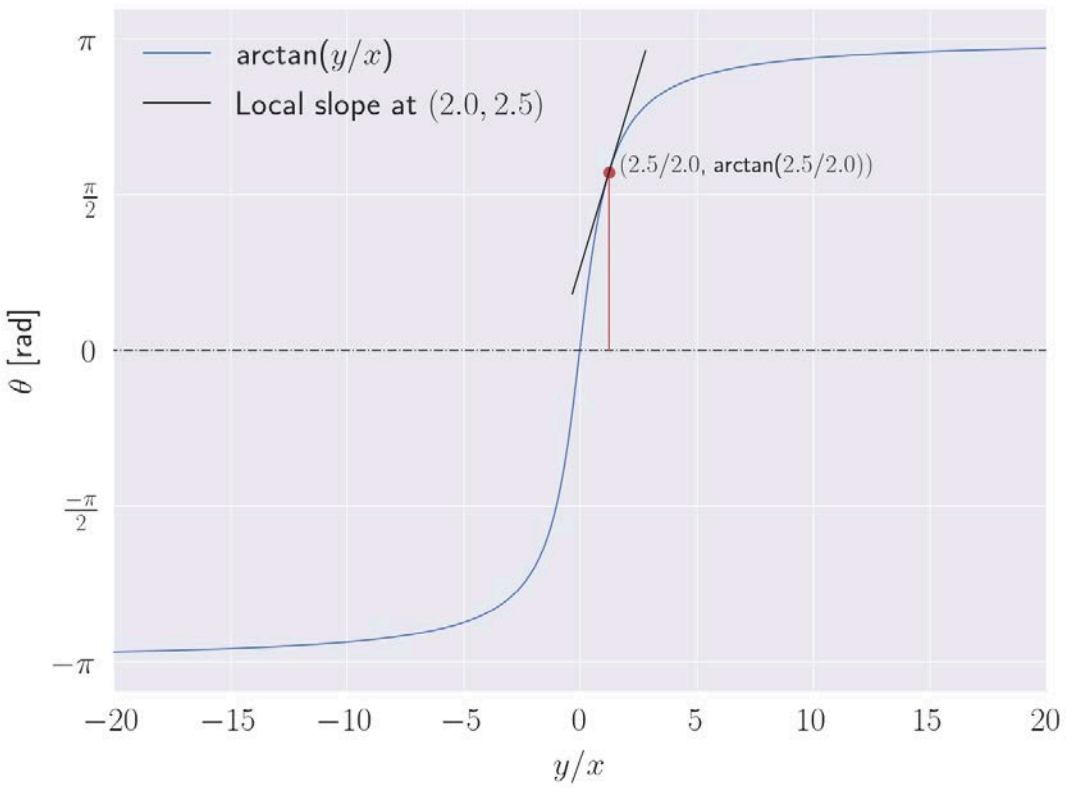


Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \to \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a}-\mathbf{s})$$

Jacobian ($M \times N$ matrix): $\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$



Applied to non-linear, Gaussian state-space model:

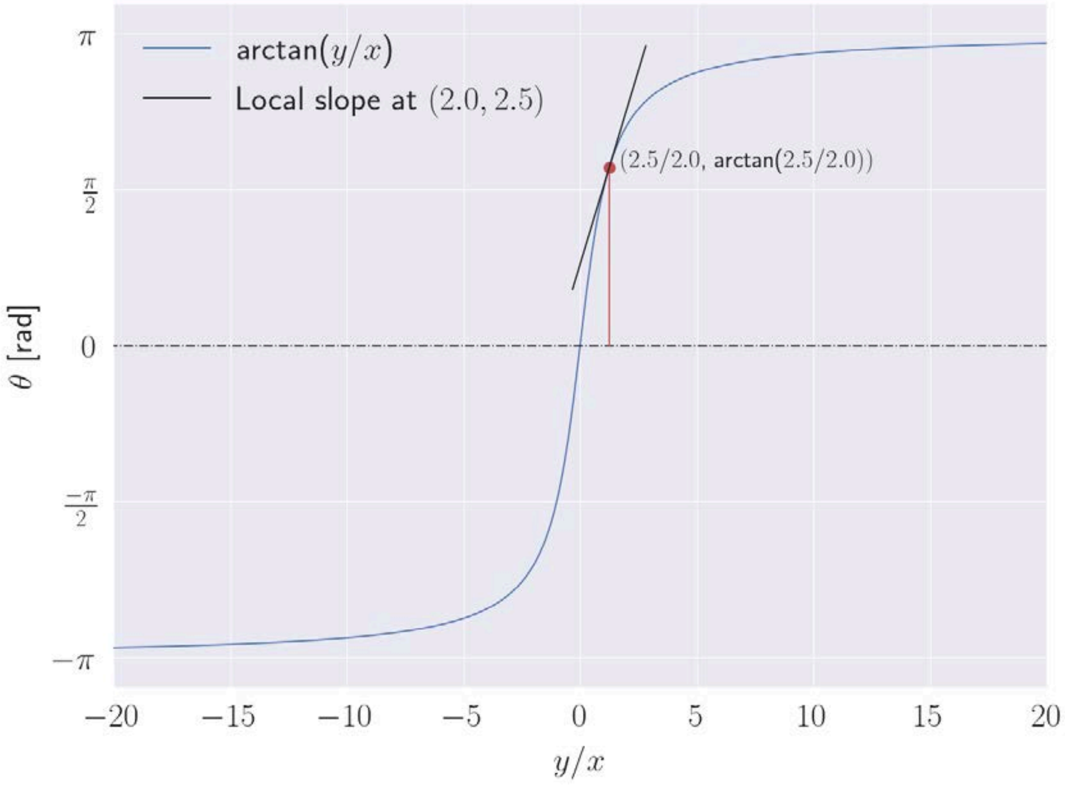
$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t$$

Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\Big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a} - \mathbf{s})$$

Jacobian ($M \times N$ matrix): $$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$$



Applied to non-linear, Gaussian state-space model:

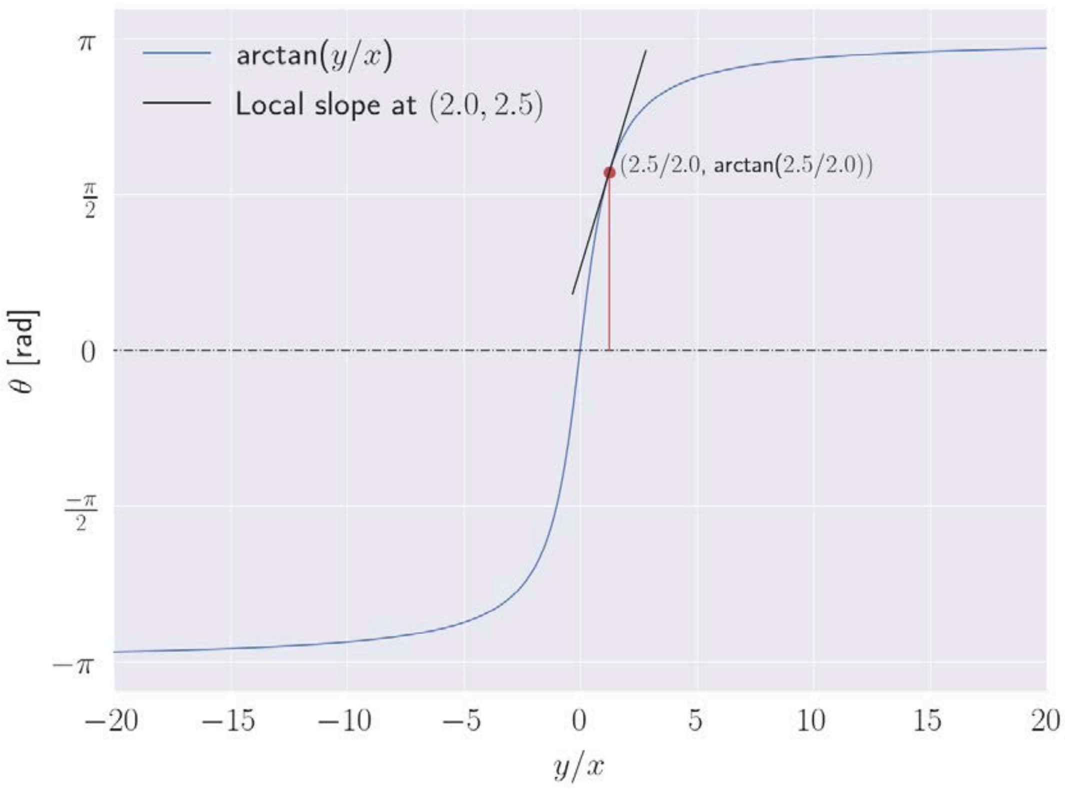
$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t \approx f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right) + \mathbf{v}_t$$

Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a}-\mathbf{s})$$

Jacobian ($M \times N$ matrix): $$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$$



Applied to non-linear, Gaussian state-space model:

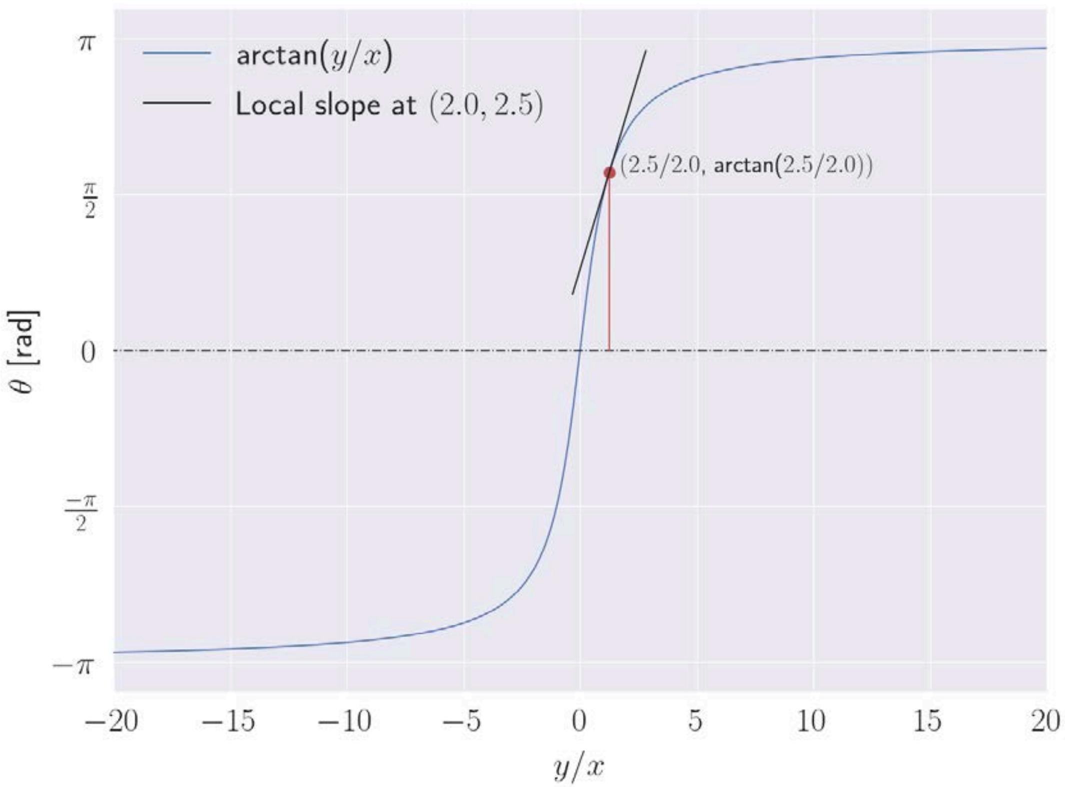
$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t \approx f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right) + \mathbf{v}_t \Rightarrow p\left(\mathbf{z}_t \mid \mathbf{z}_{t-1}\right) \approx \mathcal{N}\left(f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right), \mathbf{Q}\right)$$

Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\Big|_{\mathbf{a}=\mathbf{s}} (\mathbf{a} - \mathbf{s})$$

Jacobian ($M \times N$ matrix): $\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$



Applied to non-linear, Gaussian state-space model:

$$\mathbf{z}_t = f(\mathbf{z}_{t-1}) + \mathbf{v}_t \approx f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}} (\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}) + \mathbf{v}_t \Rightarrow p(\mathbf{z}_t \mid \mathbf{z}_{t-1}) \approx \mathcal{N}\left(f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}} (\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}), \mathbf{Q} \right)$$

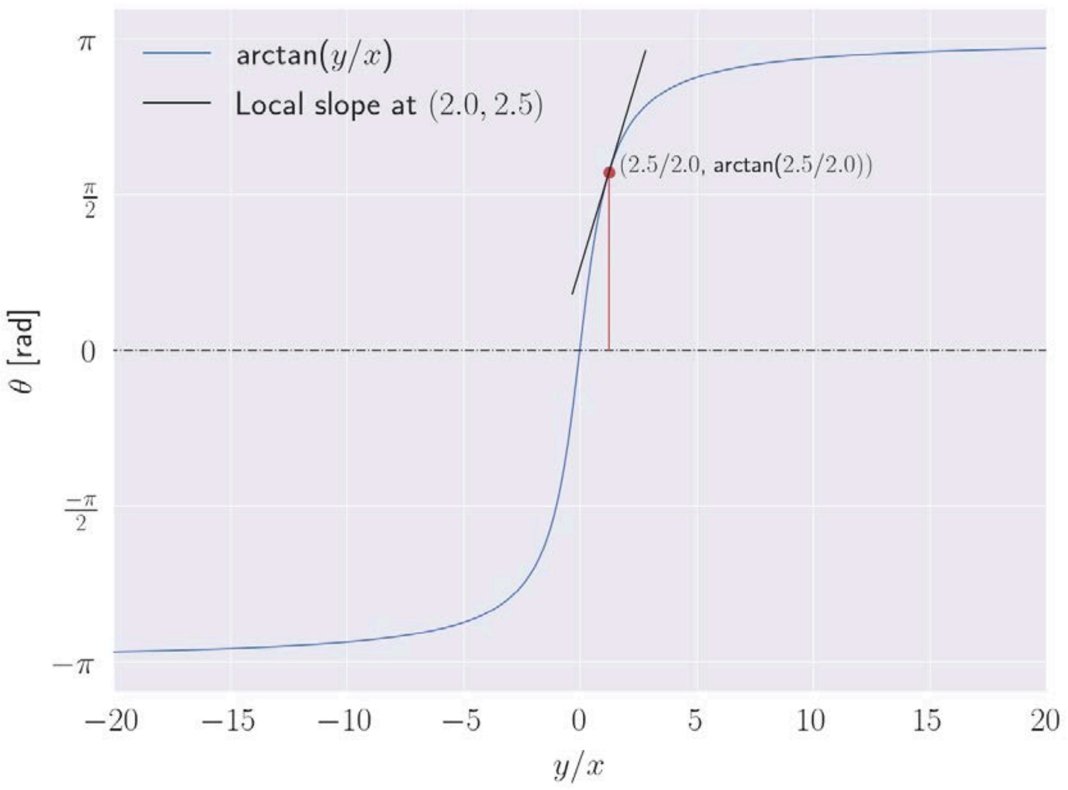
$$\mathbf{x}_t = h(\mathbf{z}_t) + \mathbf{w}_t$$

Local Linearisation

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\Big|_{\mathbf{a}=\mathbf{s}} (\mathbf{a} - \mathbf{s})$$

Jacobian ($M \times N$ matrix): $\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$



Applied to non-linear, Gaussian state-space model:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t \approx f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}} \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right) + \mathbf{v}_t \Rightarrow p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) \approx \mathcal{N}\left(f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}} \left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}} \left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + \mathbf{w}_t$$

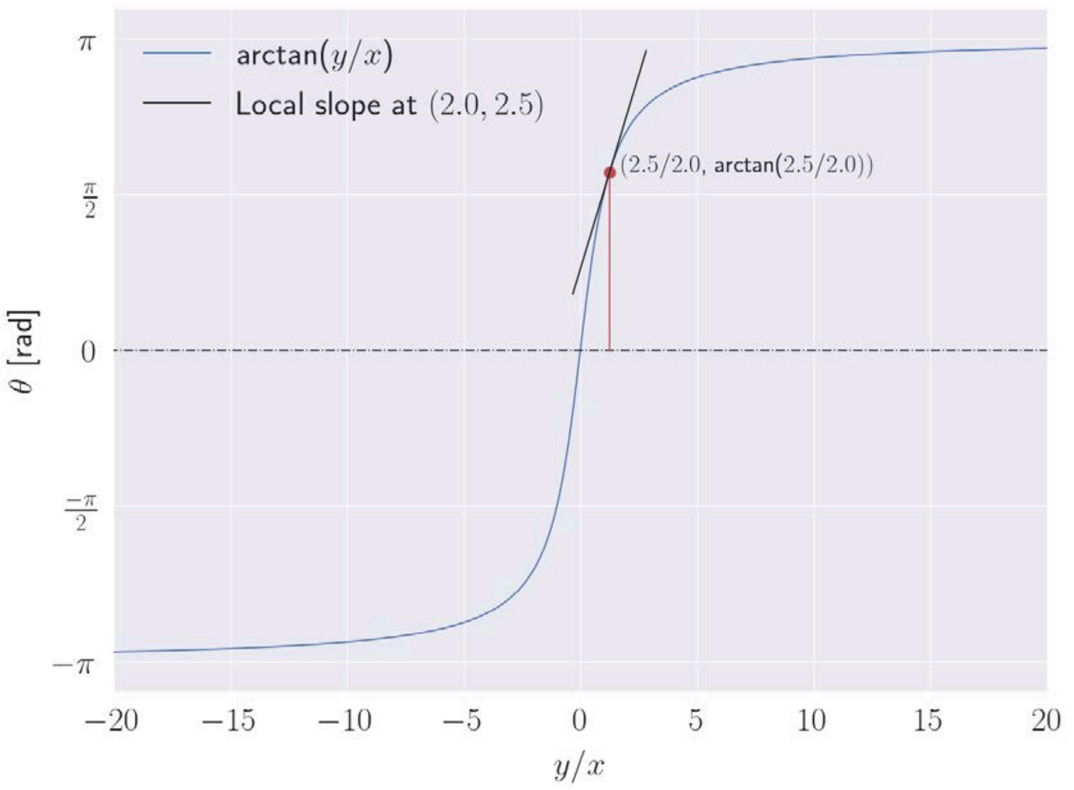
Local Linearisation

UNIVERSITY OF Southampton

First-order Taylor Expansion of $g : \mathbb{R}^N \rightarrow \mathbb{R}^M$:

$$g(\mathbf{a}) \approx g(\mathbf{s}) + \mathbf{J}\Big|_{\mathbf{a}=\mathbf{s}}(\mathbf{a} - \mathbf{s})$$

Jacobian ($M \times N$ matrix): $$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M}{\partial a_1} & \cdots & \frac{\partial g_M}{\partial a_N} \end{bmatrix}$$



Applied to non-linear, Gaussian state-space model:

$$\mathbf{z}_t = f\left(\mathbf{z}_{t-1}\right) + \mathbf{v}_t \approx f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right) + \mathbf{v}_t \quad \Rightarrow \quad p\left(\mathbf{z}_t \,|\, \mathbf{z}_{t-1}\right) \approx \mathcal{N}\left(f(\boldsymbol{\mu}_{t-1}) + \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}\right), \mathbf{Q}\right)$$

$$\mathbf{x}_t = h\left(\mathbf{z}_t\right) + \mathbf{w}_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + \mathbf{w}_t \quad \Rightarrow \quad p\left(\mathbf{x}_t \,|\, \mathbf{z}_t\right) \approx \mathcal{N}\left(h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right), \mathbf{R}\right)$$

Extended Kalman Filter (EKF)

Predictive pdf: $$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$$

Posterior pdf: $$p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \boldsymbol{\Sigma}_{t-1} \hat{\mathbf{F}}_t^T$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \boldsymbol{\Sigma}_{t-1} \hat{\mathbf{F}}_t^T$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \hat{\mathbf{H}}_t^T \left(\hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1} \, \hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F} \, \boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F} \, \boldsymbol{\Sigma}_{t-1} \, \mathbf{F}^T$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \hat{\mathbf{H}}_t^T \left(\hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1} \, \hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F} \, \boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F} \, \boldsymbol{\Sigma}_{t-1} \, \mathbf{F}^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - \mathbf{H} \, \boldsymbol{\mu}_{t|t-1}\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}^T \left(\mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1} \, \mathbf{H}^T + \mathbf{R}\right)^{-1}$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \mid \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \mid \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \hat{\mathbf{H}}_t^T \left(\hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1} \, \hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F} \, \boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F} \, \boldsymbol{\Sigma}_{t-1} \, \mathbf{F}^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - \mathbf{H} \, \boldsymbol{\mu}_{t|t-1}\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}^T \left(\mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1} \, \mathbf{H}^T + \mathbf{R}\right)^{-1}$$

Extended Kalman Filter (EKF)

Predictive pdf: $p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t-1}) \approx \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}\right)$

Posterior pdf: $p(\mathbf{z}_t \,|\, \mathbf{x}_{1:t}) \approx \mathcal{N}\left(\mathbf{z}_t \,|\, \boldsymbol{\mu}_{t}, \boldsymbol{\Sigma}_{t}\right)$

EKF: Non-linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = f\left(\boldsymbol{\mu}_{t-1}\right)$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \hat{\mathbf{F}}_t \, \boldsymbol{\Sigma}_{t-1} \, \hat{\mathbf{F}}_t^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \hat{\mathbf{H}}_t^T \left(\hat{\mathbf{H}}_t \, \boldsymbol{\Sigma}_{t|t-1} \, \hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$$

where $\hat{\mathbf{F}}_t \triangleq \hat{\mathbf{F}}\Big|_{\mathbf{z}_{t-1}=\boldsymbol{\mu}_{t-1}}$ and $\hat{\mathbf{H}}_t \triangleq \hat{\mathbf{H}}\Big|_{\mathbf{z}_{t}=\boldsymbol{\mu}_{t|t-1}}$

Kalman Filter: Linear, Gaussian state spaces

$$\boldsymbol{\mu}_{t|t-1} = \mathbf{F} \, \boldsymbol{\mu}_{t-1}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F} \, \boldsymbol{\Sigma}_{t-1} \, \mathbf{F}^T$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \left(\mathbf{x}_t - \mathbf{H} \, \boldsymbol{\mu}_{t|t-1}\right)$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t \, \mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1}$$

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}^T \left(\mathbf{H} \, \boldsymbol{\Sigma}_{t|t-1} \, \mathbf{H}^T + \mathbf{R}\right)^{-1}$$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

for $t = 0, \ldots, \infty$:

Prediction:

1. Mean: $\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$

2. Covariance: $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

for $t = 0, \ldots, \infty$:

Prediction:

1. Mean: $\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$
2. Covariance: $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$

Update:

3. Jacobian: $\hat{\mathbf{H}}_t = \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2+y_k^2} \quad \frac{x_k}{x_k^2+y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

for $t = 0,\ldots,\infty$:

Prediction:

1. Mean: $\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$
2. Covariance: $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$

Update:

3. Jacobian: $\hat{\mathbf{H}}_t = \left[-\frac{y_k}{x_k^2+y_k^2} \quad \frac{x_k}{x_k^2+y_k^2} \quad 0 \quad 0\right]\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$
4. Kalman gain: $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\hat{\mathbf{H}}_t^T\left(\hat{\mathbf{H}}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

for $t = 0,\ldots,\infty$:

Prediction:

1. Mean: $\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$
2. Covariance: $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$

Update:

3. Jacobian: $\hat{\mathbf{H}}_t = \left[-\frac{y_k}{x_k^2 + y_k^2} \quad \frac{x_k}{x_k^2 + y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$
4. Kalman gain: $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\hat{\mathbf{H}}_t^T\left(\hat{\mathbf{H}}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$
5. Mean: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\left(\theta_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$

EKF for Bearing-Only Tracking

Non-linear, Gaussian state space:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t$$

$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t \approx h(\boldsymbol{\mu}_{t|t-1}) + \hat{\mathbf{H}}_t\left(\mathbf{z}_t - \boldsymbol{\mu}_{t|t-1}\right) + w_t$$

Jacobian for bearing-only observations:

$$\hat{\mathbf{H}}_t = \left[\frac{\partial h}{\partial x_t} \quad \frac{\partial h}{\partial y_t} \quad \frac{\partial h}{\partial \dot{x}_t} \quad \frac{\partial h}{\partial \dot{y}_t}\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

$$= \left[-\frac{y_k}{x_k^2+y_k^2} \quad \frac{x_k}{x_k^2+y_k^2} \quad 0 \quad 0\right]\Bigg|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$$

for $t = 0,\ldots,\infty$:

Prediction:

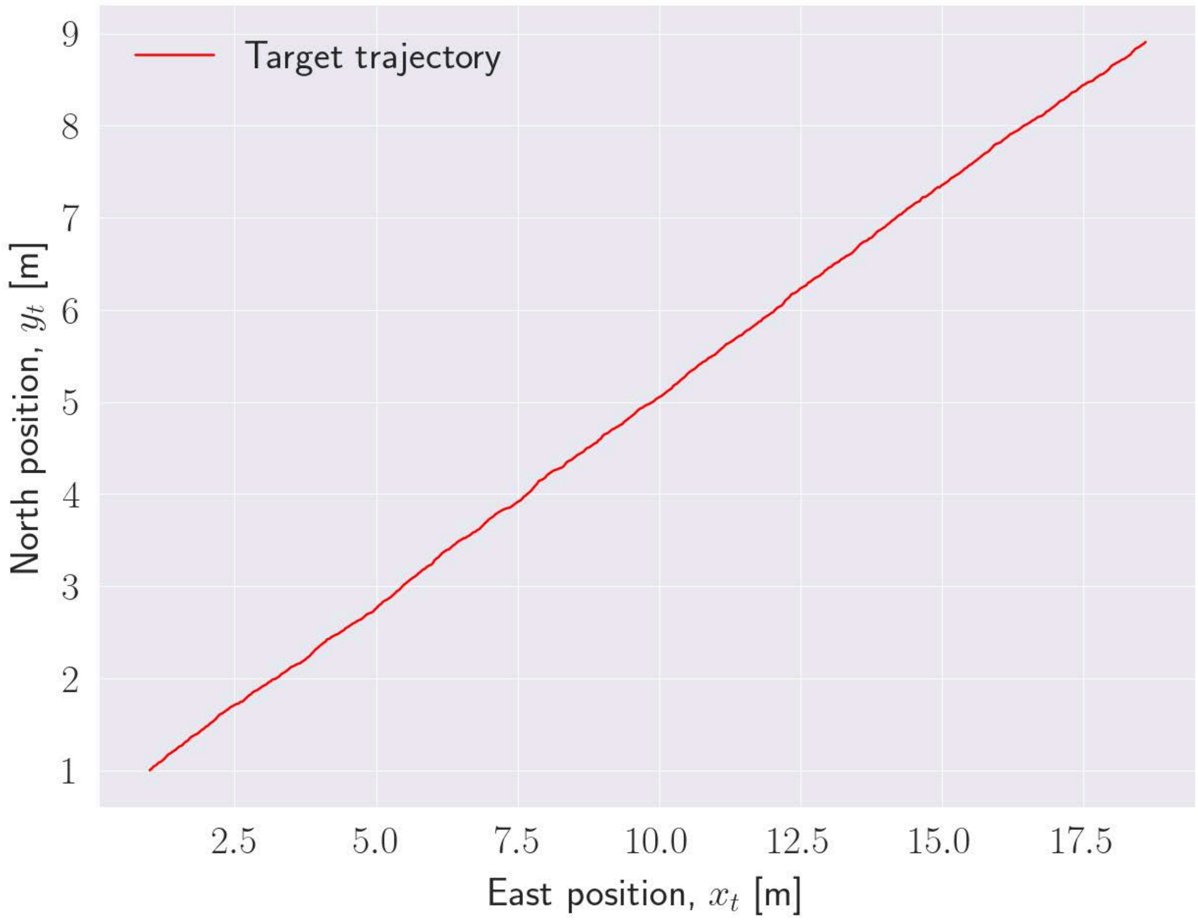
1. Mean: $\boldsymbol{\mu}_{t|t-1} = \mathbf{F}\,\boldsymbol{\mu}_{t-1}$
2. Covariance: $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{Q} + \mathbf{F}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{F}^T$

Update:

3. Jacobian: $\hat{\mathbf{H}}_t = \left[-\frac{y_k}{x_k^2+y_k^2} \quad \frac{x_k}{x_k^2+y_k^2} \quad 0 \quad 0\right]\Big|_{\mathbf{z}_t=\boldsymbol{\mu}_{t|t-1}}$
4. Kalman gain: $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\hat{\mathbf{H}}_t^T\left(\hat{\mathbf{H}}_t\,\boldsymbol{\Sigma}_{t|t-1}\,\hat{\mathbf{H}}_t^T + \mathbf{R}\right)^{-1}$
5. Mean: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\left(\theta_t - h\left(\boldsymbol{\mu}_{t|t-1}\right)\right)$
6. Covariance: $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\,\hat{\mathbf{H}}_t\,\boldsymbol{\Sigma}_{t|t-1}$

Bearing-Only EKF Demo

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t, \text{ where } \left[x_t, y_t, \dot{x}_{t-1}, \dot{y}_{t-1}\right]^T$$

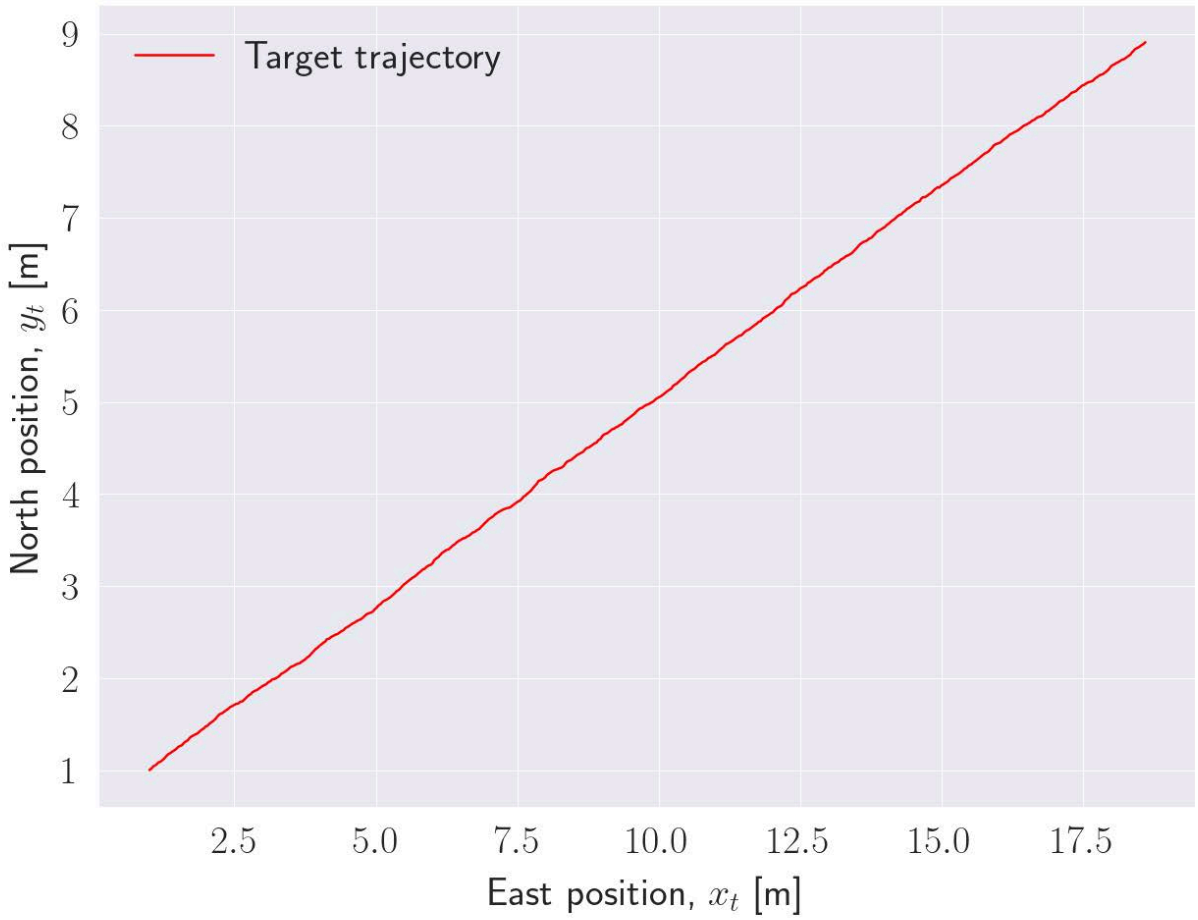


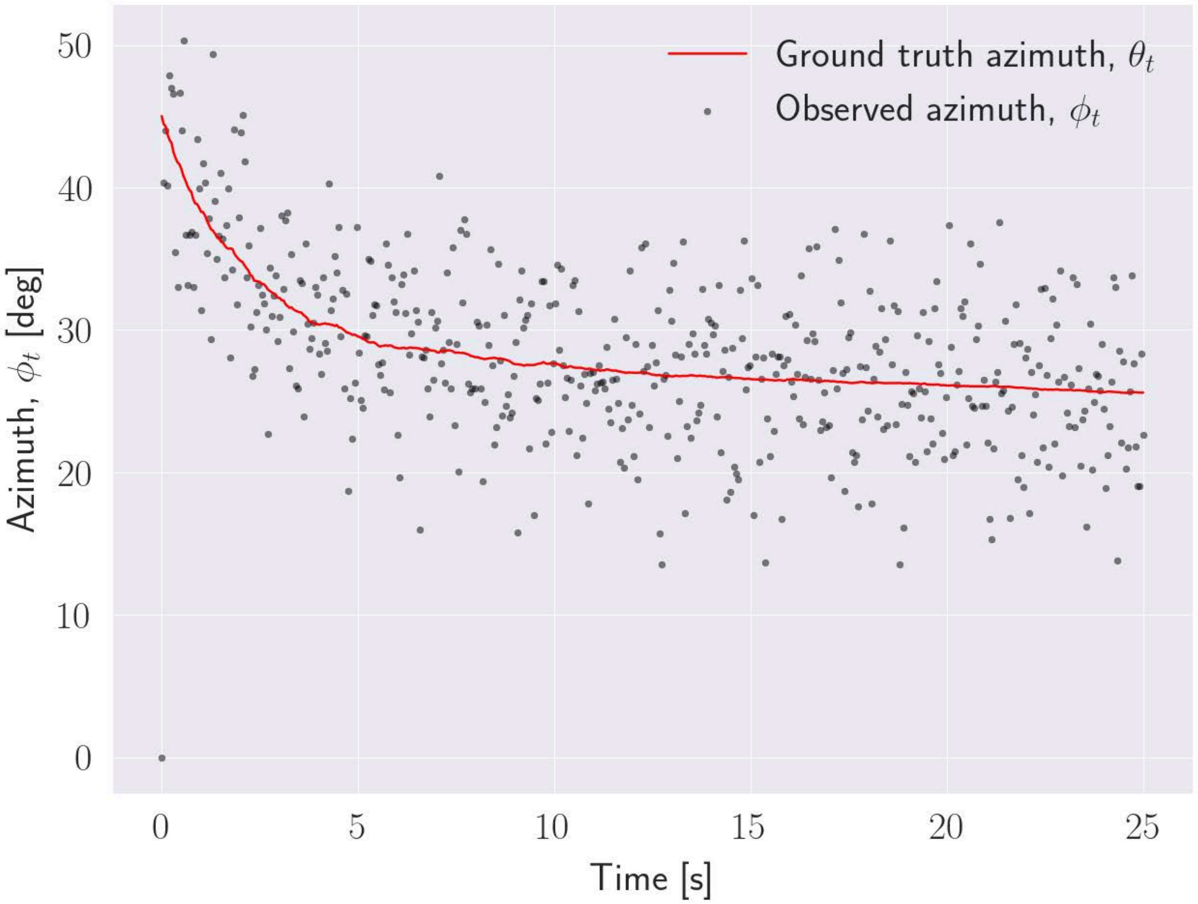
Bearing-Only EKF Demo

UNIVERSITY OF Southampton

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{v}_t, \text{ where } \left[x_t, y_t, \dot{x}_{t-1}, \dot{y}_{t-1}\right]^T$$

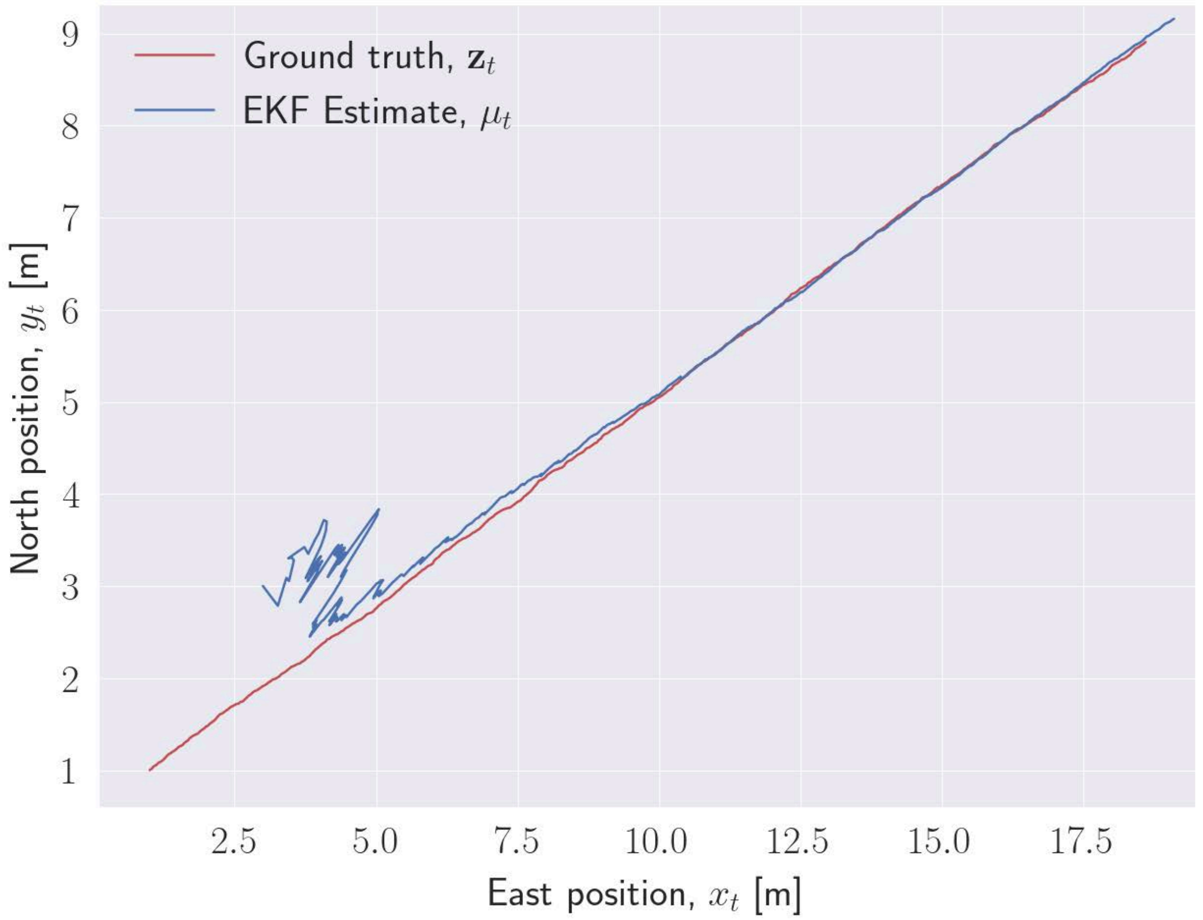
$$\theta_t = \text{atan2}\left(\frac{y_t}{x_t}\right) + w_t$$

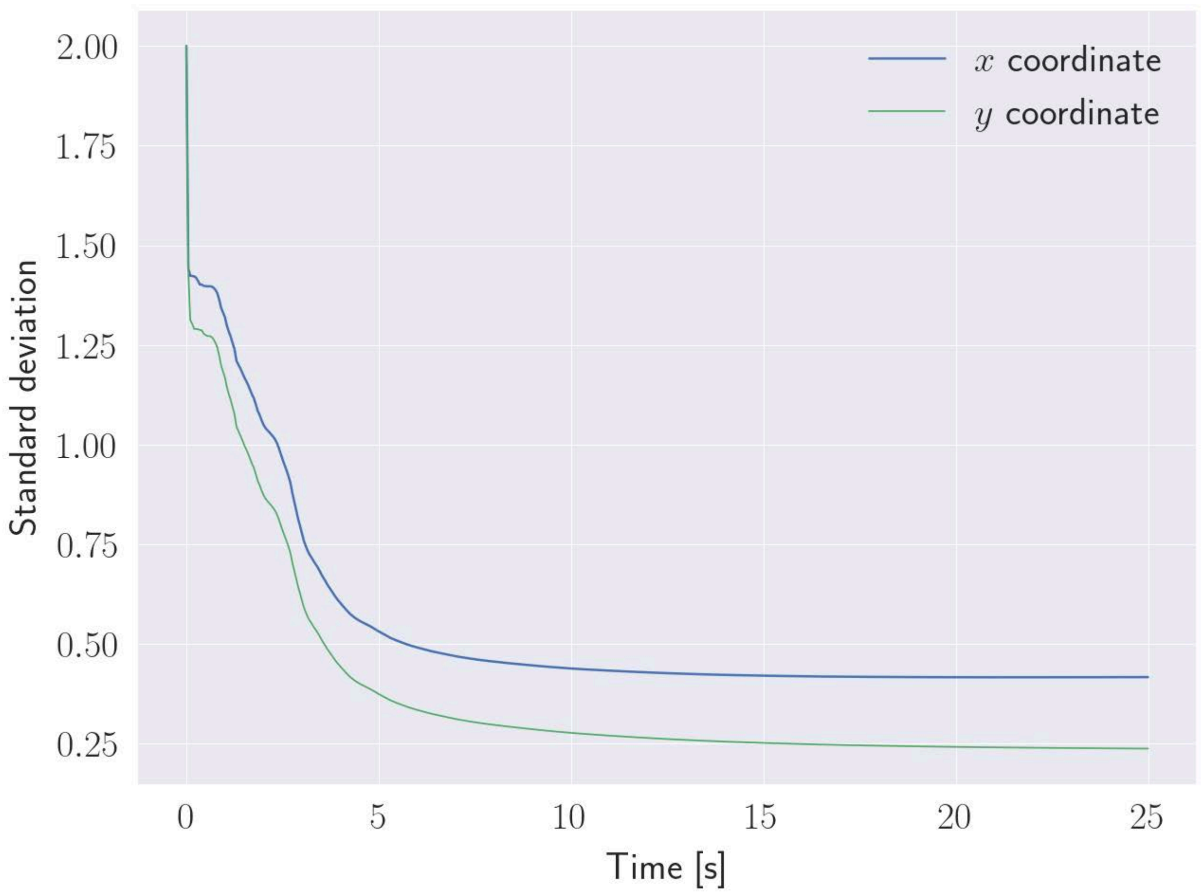




Bearing-Only EKF Demo

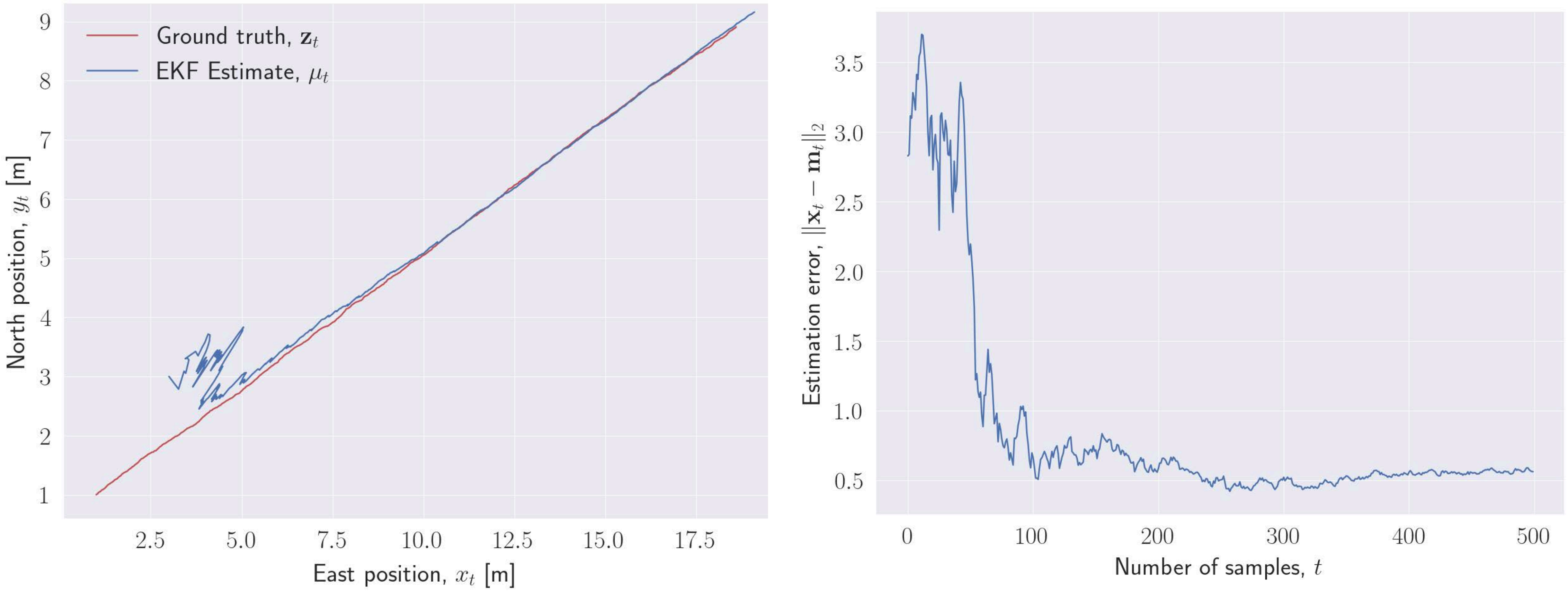
Initial state: 2m error in position, velocity assumed known





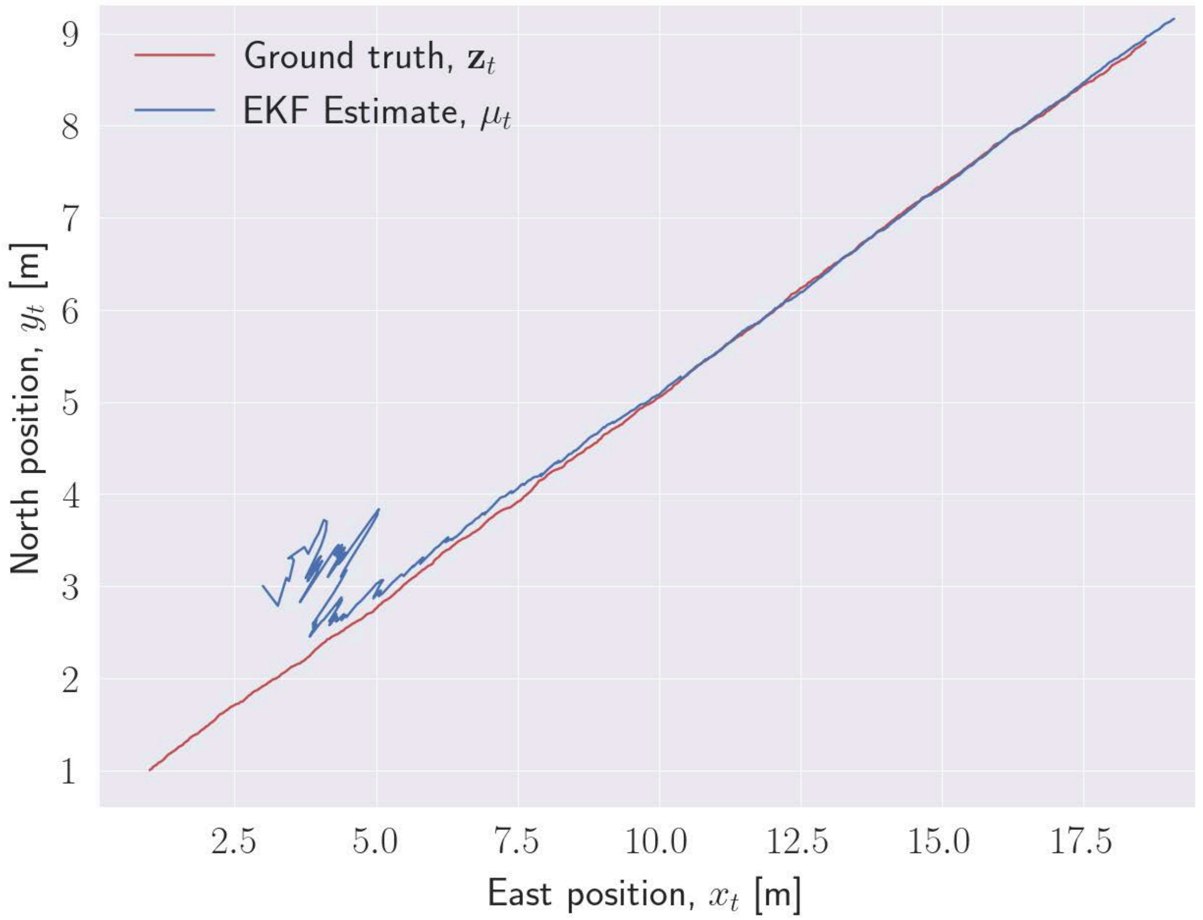
Bearing-Only EKF Demo

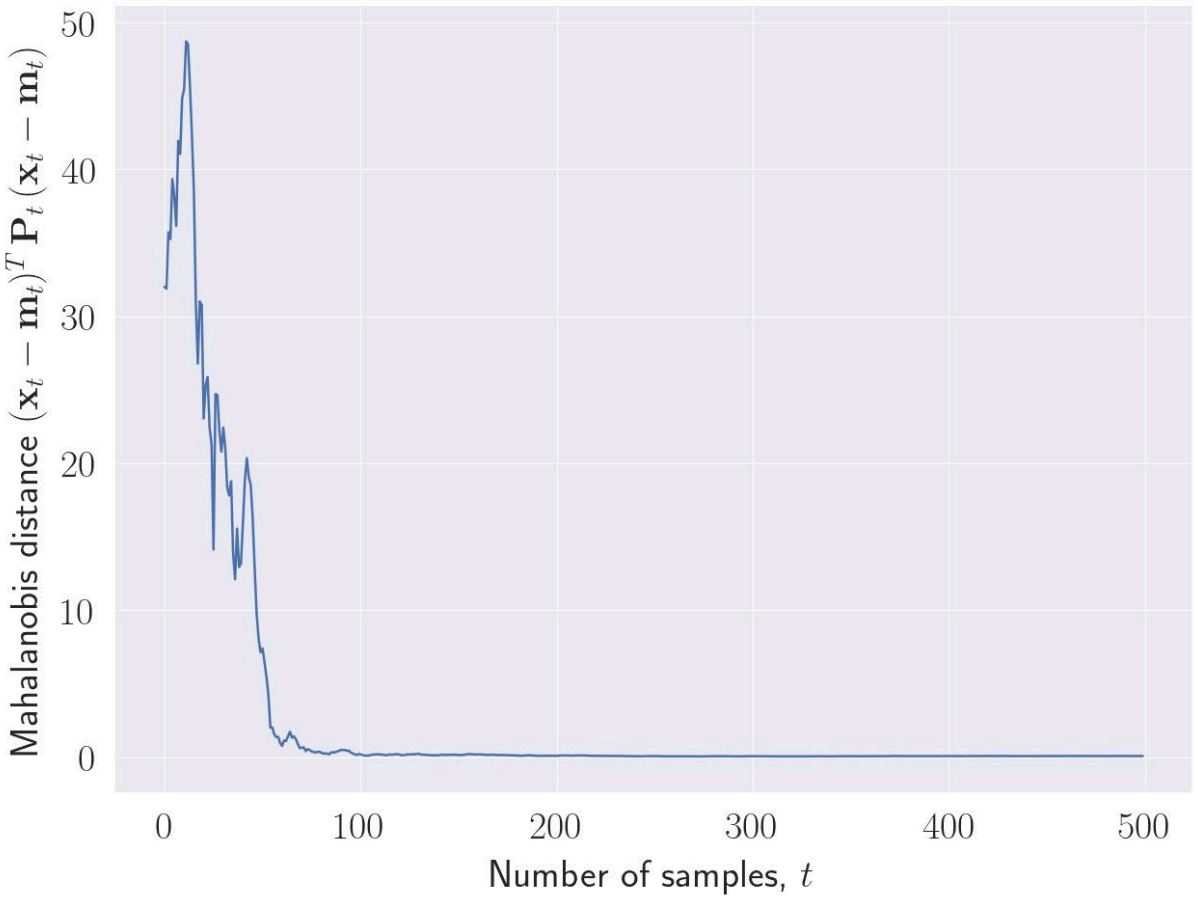
Initial state: 2m error in position, velocity assumed known



Bearing-Only EKF Demo

Initial state: 2m error in position, velocity assumed known

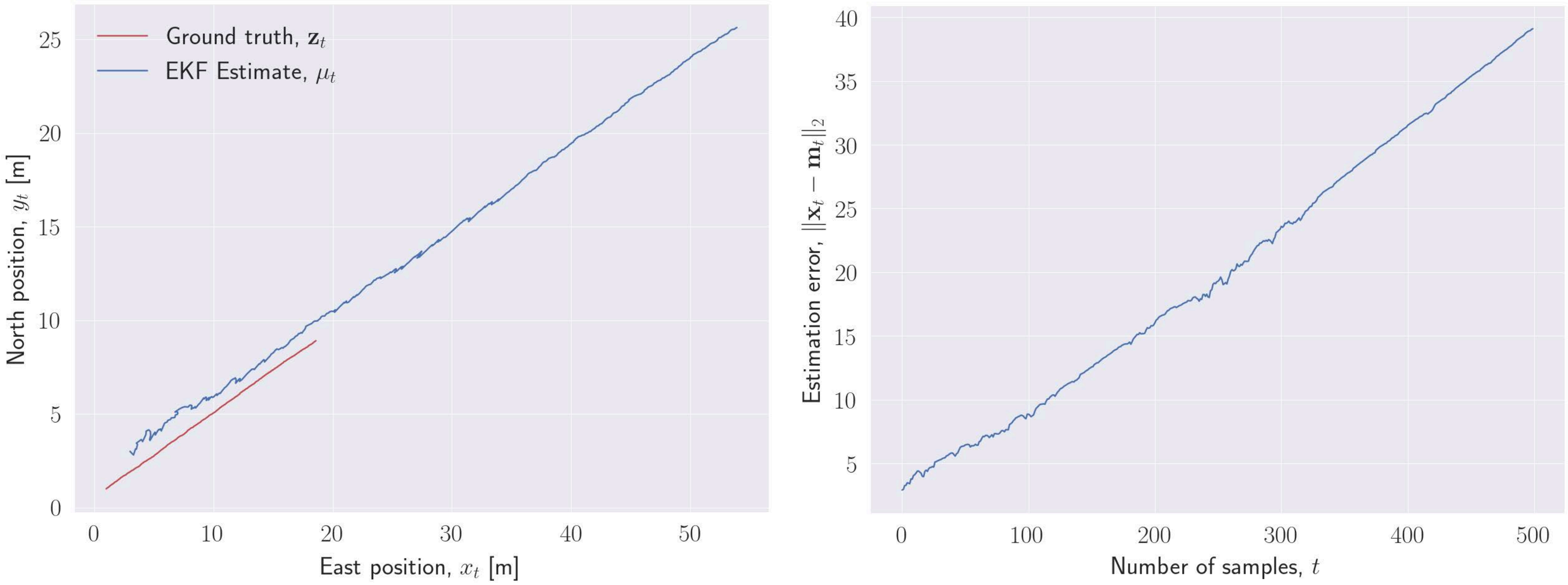




Bearing-Only EKF Demo

UNIVERSITY OF Southampton

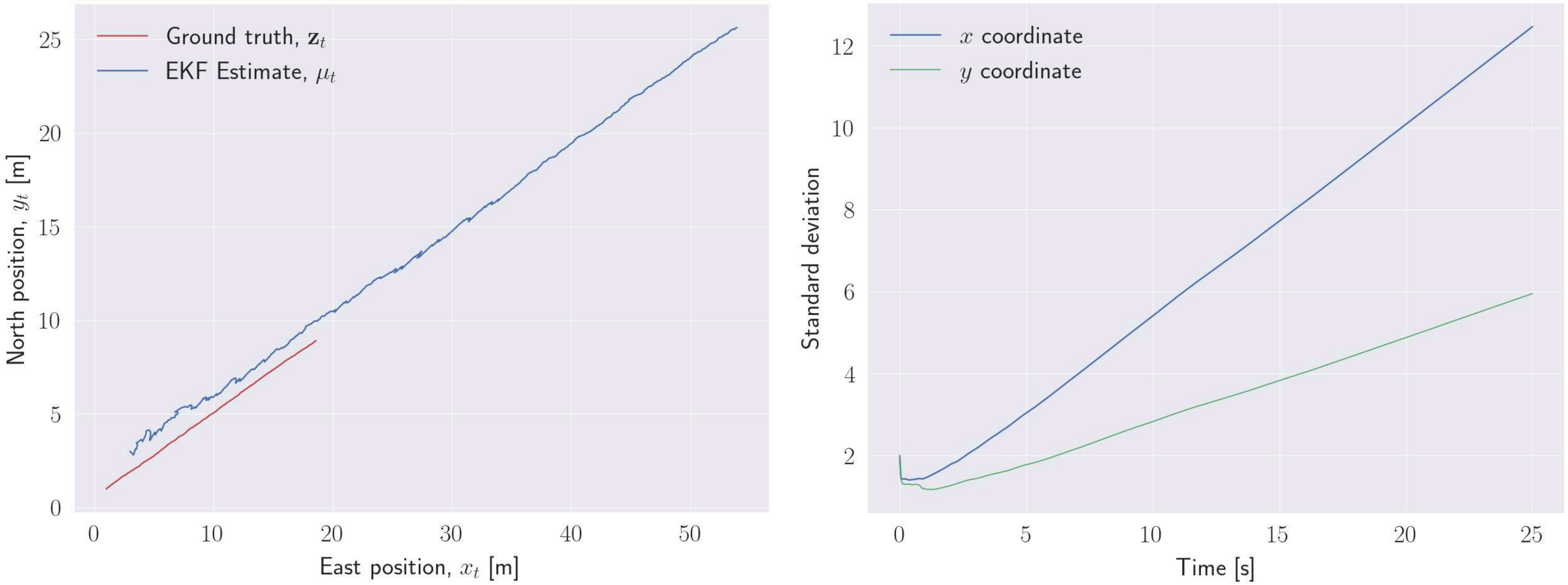
Initial state: 2m error in position, 0.5m/s error in velocity



Bearing-Only EKF Demo

UNIVERSITY OF Southampton

Initial state: 2m error in position, 0.5m/s error in velocity



Local Linearisation - Conclusion

Approximation of non-linearities:

- Local linearisation using on first-order Taylor expansion

Local Linearisation - Conclusion

Approximation of non-linearities:
- Local linearisation using on first-order Taylor expansion

Advantages:
- Easy to implement, efficient to run

Local Linearisation - Conclusion

Approximation of non-linearities:
- Local linearisation using on first-order Taylor expansion

Advantages:
- Easy to implement, efficient to run

Disadvantages:
- Accuracy of the linearization depends on two factors:
    - Degree of uncertainty
    - Degree of local nonlinearity in the functions being approximated.
    - In practice, often a poor approximation of most non-linear functions

Local Linearisation - Conclusion

Approximation of non-linearities:
- Local linearisation using on first-order Taylor expansion

Advantages:
- Easy to implement, efficient to run

Disadvantages:
- Accuracy of the linearization depends on two factors:
    - Degree of uncertainty
    - Degree of local nonlinearity in the functions being approximated.
    - In practice, often a poor approximation of most non-linear functions
- Is Gaussianity an appropriate assumption?

Learning Outcomes

Learning Outcomes

Following this week's lecture, you should be able to:

Learning Outcomes

Following this week's lecture, you should be able to:

1) **Distinguish between the frequentist and Bayesian paradigms**
   - Frequency of occurrences vs belief in models
   - Likelihood vs posterior pdf

Learning Outcomes

Southampton UNIVERSITY OF

Following this week's lecture, you should be able to:

1) **Distinguish between the frequentist and Bayesian paradigms**
   - Frequency of occurrences vs belief in models
   - Likelihood vs posterior pdf
2) **Understand the mathematical framework for Bayesian inference**
   - Prior, likelihood, Posterior → Prediction & Update
   - *We need to solve integrals!* Analytically tractable for linear, Gaussian state-spaces, but generally intractable for non-linear state-spaces.

Learning Outcomes

Following this week's lecture, you should be able to:

1) **Distinguish between the frequentist and Bayesian paradigms**
   - Frequency of occurrences vs belief in models
   - Likelihood vs posterior pdf
2) **Understand the mathematical framework for Bayesian inference**
   - Prior, likelihood, Posterior $\rightarrow$ Prediction & Update
   - *We need to solve integrals!* Analytically tractable for linear, Gaussian state-spaces, but generally intractable for non-linear state-spaces.
3) **Apply techniques for local linearisation to approximate integrals in non-linear state-spaces**
   - First order Taylor expansion $\rightarrow$ Jacobian
   - Extended Kalman filter

Next Week

Tuesday:
Q&A Week 4 material
Lab / Q&A Kalman Filter Coursework

Week 5: Monte Carlo methods
Methods that rely on random sampling

Week 6: Sequential Monte Carlo
Online learning for sequential data