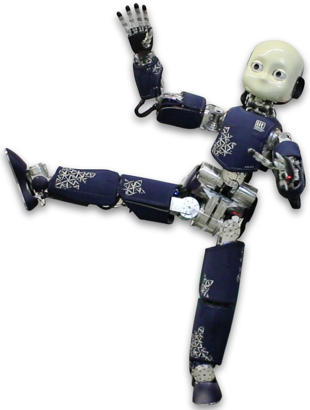
Time Series Models & Databases

Applied Machine & Deep Learning (190.015)

Univ.-Prof. Dr. Elmar Rueckert

Telefon: +43 3842 402 - 1901
Email: teaching@ai-lab.science

WO AUS FORSCHUNG ZUKUNFT WIRD



Chair of Cyber-Physical-Systems

1st Week:

	MON 02.10.2023	TUE 03.10.2023	WED 04.10.2023	THUR 05.10.2023	FRI 06.10.2023
Topic	Intro to ML Organisation	Neural Networks	Representation Learning	Robot Learning	AML Projects
9 am					
:15					
:30					
:45					
10 am					
:15	Quizz on ML	Quizz on Neural Nets	Introduction to Deep Representation Learning		Quizz on AML
:30	Introduction to ML	Introduction to Multi-Layer-Perceptrons			Project Topic Presentations
:45					
11 am	15 min Break	15 min Break	JupyterHub NB on Rep. Learning		
:15	Statistics, Model Validation, Figures & Evaluations	Handout on Neural Networks using playground.tensorflow			Team Ass., Git Repos & Wiki Instructions
:30			30 min Break		
:45					AML Summary
12 pm	30 min Lunch Break	30 min Lunch Break	Curiosity (MLPs), Imagination (Dreamer) and Information (Empowerment)		
:15				Quizz on Robotics	
:30	Course Organisation & Grading	Introduction to CNNs		Introduction to Robot Learning	
:45					
1 pm	15 min Break	15 min Break	Quizz Summary		
:15	Python Programming with our JupyterHub	JupyterHub NB on MLPs CNNs		15 min Break	
:30				Handout on Robot Learning (Model Learning & RL)	
:45	Quizz Summary	Quizz Summary			
2 pm					
:15				15 min Break	
:30				Introduction to Mobile Robotics & SLAM	
:45					
3 pm				JupyterHub NB on Path Planning	
:15					
:30				Quizz Summary	
:45					

Legend	
Quizz on ML	Online Quizz using https://tweedback.de
Course Content Presentation	Using google slides, etc.
15 min Break	Breaks to recover or to continue programming
Organisation & Instructions	Using google slides, etc.
Practical Exercise	Using online tools, our JupyterHub, etc.
Latest Research	State-of-the-art research

Outlook of this lecture

- An introduction to Probabilistic Time Series Models
    - Single Time Step Model
    - Multi Time step Model
- Some Applications in Research
- Let’s implement Databases in Jupyter Hub

Probabilistic Time-Series Models (L8)

Machine Learning (190.012)

Univ.-Prof. Dr. Elmar Rueckert[1]

[1] *teaching@ai-lab.science*, Montanuniversität Leoben, Austria

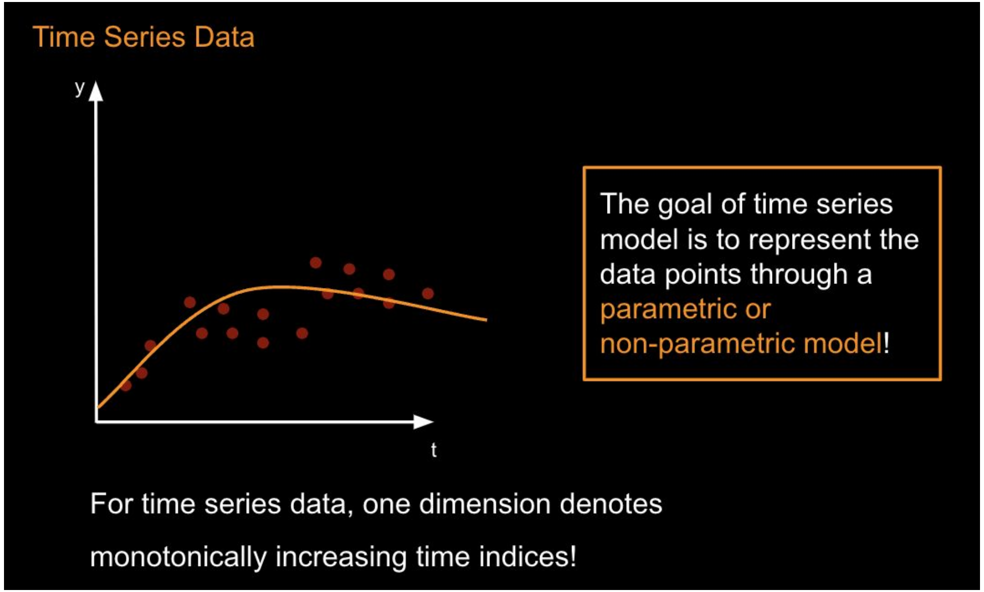
June 3, 2022

Chair of Cyber-Physical-Systems

For more details, please have a look at these slides: https://cloud.cps.unileoben.ac.at/index.php/s/YTNmPbgsSbTFdsZ.

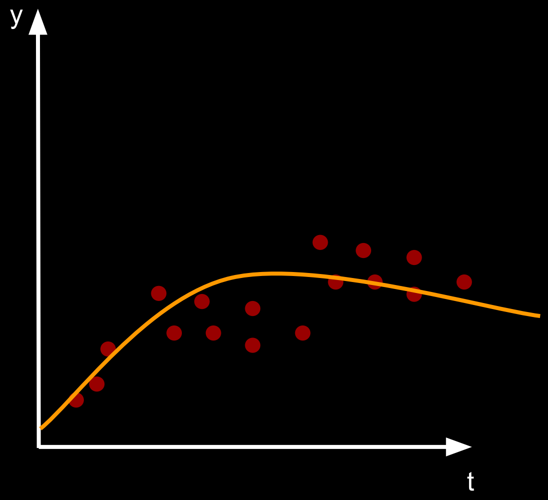
Definition of Time Series

Time series are defined as a **series of data points** indexed or listed **in time order**.



[Slides on Time Series by Prof. Rueckert.](#)

Time Series Data



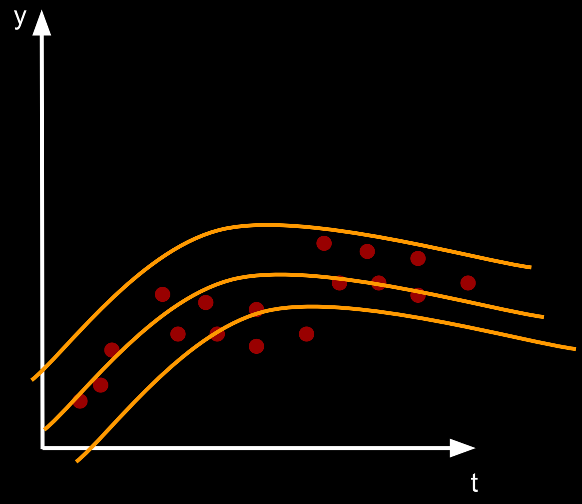
<u>Definition:</u> A **trajectory** is a vector of time series data of any type (e.g., joint angles, forces, Cartesian coordinates, or inhomogeneous mixtures)!

Consider D as vector with $D = \{y_1, y_2, \ldots, y_T\}$

representing 1-dimensional time series data with the

time indices t=1...T!

Time Series Data

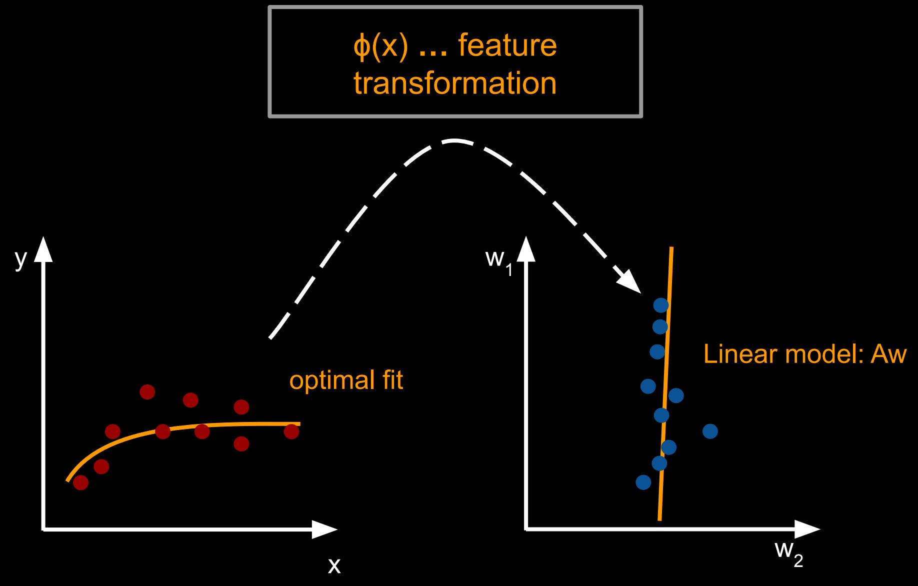


Attention: Individual dimensions (e.g. the x and y Cartesian coord.) are organized in a sequential order where first the complete dimension one is stored!

For multiple trajectories (e.g., recordings), we define D as vector with $D = \{y_{1,1}, y_{2,1}, \ldots, y_{T,1}, y_{2,1}, \ldots, y_{T,n}\}$, where n denotes the number of trajectories!

Single-time step Model

Well that are our simple probabilistic regression models, summarized here!



Ridge Regression	Maximum A-Posteriori
$\boldsymbol{w} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{y}$	$\boldsymbol{w} = (\boldsymbol{A}^T\boldsymbol{A} + \sigma^2\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{y}$

2 Steps:

1. Learning $p(w) = N(w|\, \mu_{w|y}, \Sigma_{w|y})$

$$\mu_{w|y} = \sum_{j=1}^{n} \boldsymbol{w_j} \qquad \Sigma_{w|y} = cov\,([w_1, \dots, w_n])$$

2 Steps:

1. Learning p(w) = $N(w|\ \mu_{w|y}, \Sigma_{w|y})$
2. Prediction y* = ?

The Marginal distribution is defined as:

p(A) = $\sum_b$ p(A|B) p(B=b) … for discrete distributions

p(a) = $\int_b$ p(a|b) p(b) db … *for continuous distributions*

p(y) = $\int_w$ p(y*|w) p(w) dw = $\int_w$ N(A w, …) N(w*, $\Sigma_{w|y}$) dw*

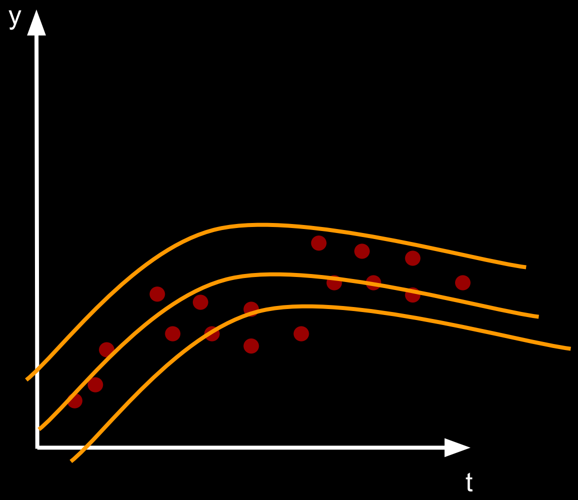
$$p(y^*|\boldsymbol{x}^*, \boldsymbol{X}, \boldsymbol{w}) = \int \mathcal{N}(y^*|\boldsymbol{\phi}(\boldsymbol{x}^*)^T\boldsymbol{w}, \sigma_y^2)\mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_{\boldsymbol{w}|\mathbf{y}}, \boldsymbol{\Sigma}_{\boldsymbol{w}|\mathbf{y}})d\boldsymbol{w}.$$

Marginal or predictive distribution

Likelihood

prior

Multi-time step Model



For simplicity, we consider only 1-dimensional time series data!

D = {$y_{1,1}$, $y_{2,1,}$ ..., $y_{T,1}$, $y_{2,1}$, ..., $y_{T,n}$},
where n denotes the number of trajectories!

The Model (the likelihood):

$$p(y_t|\boldsymbol{w}) = \mathcal{N}(y_t|\boldsymbol{\phi}_t^T \boldsymbol{w}, \sigma_y^2)$$

The Model (the likelihood):

$$p(y_t|w) = \mathcal{N}(y_t|\boldsymbol{\phi}_t^T w, \sigma_y^2)$$

The Basis Functions
(our model assumption):

$$\boldsymbol{\phi}_t : \mathbb{R}^1 \rightarrow \mathbb{R}^M$$

per time-step & per feature i:

$$\phi_t^i = \exp^{\{-\frac{1}{2h}(z_t - c_i)^2\}}$$

per time-step over M features:

$$\boldsymbol{\phi}_t = \frac{1}{\sum_{i=1}^M \phi_t^i}[\phi_t^1, \phi_t^2, \ldots, \phi_t^M]$$

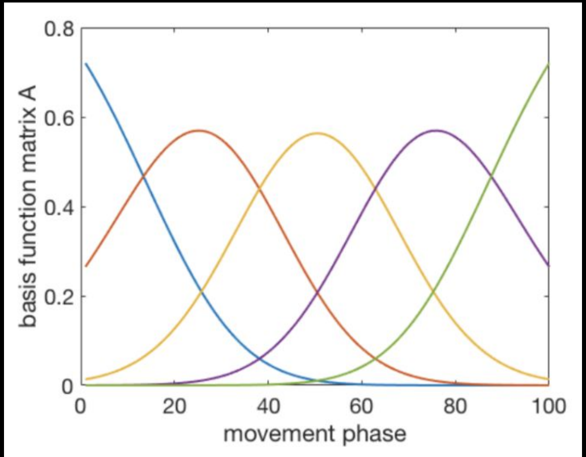
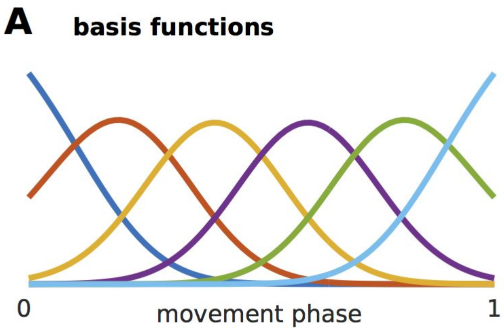
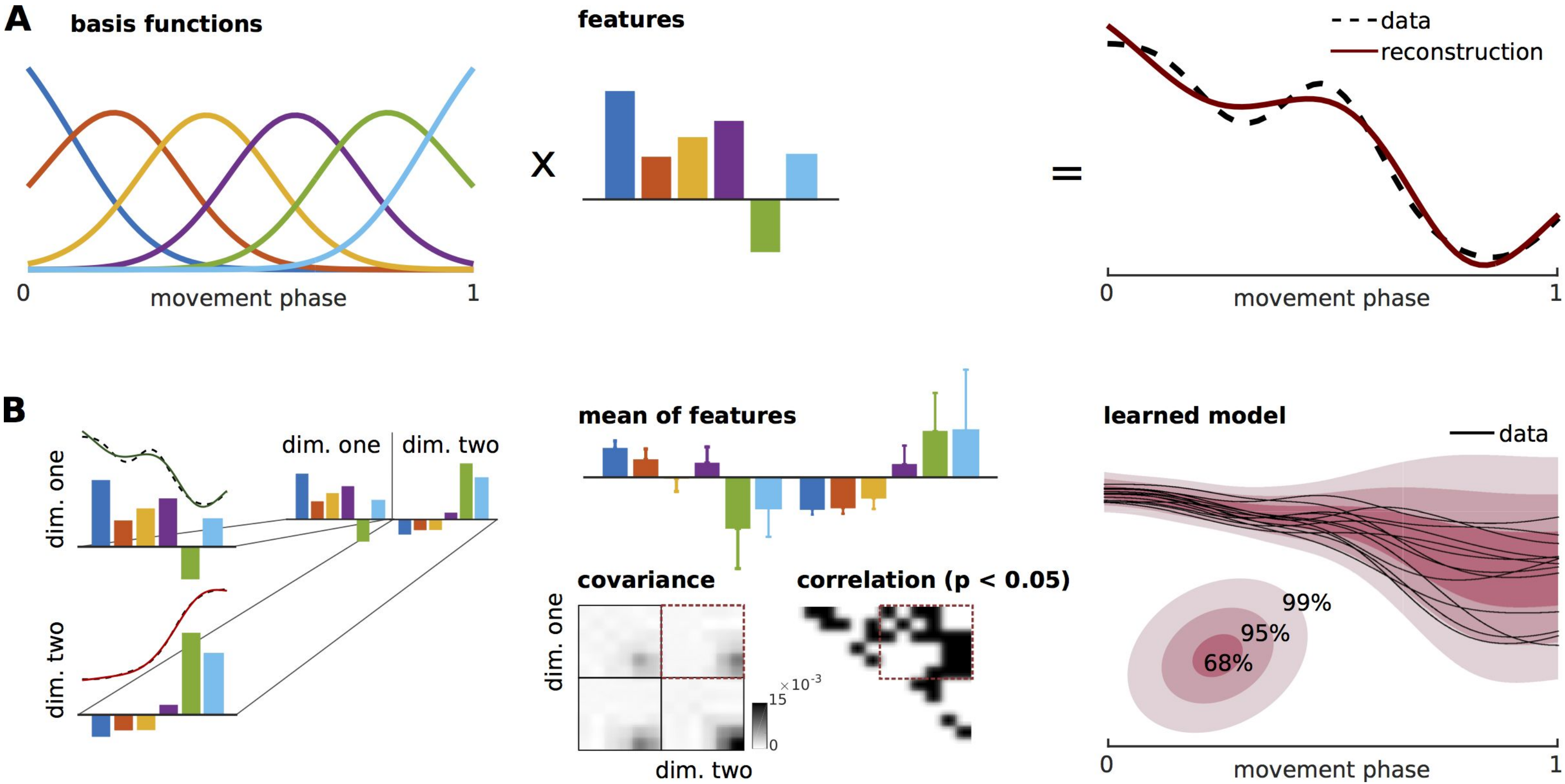


Figure 4.5: Shown are five *normalized* basis functions arranged in the movement phase interval $[0, 1]$.

The model can be used to represent any non-linear function, see the next slide!


A
basis functions
0
movement phase
1



A
basis functions
0
movement phase
1
features
x
=
data
reconstruction
0
movement phase
1
B
dim. one
dim. two
dim. one
dim. two
mean of features
covariance
dim. one
dim. two
×10-3
15
0
correlation (p < 0.05)
learned model
data
99%
95%
68%
0
movement phase
1

The Multi-time Step Model

$$p(\boldsymbol{\tau}_j|\boldsymbol{w}) = \prod_{t=1}^{T} p(y_{t,j}|\boldsymbol{w}) ,$$
$$= \prod_{t=1}^{T} \mathcal{N}(y_{t,j}|\boldsymbol{\phi}_t^T \boldsymbol{w}, \sigma_y^2) ,$$
$$= \mathcal{N}(\boldsymbol{\tau}_j|\boldsymbol{A}\boldsymbol{w}, \sigma_y^2 \boldsymbol{I}) .$$

$$\boldsymbol{A} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_T]^T \in \mathbb{R}^{T \times M}$$

$$\boldsymbol{w}_{LS}^j = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{\tau}_j$$

Assumption: i.i.d. time series data samples!

$$\boxed{\mathcal{N}(a|\mu_1, \sigma_1)\, \mathcal{N}(b|\mu_2, \sigma_2) = \mathcal{N}\left(\begin{bmatrix} a \\ b \end{bmatrix} \middle| \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)}$$

Attention: Individual dimensions (e.g. the x and y Cartesian coord.) are organized in a sequential order where first the complete dimension one is stored, see Slide 5!

Multi-time Step Model

$$
\begin{aligned}
p(\boldsymbol{\tau}_j|\boldsymbol{w}) &= \prod_{t=1}^{T} p(y_{t,j}|\boldsymbol{w}) \,, \\
&= \prod_{t=1}^{T} \mathcal{N}(y_{t,j}|\boldsymbol{\phi}_t^T \boldsymbol{w}, \sigma_y^2) \,, \\
&= \mathcal{N}(\boldsymbol{\tau}_j|\boldsymbol{A}\boldsymbol{w}, \sigma_y^2 \boldsymbol{I}) \,.
\end{aligned}
$$

$$\boldsymbol{A} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_T]^T \in \mathbb{R}^{T\times M}$$

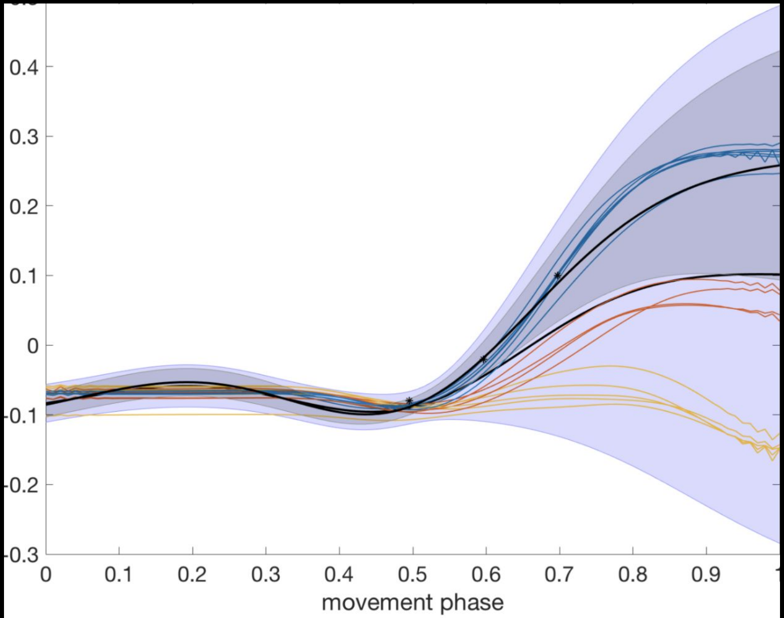
$$\boldsymbol{w}_{LS}^{j} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{\tau}_j$$

Predictions with the multi-time step model

$$p(\boldsymbol{\tau}) = \int \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{A}\boldsymbol{w}, \sigma_y^2 \boldsymbol{I})\, \mathcal{N}(\boldsymbol{w}|\boldsymbol{\mu}_{w|y}, \boldsymbol{\Sigma}_{w|y})\, d\boldsymbol{w},$$
$$= \mathcal{N}(\boldsymbol{\tau} \mid \boldsymbol{A}\boldsymbol{\mu}_{w|y},\ \sigma_y^2 \boldsymbol{I} + \boldsymbol{A}\boldsymbol{\Sigma}_{w|y}\boldsymbol{A}^T).$$

$$\boldsymbol{\mu}_{w|y} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{w}^j \text{ and ,}$$

$$\boldsymbol{\Sigma}_{w|y} = \frac{1}{n-1}\sum_{j=1}^{n} (\boldsymbol{w}^j - \boldsymbol{\mu}_{w|y})(\boldsymbol{w}^j - \boldsymbol{\mu}_{w|y})^T$$



Conditional Rule to predict future time series data points, see [page 31 in the script]:

$$p(a|b) = N(x_a|u_a + \Sigma_{ab}\, \Sigma_{bb}^{-1}\, (b - u_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}^{T})$$

Conditional Rule to predict future time series data points, see page 31 in the script:

$$p(a|b) = N(x_a|u_a + \Sigma_{ab}\,\Sigma_{bb}^{-1}\,(b - u_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}^{T})$$

$$p(\boldsymbol{\tau}^o) = N(\boldsymbol{\tau}^o \mid \boldsymbol{A}^o \boldsymbol{\mu}_{w|o},\ \sigma_y^2 \boldsymbol{I} + \boldsymbol{A}^o \boldsymbol{\Sigma}_{w|o} \boldsymbol{A}^{oT}).$$

$$\begin{aligned}
\boldsymbol{\mu}_{w|o} &= \boldsymbol{\mu}_{w|y} + \boldsymbol{L}(\boldsymbol{o} - \boldsymbol{A}^o \boldsymbol{\mu}_{w|y}) \text{ and ,} \\
\boldsymbol{\Sigma}_{w|o} &= \boldsymbol{\Sigma}_{w|y} - \boldsymbol{L}\ \boldsymbol{A}^o \boldsymbol{\Sigma}_{w|y} \text{ and ,} \\
\boldsymbol{L} &= \boldsymbol{\Sigma}_{w|y} \boldsymbol{A}^{oT} (\sigma_o \boldsymbol{I} + \boldsymbol{A}^o \boldsymbol{\Sigma}_{w|y} \boldsymbol{A}^{oT})^{-1}
\end{aligned}$$

A reformulated version using definitions of the matrix cookbook.

Multi-dimensional multi-time step models

$$A = [\phi_1, \phi_2, \ldots, \phi_T]^T \in \mathbb{R}^{T \times M}$$

$$A \in \mathbb{R}^{TD \times MD} \text{ for } D\text{-dimensional data}$$

$$w^j_{LS} = (A^T A)^{-1} A^T \tau_j$$

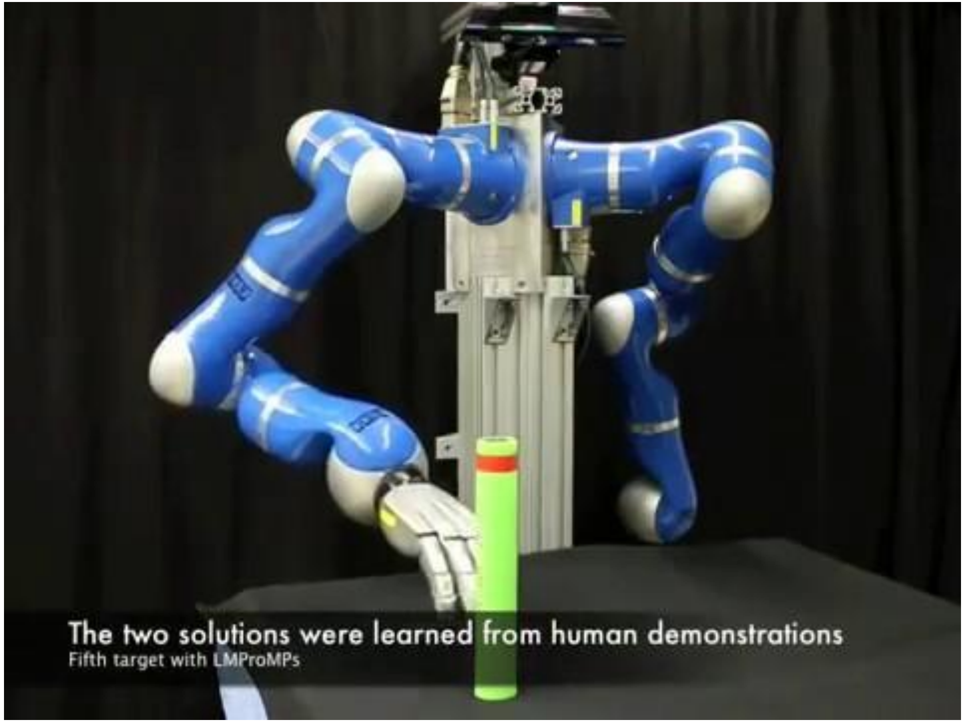
$$w \in \mathbb{R}^{MD}$$

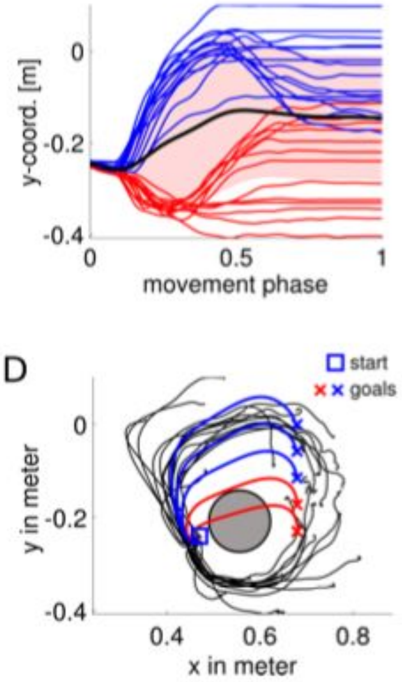
Attention: Individual dimensions (e.g. the x and y Cartesian coord.) are organized in a sequential order where first the complete dimension one is stored!

Assumption: The model captures the correlation between the multiple dimensions through the covariance matrix $\Sigma_{w|y}$. This implies a local linear relationship between the individual dimensions.

Enough Theory - Some Applications

When a single primitive is not sufficient





Rueckert, Elmar; Mundo, Jan; Paraschos, Alexandros; Peters, Jan; Neumann, Gerhard. [Extracting Low-Dimensional Control Variables for Movement Primitives.]() Proceedings of the International Conference on Robotics and Automation (ICRA), 2015.

Gaussian Mixture Model as Prior

$$p(\boldsymbol{w}^{[i]}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{w}^{[i]} \middle| \boldsymbol{b}_k + \boxed{\boldsymbol{M}_k \boldsymbol{h}_k^{[i]}}, \alpha^{-1}\boldsymbol{I}\right)$$

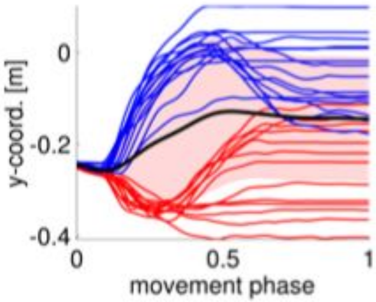
projection matrix M and low-dim. latent variables h

weights as before

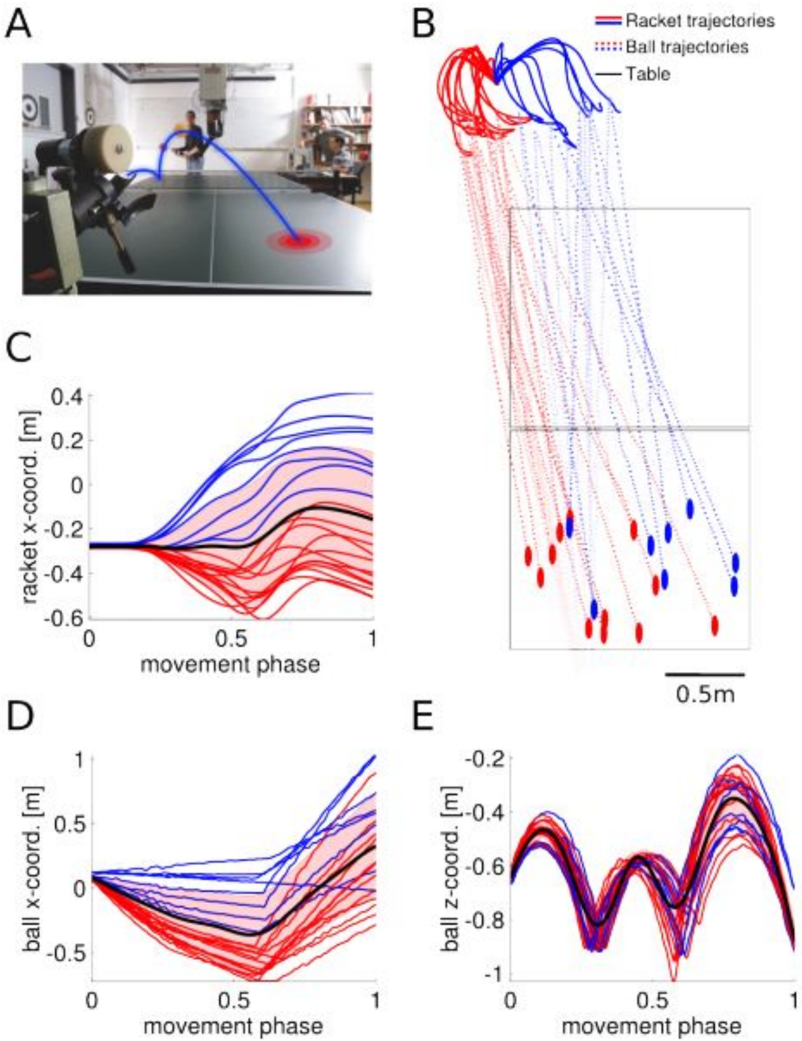
K mixture components

$$\boldsymbol{\mu}_{\boldsymbol{w}^{[i]}} = \boldsymbol{\Sigma}_{\boldsymbol{w}^{[i]}} \left(\beta \boldsymbol{\Psi}_{1:T}^{[i]\,T} \boldsymbol{y}_{1:T}^{[i]} + \sum_{k=1}^{K} \bar{\alpha}_k \mu_{z_k^{[i]}} \left(\boldsymbol{\mu}_{\boldsymbol{b}_k} + \bar{\boldsymbol{M}}_k \boldsymbol{\mu}_{\boldsymbol{h}_k^{[i]}} \right) \right),$$

$$\boldsymbol{\Sigma}_{\boldsymbol{w}^{[i]}} = \left(\beta \boldsymbol{\Psi}_{1:T}^{[i]\,T} \boldsymbol{\Psi}_{1:T}^{[i]} + \sum_{k=1}^{K} \bar{\alpha}_k \mu_{z_k^{[i]}} \boldsymbol{I} \right)^{-1}$$

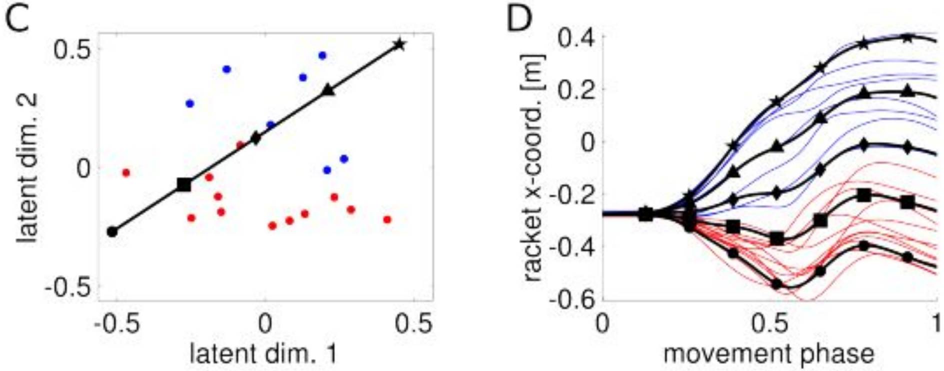


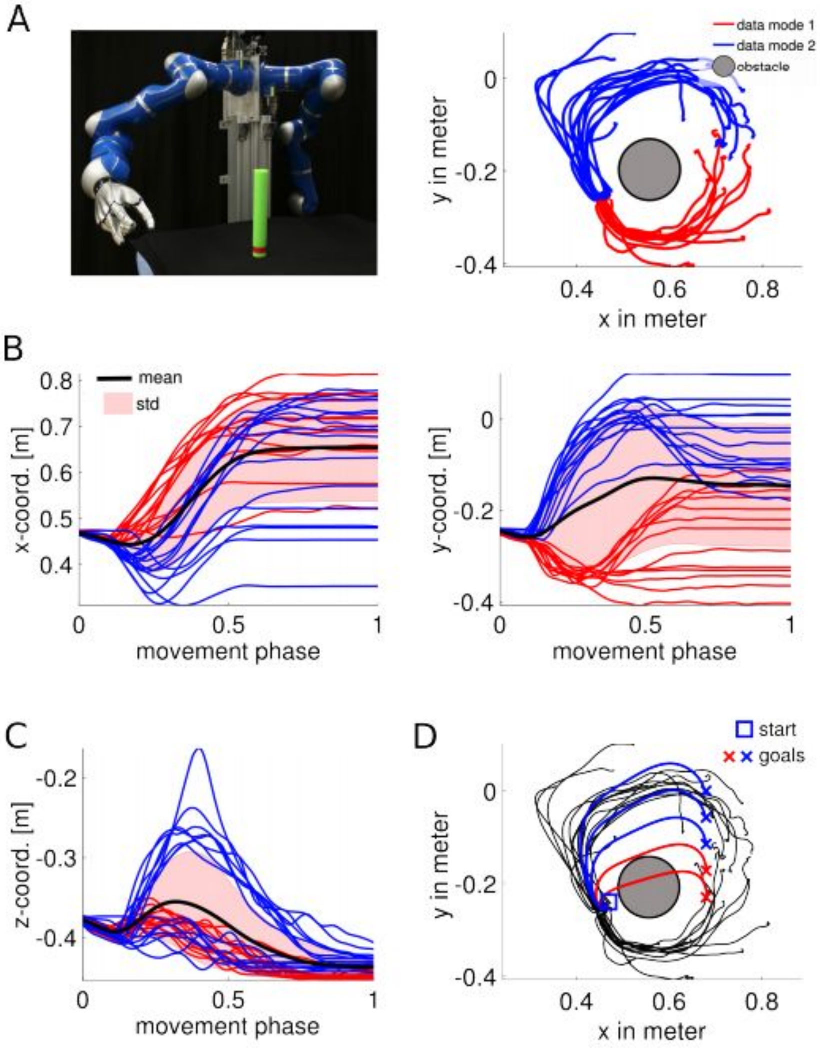
Rueckert, Elmar; Mundo, Jan; Paraschos, Alexandros; Peters, Jan; Neumann, Gerhard. [Extracting Low-Dimensional Control Variables for Movement Primitives.]() Proceedings of the International Conference on Robotics and Automation (ICRA), 2015.



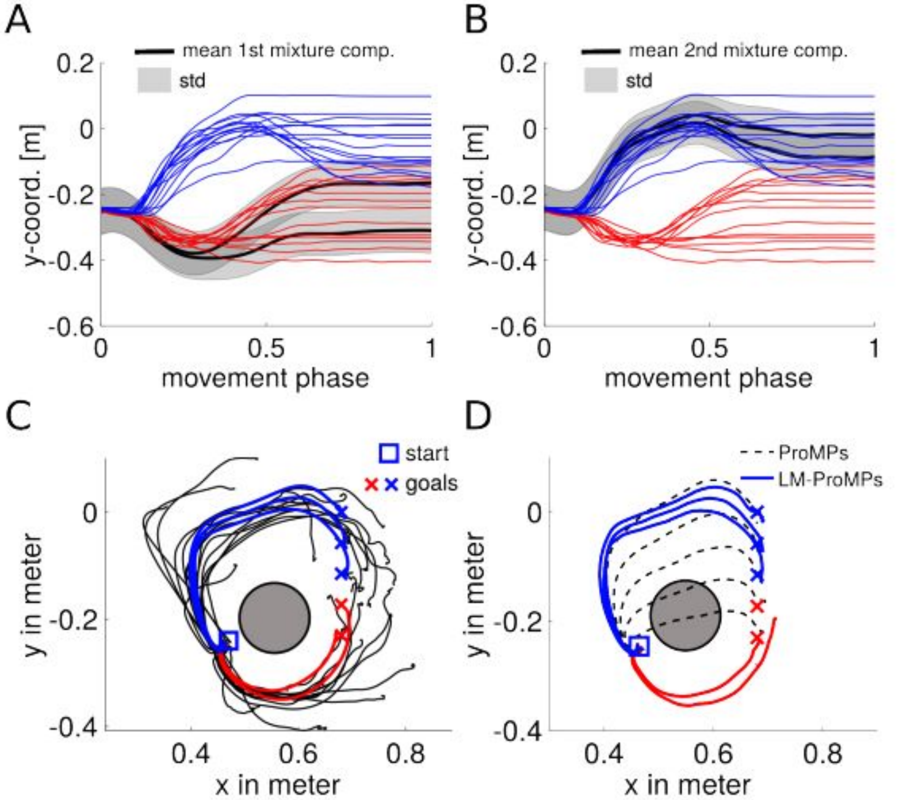
K=1 example: projection matrix M and low-dim. latent variables h

$$p(\boldsymbol{w}^{[i]}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{w}^{[i]} \middle| \boldsymbol{b}_k + \boldsymbol{M}_k \boldsymbol{h}_k^{[i]}, \alpha^{-1}\boldsymbol{I}\right)$$





K=2 example: projection matrix M and low-dim. latent variables h

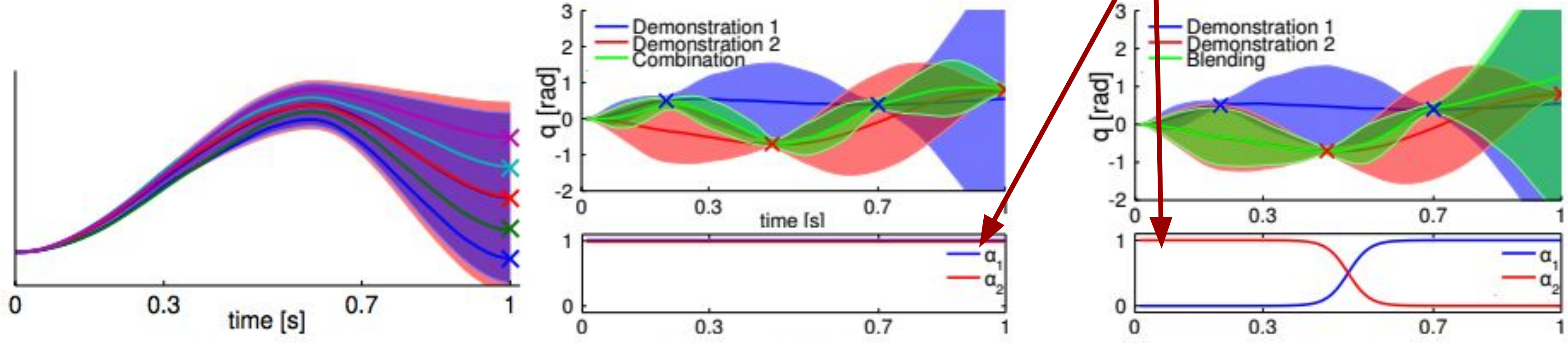


Conditioning

Combination and Blending

$$p(\boldsymbol{w}_o|\boldsymbol{o}) \propto \mathcal{N}(\boldsymbol{o}|\boldsymbol{\Phi}_o \boldsymbol{w}_o, \boldsymbol{\Sigma}_o)\, p(\boldsymbol{w})$$

$$p_{new}(\boldsymbol{\tau}) \propto \prod_i p_i(\boldsymbol{\tau})^{\alpha^{[i]}}$$

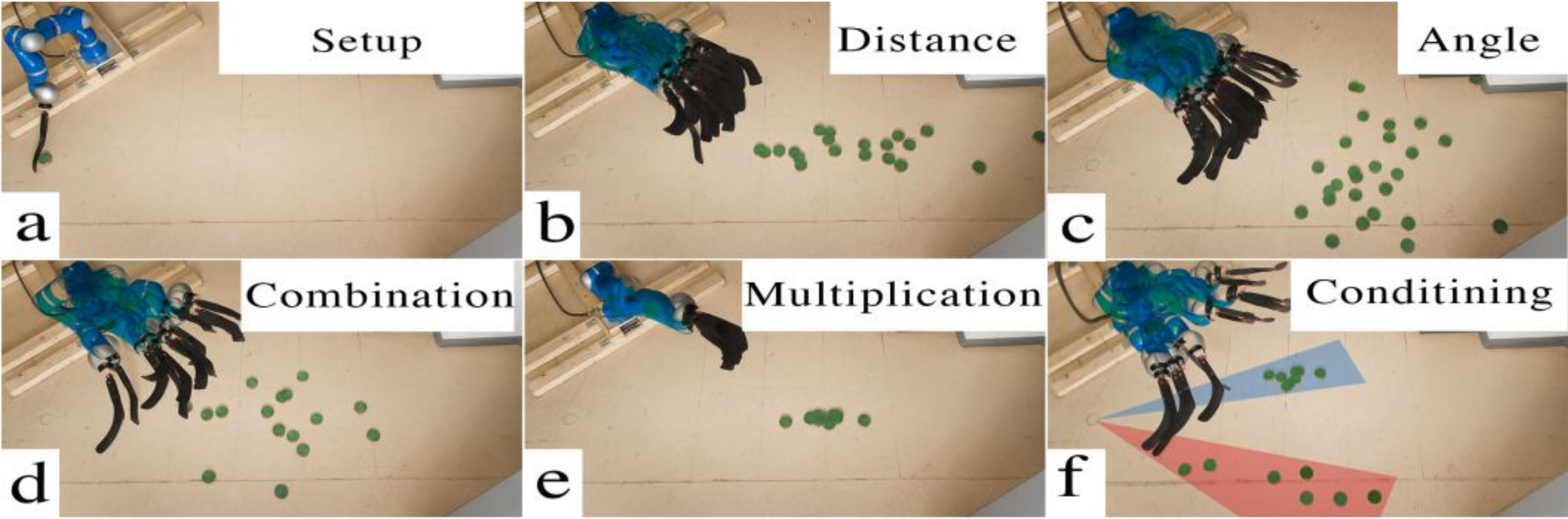


(a) Conditioning (b) Combination (c) Blending

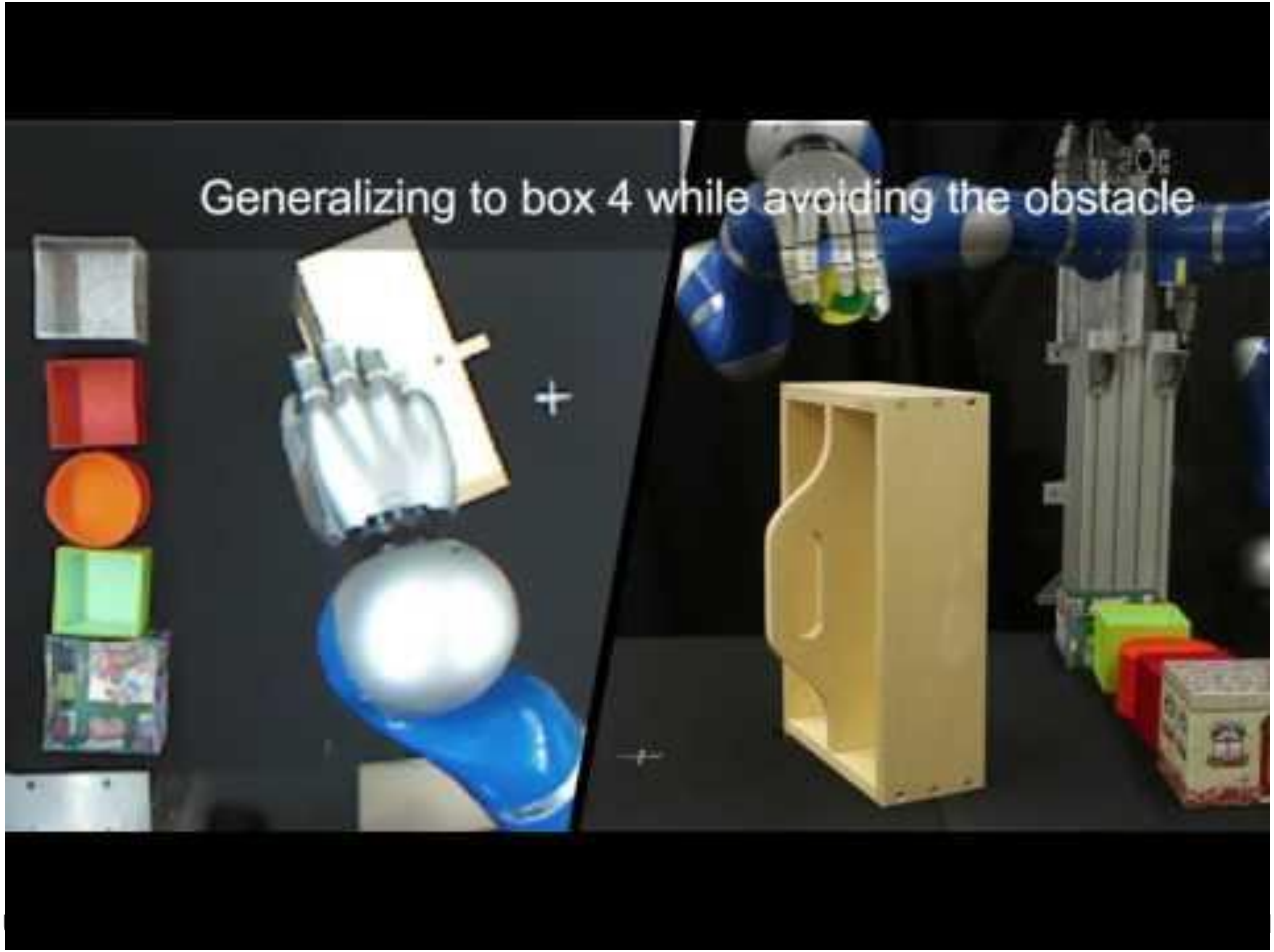
Paraschos, Alexandros; Daniel, Christian; Peters, Jan; Neumann, Gerhard. Probabilistic Movement Primitives, *Advances in Neural Information Processing Systems (NIPS)*, MIT Press, 2013.

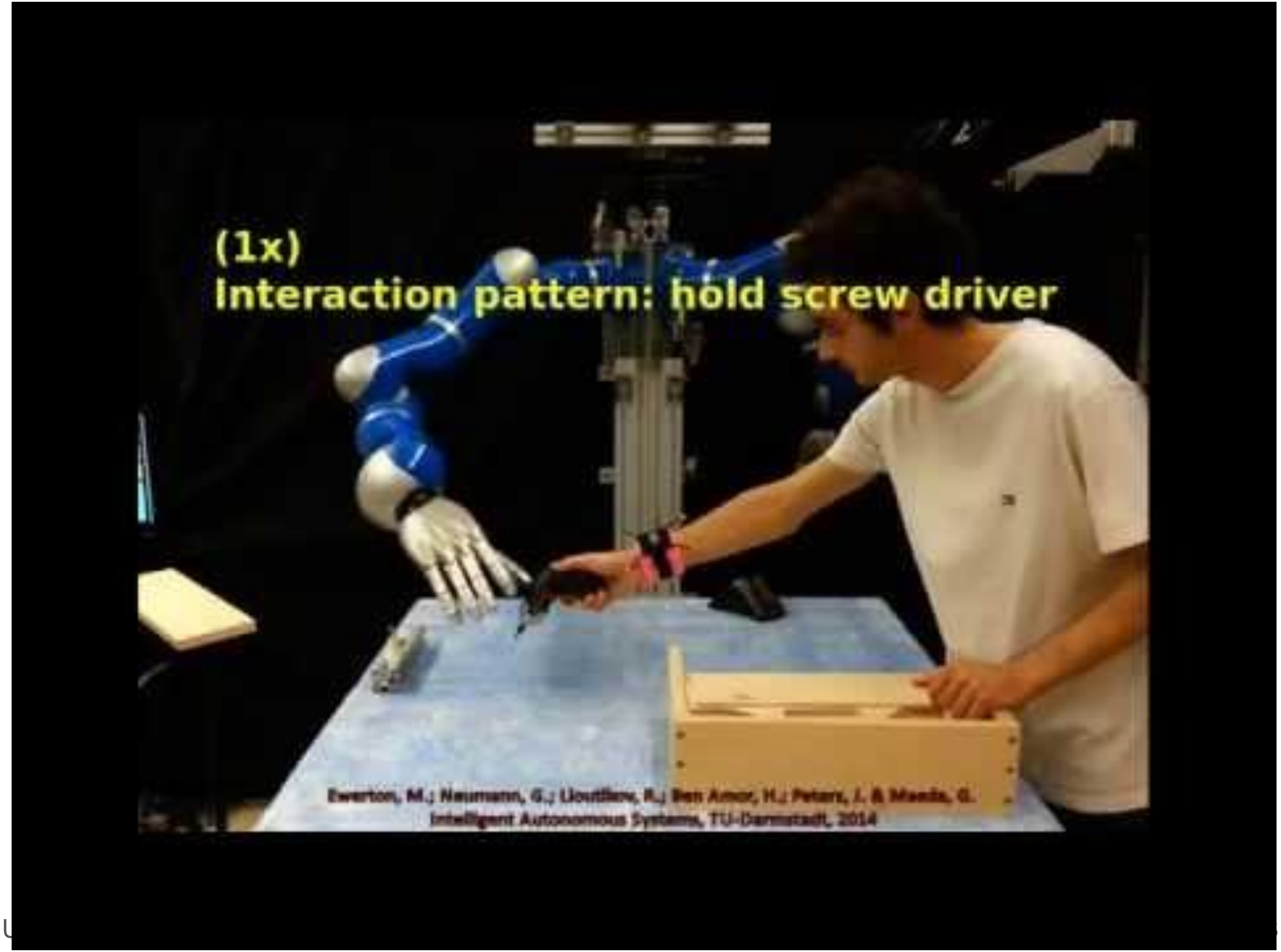
Multiplication

$$p(\underset{o}{\overset{\text{angle}}{w}} | o) \; p(\underset{o}{\overset{\text{distance}}{w}} | o)$$



Paraschos, A.; Daniel, C.; Peters, J.; Neumann, G. (2018). Using Probabilistic Movement Primitives in Robotics, *Autonomous Robots (AURO), **42**, **3**, pp.529-551*.


Generalizing to box 4 while avoiding the obstacle

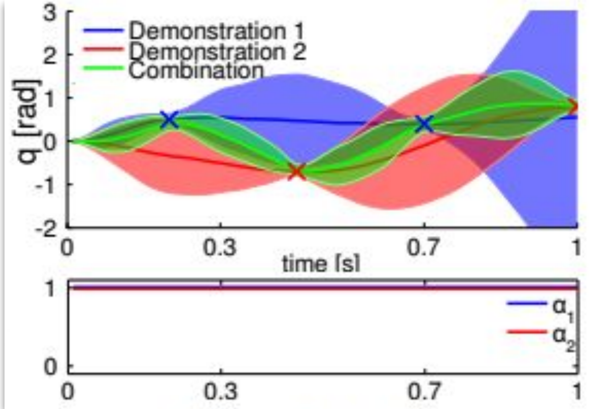

(1x)
Interaction pattern: hold screw driver
Ewerton, M.; Neumann, G.; Lioutikov, R.; Ben Amor, H.; Peters, J. & Maeda, G.
Intelligent Autonomous Systems, TU-Darmstadt, 2014


3rd requested
demonstration: screw driver

Summary of Time Series & Databases

- An introduction to Probabilistic Time Series Models
  - Single Time Step Model
  - Multi Time step Model
- Some Applications in Research
- Let’s implement Databases in Jupyter Hub

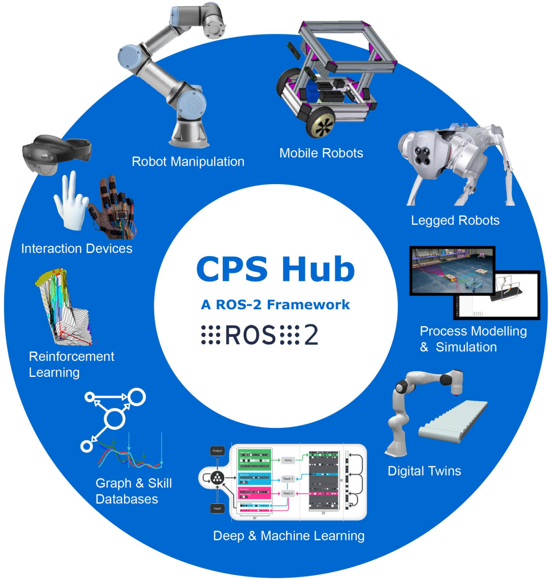
$$
\begin{aligned}
p(\boldsymbol{\tau}_j|\boldsymbol{w}) &= \prod_{t=1}^{T} p(y_{t,j}|\boldsymbol{w}) \,, \\
&= \prod_{t=1}^{T} \mathcal{N}(y_{t,j}|\boldsymbol{\phi}_t^T \boldsymbol{w}, \sigma_y^2) \,, \\
&= \mathcal{N}(\boldsymbol{\tau}_j|\boldsymbol{A}\boldsymbol{w}, \sigma_y^2 \boldsymbol{I}) \,.
\end{aligned}
$$



Thank you for your attention!

Visit our Youtube Channel:
https://youtube.com/@CPSAustria

YouTube 117



Phone: +43 3842 402 – **1901** (Sekretariat CPS)

Email: cps@unileoben.ac.at

Web: https://cps.unileoben.ac.at