Model Predictive Control and Reinforcement Learning
– Introduction –

Joschka Boedecker and Moritz Diehl

University Freiburg

July 26, 2021

Aim of this course

Understanding the main concepts of model predictive control (MPC) and reinforcement learning (RL) and their similarities and differences.

Applying the methods to practical optimal control problems in hands-on exercises and project work.

Agenda

Time Slots	Monday	Tuesday	Wednesday	Thursday	Friday
09:00-09:45	Lecture 1 - Introduction - Joschka Boedecker and Moritz Diehl	Lecture 4 - Dynamic Programming and LQR - Moritz Diehl	Lecture 7 - Numerical Optimal Control - Moritz Diehl	Lecture 10 - On-policy RL with Function Approximation - Joschka Boedecker	Microexam
10:00-10:45	Lecture 2: Dynamic Systems and Simulation (Moritz Diehl) (/files/2021ss/MPCRL/lecture-2-simulation.pdf)	Lecture 4 (continued) - Moritz Diehl	Lecture 7 (continued) - Moritz Diehl	Lecture 11 - Off-policy RL with Function Approximation - Joschka Boedecker	Lecture 14 - Robust and Stochastic MPC - Moritz Diehl
11:15-12:00	Exercise 1 - Dynamic System Simulation - Katrin Baumgärtner und Jasper Hoffmann	Exercise 3 - Dynamic Programming and LQR - Katrin Baumgärtner	Exercise 5 - Numerical Optimal Control - Katrin Baumgärtner	Exercise 7 - RL with Function Approximators - Jasper Hoffmann	Project Guidelines
14:00-14:45	Lecture 3: Numerical Optimization (Moritz Diehl) (/files/2021ss/MPCRL/lecture-3-optimization.pdf)	Lecture 5 - MDP Formalization and Monte Carlo Methods - Joschka Boedecker	Lecture 8 - MPC Stability Theory - Moritz Diehl	Lecture 12 - Policy Gradient Methods - Joschka Boedecker	Lecture 15 - Planning and Learning - Joschka Boedecker
15:00-15:45	Extended Coffee Break / Get-to-Know Each-Other-Session	Lecture 6 - Temporal Difference and Q-Learning - Joschka Boedecker	Lecture 9 - MPC Algorithms - Moritz Diehl	Lecture 13 - Advanced Value-based Methods - Joschka Boedecker	Lecture 16 - Differences and Similarities of MPC and RL - Joschka Boedecker and Moritz Diehl
16:15-17:00	Exercise 2 - Numerical Optimization - Katrin Baumgärtner	Exercise 4 - Q-Learning - Jasper Hoffmann	Exercise 6 - Model Predictive Control - Katrin Baumgärtner	Exercise 8 - Policy Gradient - Jasper Hoffmann	Project Pitch Presentations

Time Slots	Monday	Tuesday	Wednesday
09:00-09:45	Guest Lecture Sebastien Gros: Adaptation of MPC via RL: fundamental principles	Project Work	Project Work
10:00-10:45	Project Commitments	Project Status Updates	Project Presentations
11:15-12:00	Project Work Question session: Jasper Hoffmann & Katrin Baumgärtner	Project Work Question session: Jasper Hoffmann & Katrin Baumgärtner	Project Presentations
14:00-14:45	Guest Lecture Sebastien Gros: RL and MPC: safety, stability, and some more recent results	Project Work	Project Presentations
15:00-15:45	Project Work	Project Work Question Session: Jasper Hoffmann & Katrin Baumgärtner	Certificate Handout (A,B,C)
16:15-17:00	Project Work Question Session: Jasper & Katrin Baumgärtner	Guest Lecture - Sergey Levine: Model-Free and Model-Based Reinforcement Learning from Offline Data	End at 16:00

Team

Moritz Diehl
(Lecturer MPC)

Joschka Boedecker
(Lecturer RL)

Katrin Baumgärtner
(Tutor)

Jasper Hoffmann
(Tutor)

Sebastien Gros
(NTNU Trondheim, Guest)

Sergey Levine
(UC Berkeley, Guest)

Discussion

- What do you know about Model Predictive Control?
- What are characteristics of Reinforcement Learning?
- What are differences to Supervised Learning?

Characteristics of MPC & Reinforcement Learning

- Both are frameworks to solve sequential decision making problems
- Both automatically design controllers based on desired outcomes (reward / stage cost, constraints)
- Closed-loop system visits different regions of the state space than uncontrolled system

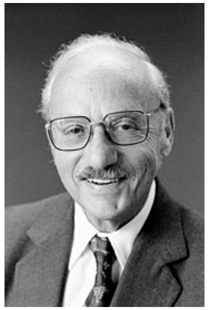
MPC

- System identification precedes control implementation, model fixed during execution
- Typically convex stage costs
- Constraints imposed explicitly
- Online optimization over prediction horizon, expensive
- Usually combined with state estimator

RL

- Controller directly learned from data, trial-and-error, exploration and exploitation trade off
- Both shaped/concave and 0-1/sparse rewards
- Constraints are imposed via penalties
- Typically parametrized controller, cheap online execution
- Usually, history included in definition of the state

MPC & Reinforcement Learning have a long history



Linear Programming … MPC

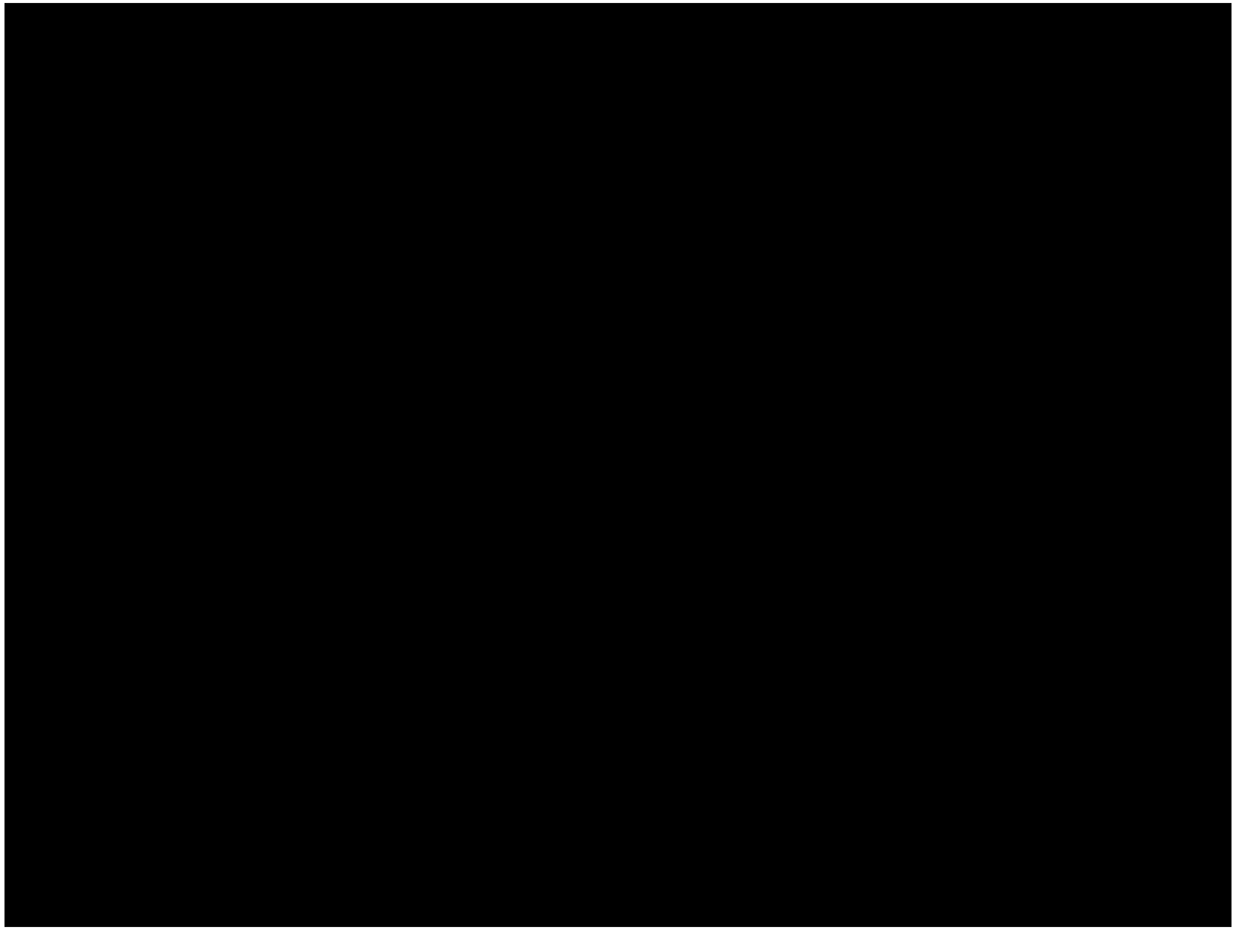
- Linear Programming (LP) developed by **G. Dantzig** in 1947
- was extended to Quadratic Programming (QP), Nonlinear Programming (NLP), Integer Programming (IP), … in field of **mathematical optimisation**
- Online solution of LP, QP, NLP, IP used for many planning problems and increasingly for industrial control problems in form of MPC

Dynamic Programming … RL

- Dynamic Programming developed by **R. Bellman** in 1950s
- was extended to approximate dynamic programming, Monte Carlo Tree Search, Q-learning, policy search … in field of **machine learning**
- Reinforcement Learning techniques are increasingly applied to solve difficult planning and decision making problems with impressive results e.g. in computer games and robotics.

Some Applications of RL

Learning to Play Atari Games from Pixel Input

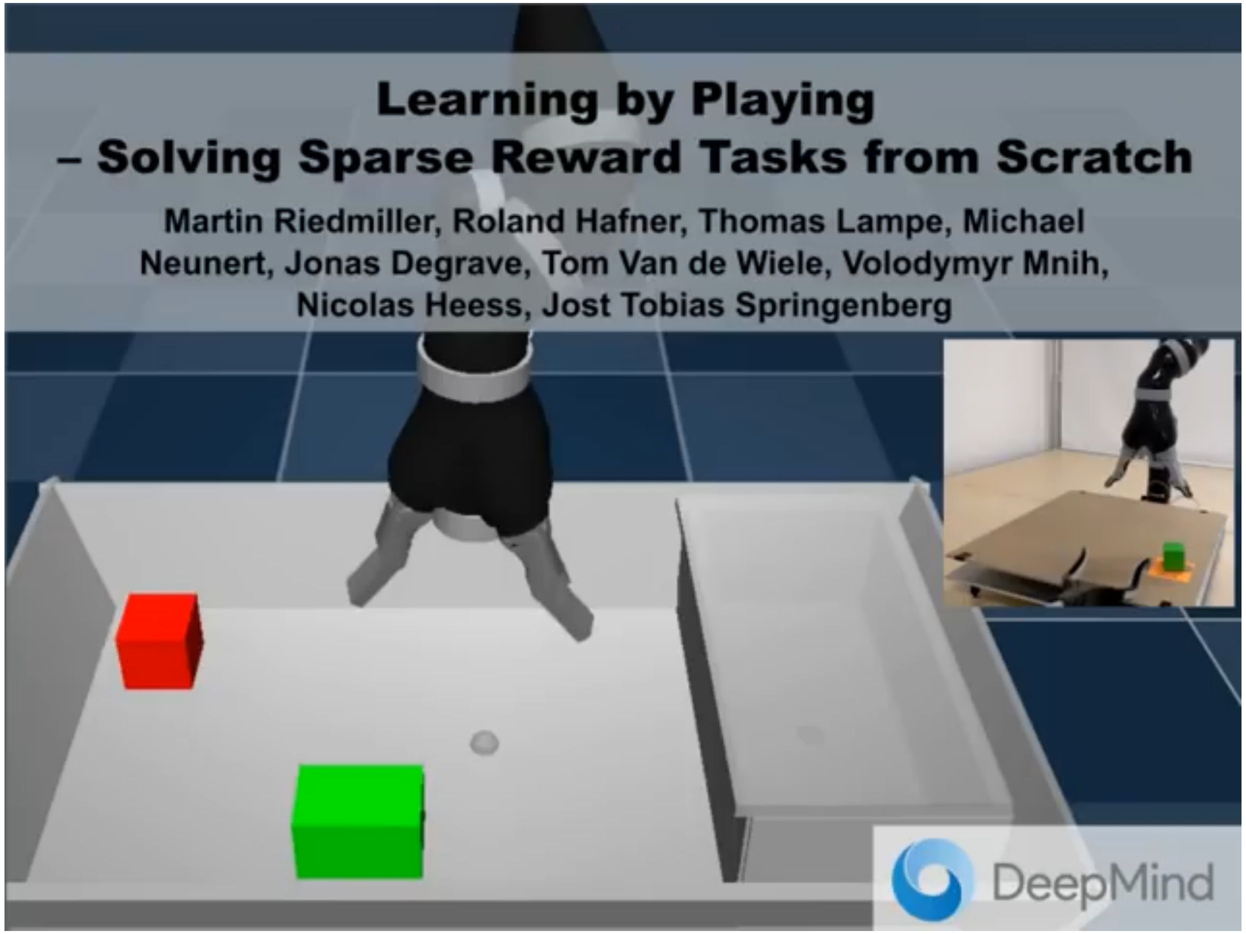


[Mnih et al., 2015]

Learning to Play the Game of Go Better Than Any Human

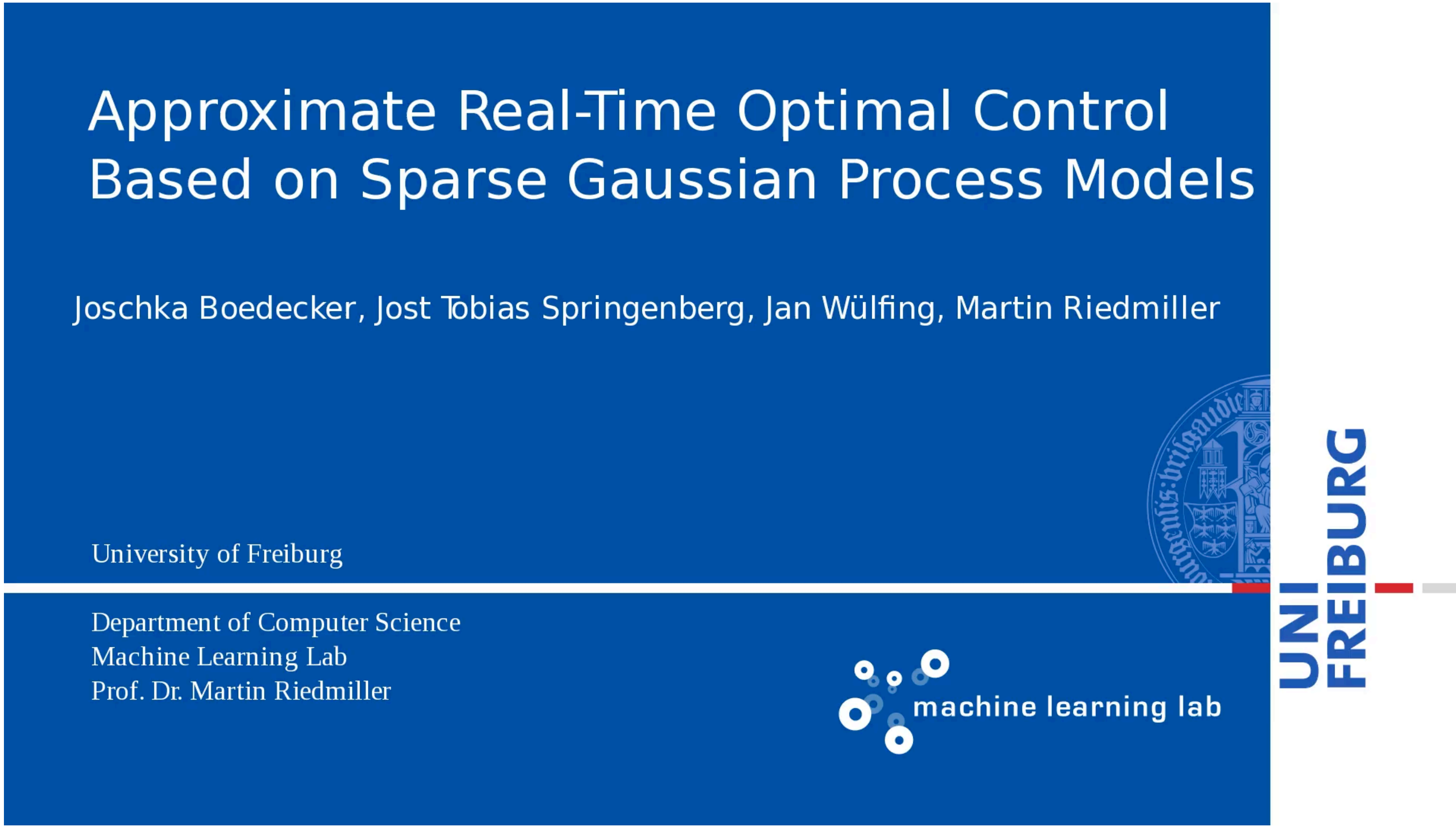
[Silver et al., 2016]

Learning Difficult Robot Manipulation Tasks from Scratch



[Riedmiller et al., 2018]

Learning to Swing Up and Balance a Pole on a Cart



[Boedecker et al., 2014]

[Kalweit et al., 2020]

Some Applications of MPC

Time-Optimal Point-To-Point Motions

Fast oscillating systems (cranes, plotters, wafer steppers, …)

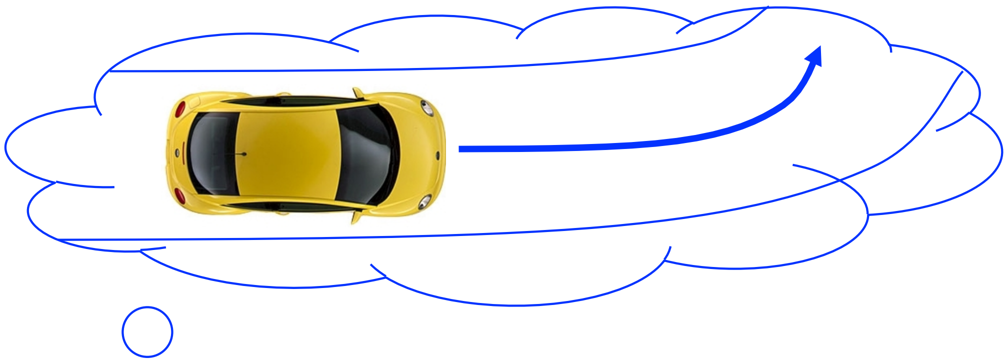
Control aims:

- reach end point as fast as possible
- do not violate constraints
- no residual vibrations

Idea: formulate as embedded optimization problem
in form of Model Predictive Control (MPC)

Model Predictive Control (MPC)

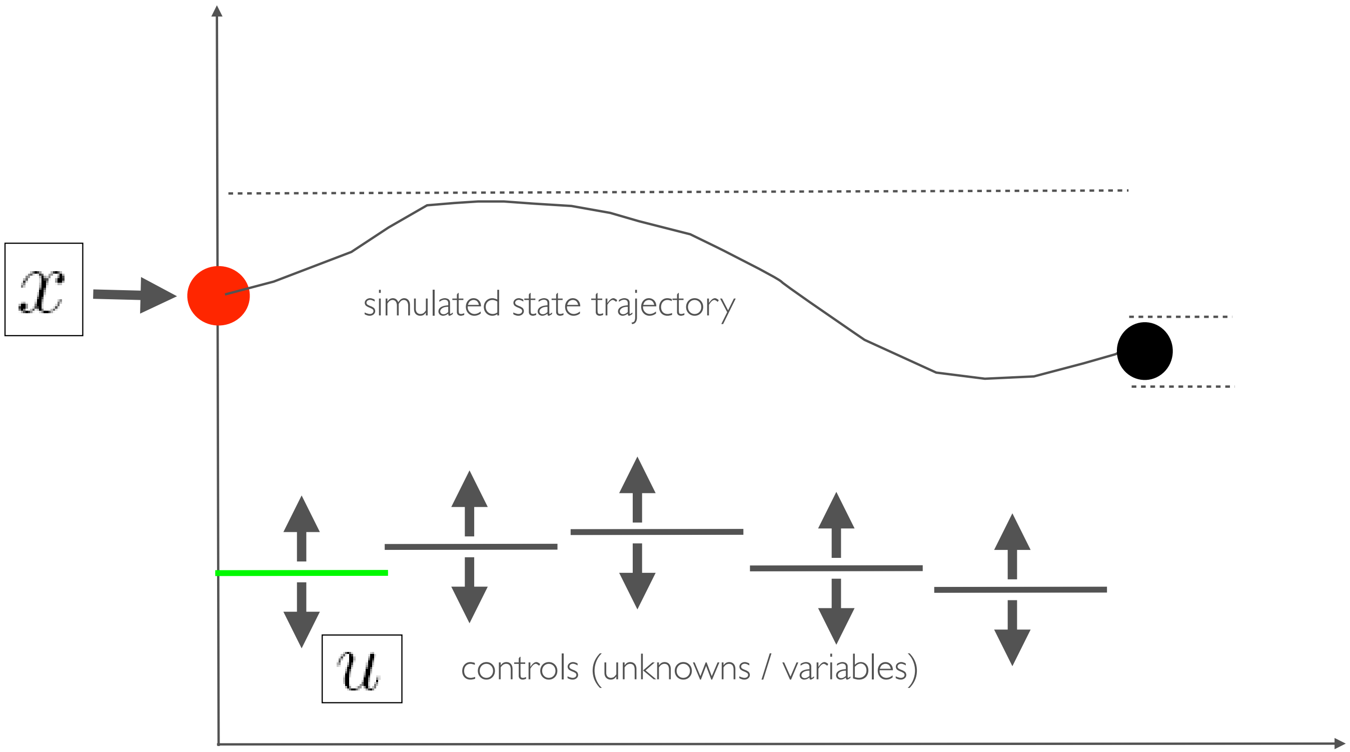
Always look a bit into the future



Example: driver predicts and optimizes,
and therefore slows down before a curve

Optimal Control Problem in MPC

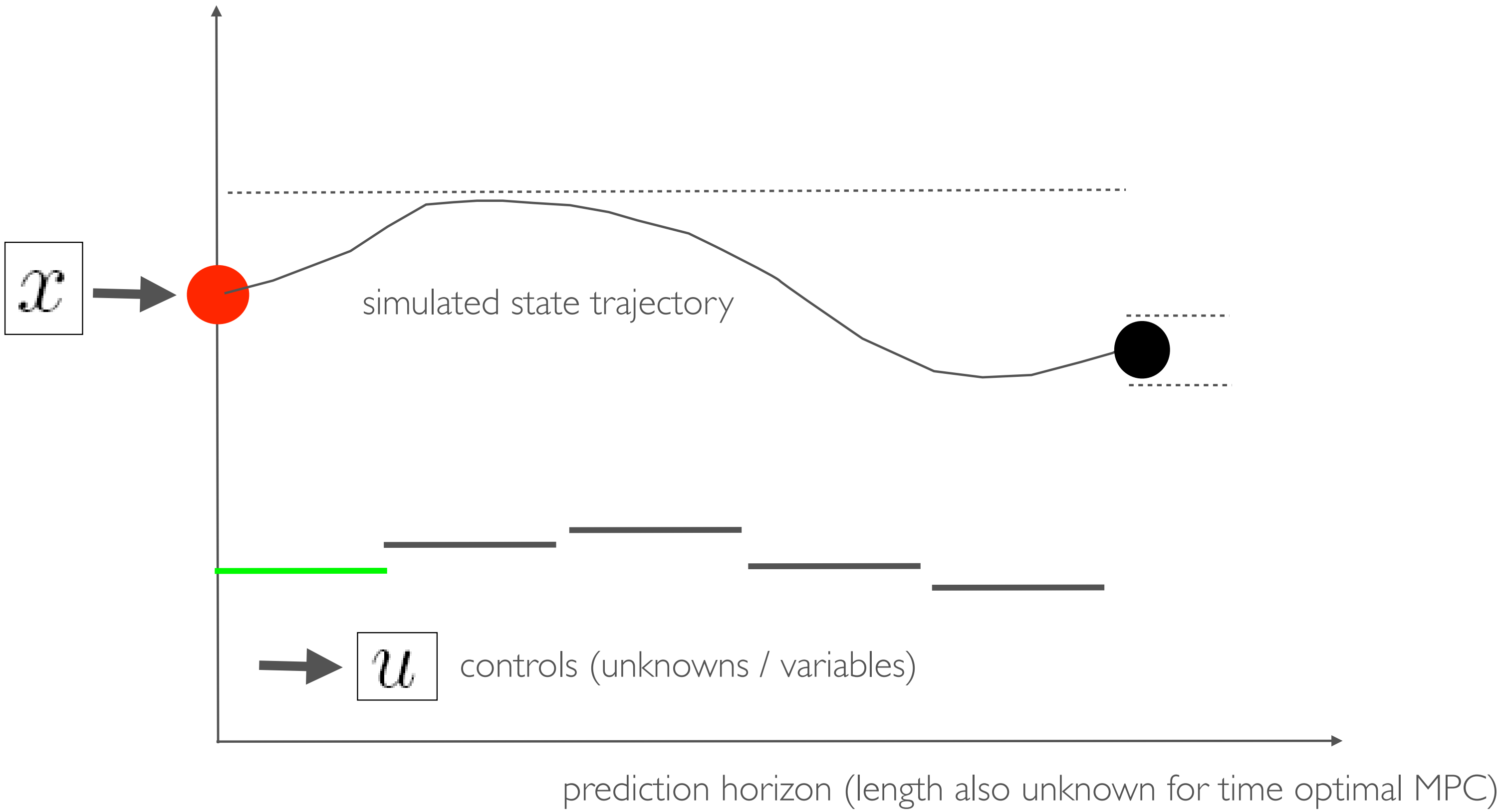
For given system state $\boldsymbol{x}$, which controls $\boldsymbol{u}$ lead to the best objective value without violation of constraints ?



prediction horizon (length also unknown for time optimal MPC)

Optimal Control Problem in MPC

For given system state $\boldsymbol{x}$, which controls $\boldsymbol{u}$ lead to the best objective value without violation of constraints ?



Time Optimal MPC of a Crane

SENSORS

- line angle
- cart position

MPC

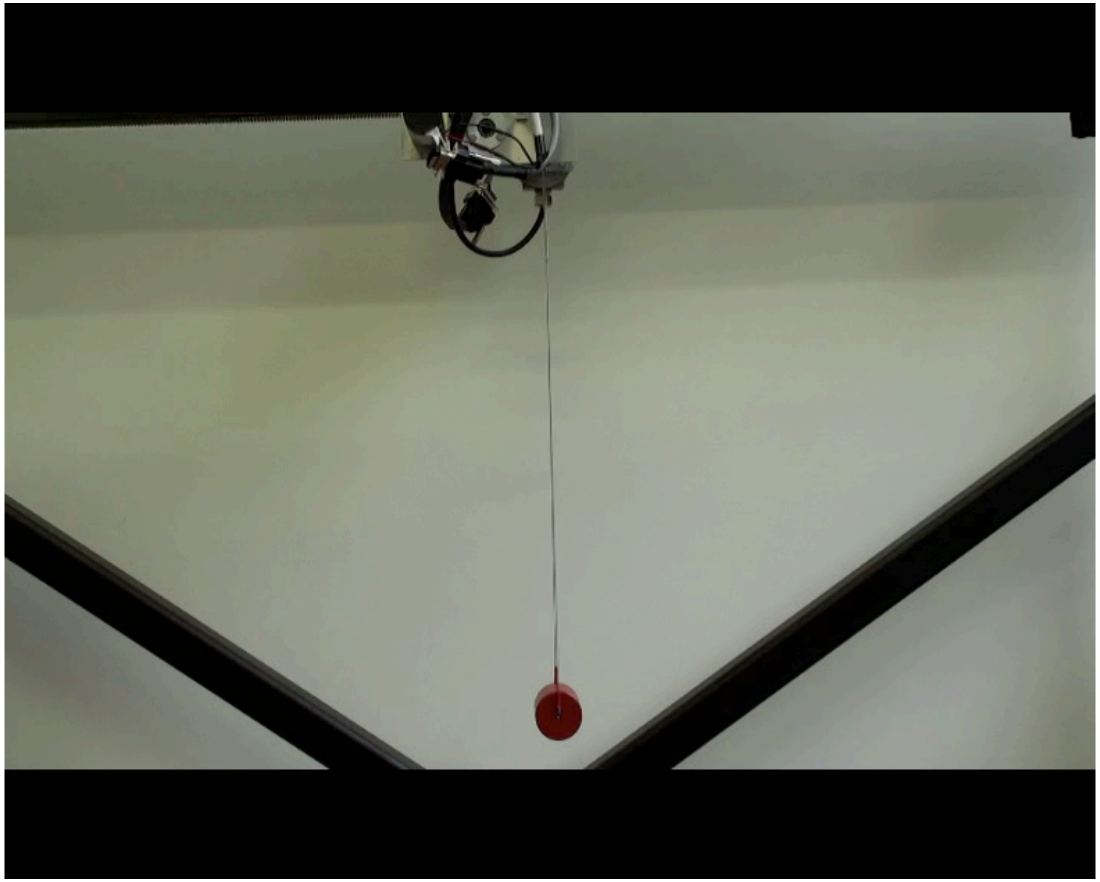
ACTUATOR

- cart motor

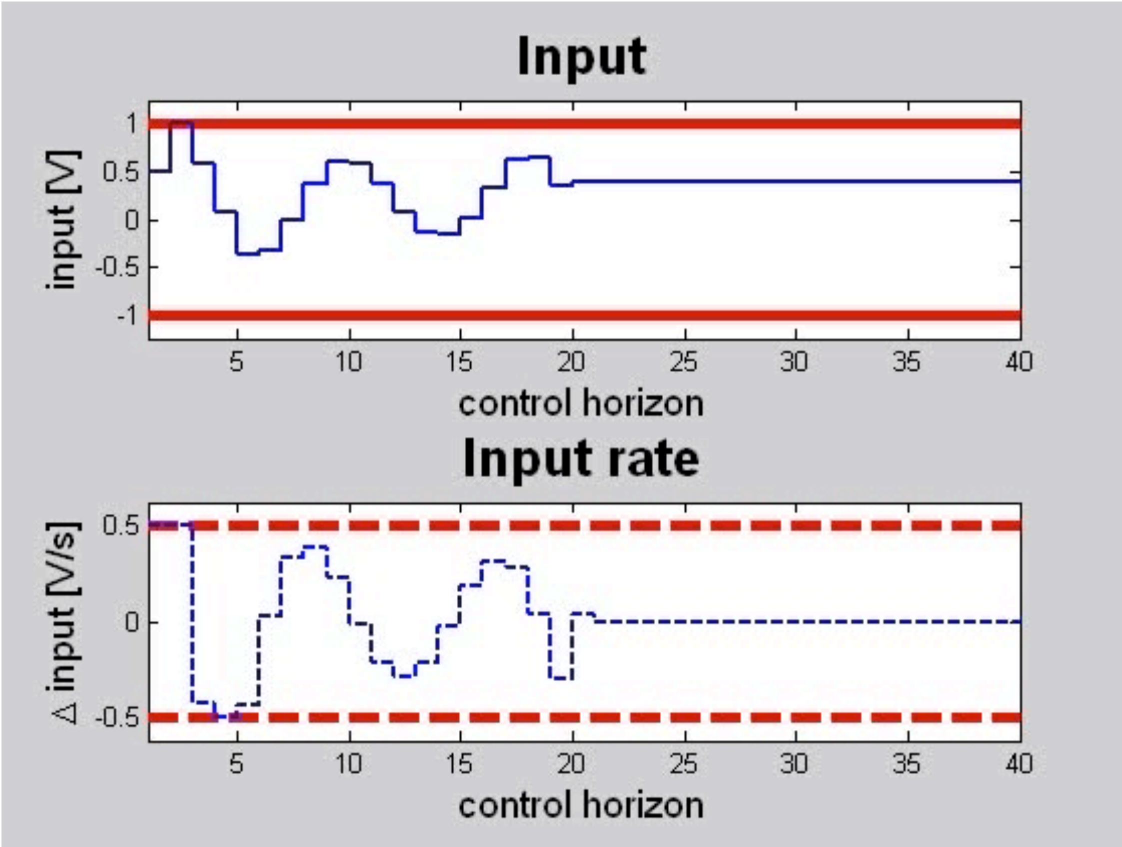
Hardware: xPC Target. Software: qpOASES [Ferreau, D., Bock, 2008]

Time Optimal MPC of a Crane

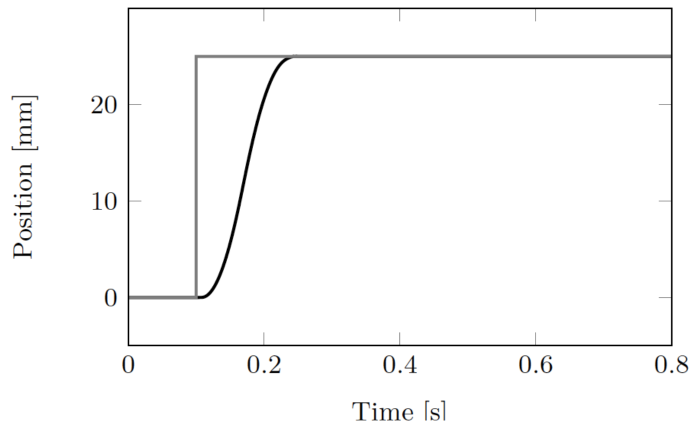
Univ. Leuven [Vandenbrouck, Swevers, D.]



Optimal Solutions in qpOASES Varying in Time

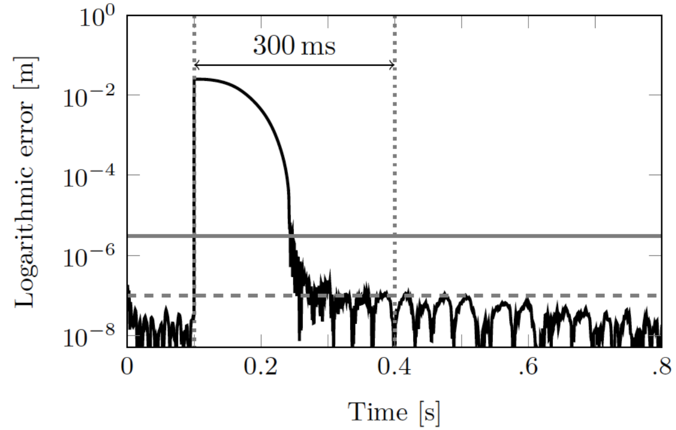


Time Optimal MPC in Industry: 25cm step, 100nm accuracy



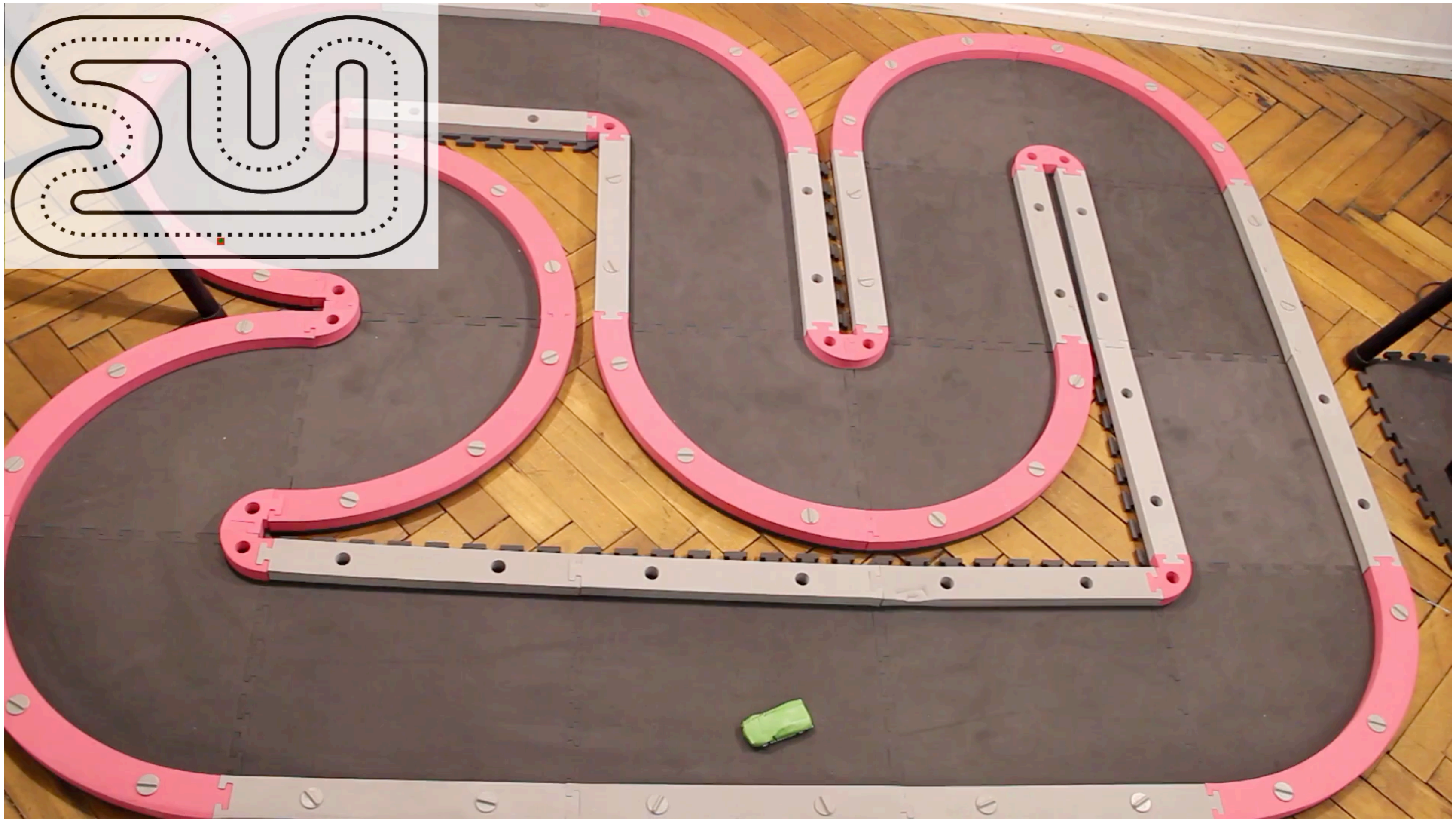
TOMPC at 250 Hz (+PID with 12 kHz)

Lieboud's results after 1 week at ETEL:
- 25 cm step in 300 ms
- 100 nm accuracy



equivalent to: „fly 2,5 km with MACH15, stop with 1 mm position accuracy"

Model Predictive Control of the Freiburg Race Cars



acados coupled into ROS, optimization every 10ms

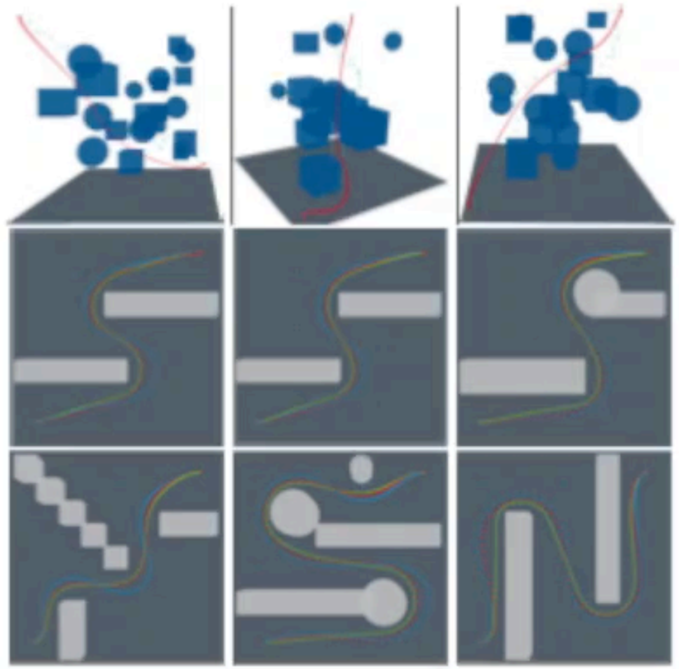
[Kloeser et al., submitted]

Safe Motion Planning at Bosch via the Convex Inner Approximation Method [Schöls et al, 2020]

An NMPC Approach using Convex Inner Approximations for Online Motion Planning with Guaranteed Collision Freedom

Tobias Schoels[1,2], Luigi Palmieri[2], Kai O. Arras[2], and Moritz Diehl[1]

Abstract— **Even though mobile robots have been around for decades, trajectory optimization and continuous time collision avoidance remains subject of active research. Existing methods trade off between path quality, computational complexity, and kinodynamic feasibility. This work approaches the problem using a model predictive control (MPC) framework, that is based on a novel convex inner approximation of the collision avoidance constraint. The proposed Convex Inner ApprOximation (CIAO) method finds a dynamically feasible and collision free trajectory in few iterations, typically one, and preserves feasibility during further iterations. CIAO scales to high-dimensional systems, is computationally efficient, and guarantees both kinodynamic feasibility and continuous-time collision avoidance. Our experimental evaluation shows that the approach outperforms state of the art baselines in terms of planning efficiency and path quality. Furthermore real-world experiments show its capability of unifying trajectory optimization and tracking for safe motion planning in dynamic environments.**

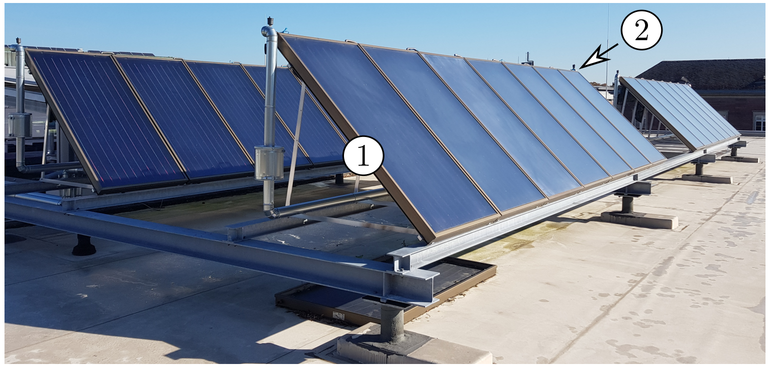


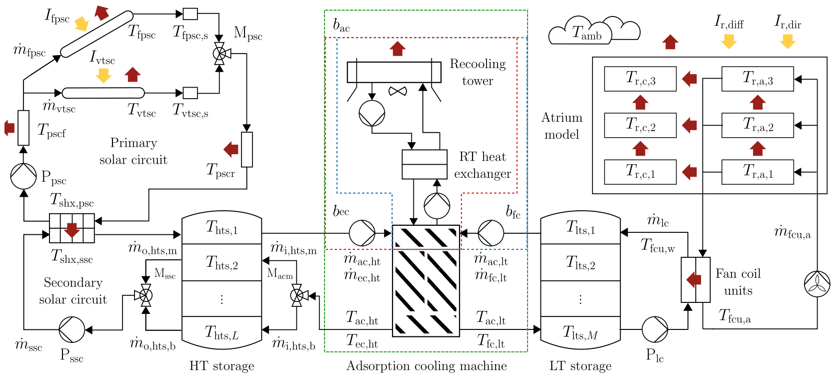
I. INTRODUCTION

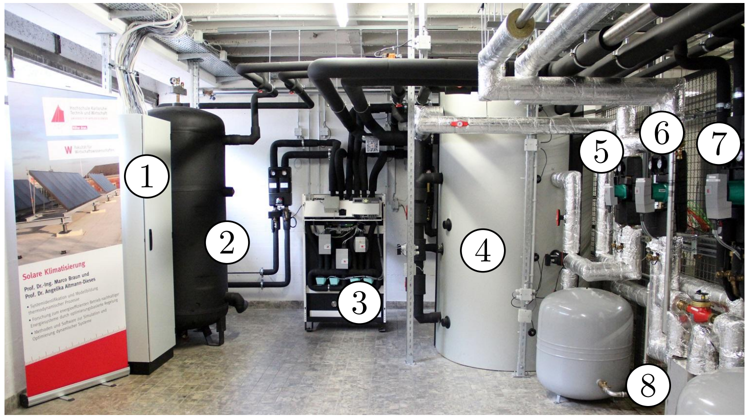
Several existing mobile robotics applications (e.g. intra-logistic and service robotics) require robots to operate in dynamic environments among other agents, such as humans or other autonomous systems. In these scenarios, the reactive

Nonlinear Mixed-Integer Control of a Solar Adsorptive Cooling Machine

[Bürger et al., 2019]



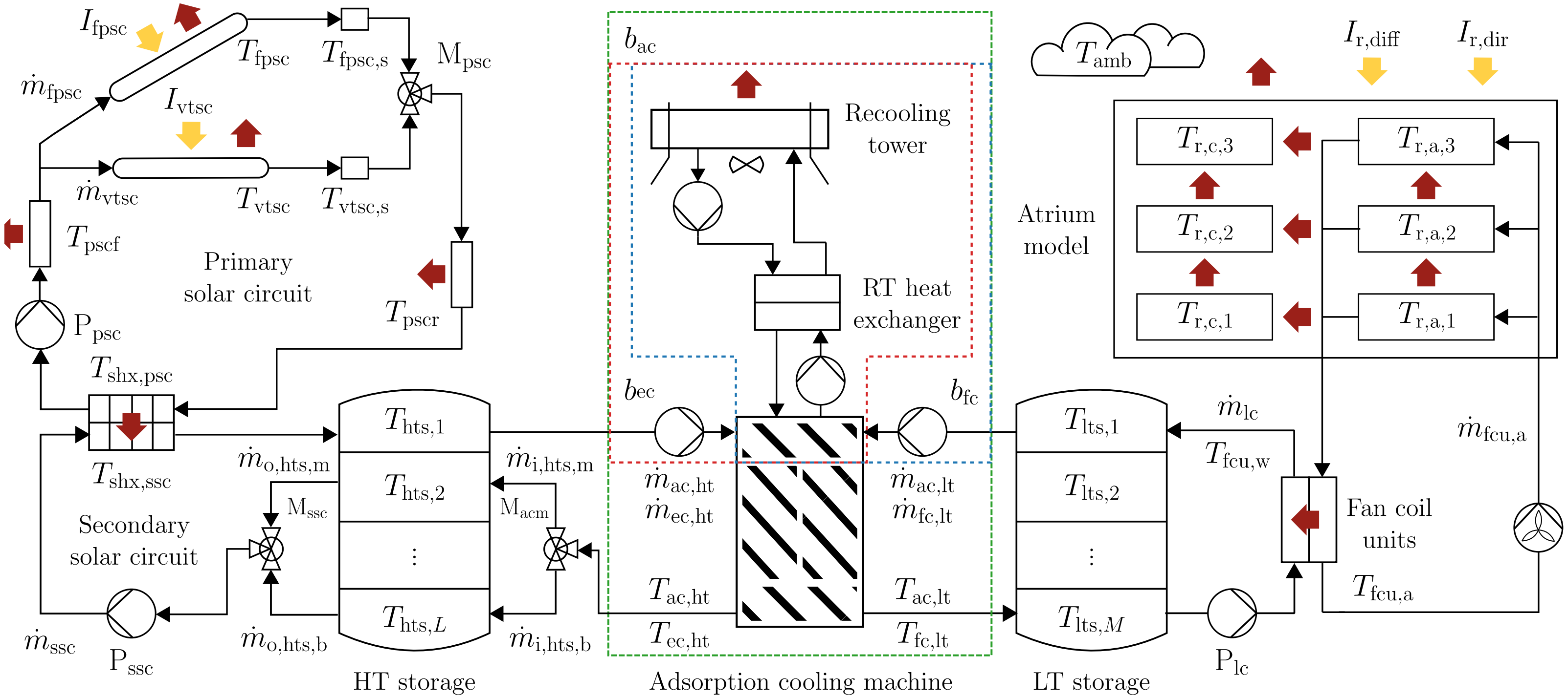




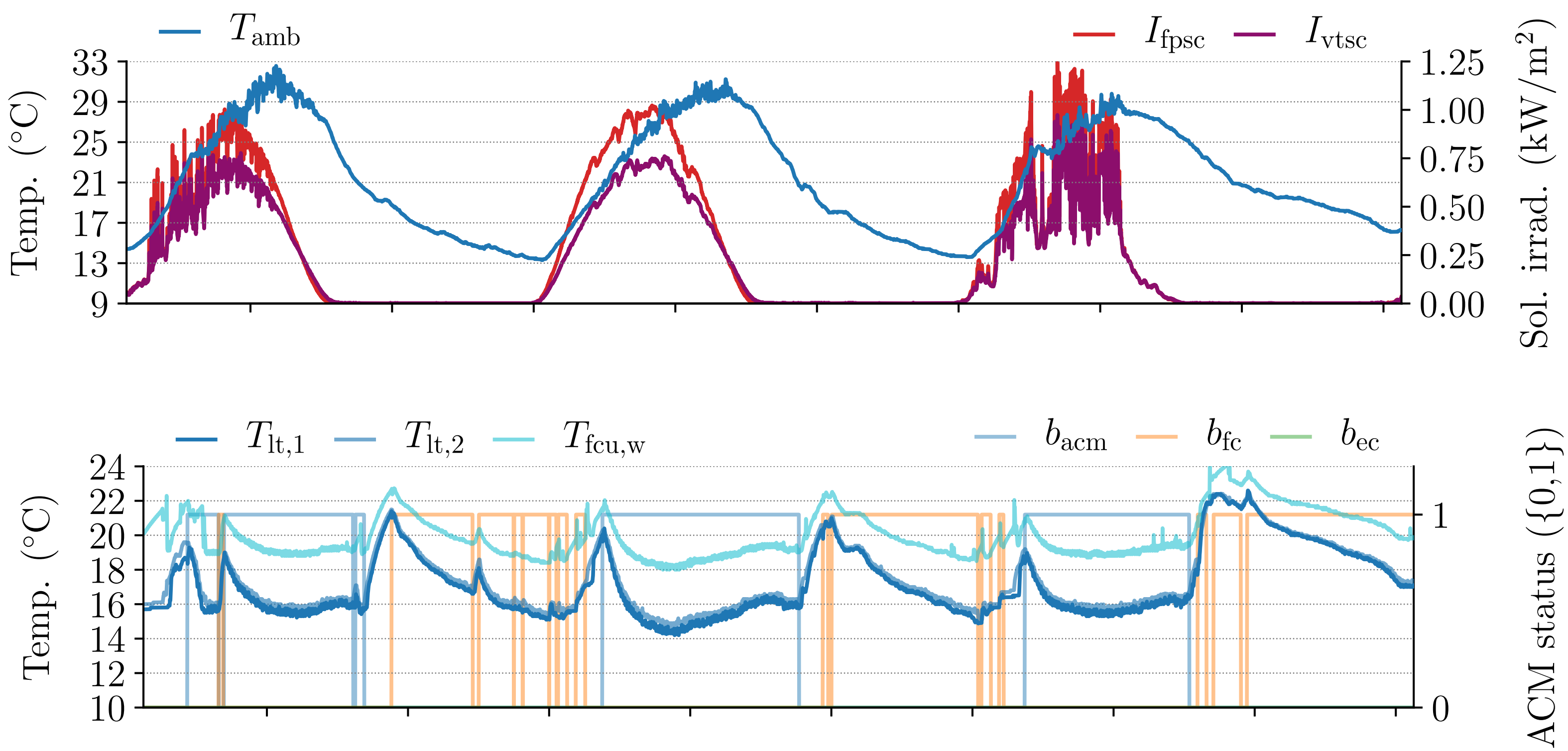
After discretisation of PDE components, obtain nonlinear ODE with 39 states, 6 continuous and 2 binary inputs.

Predict 24 hours. Aim: minimise electricity consumption.

Model Overview

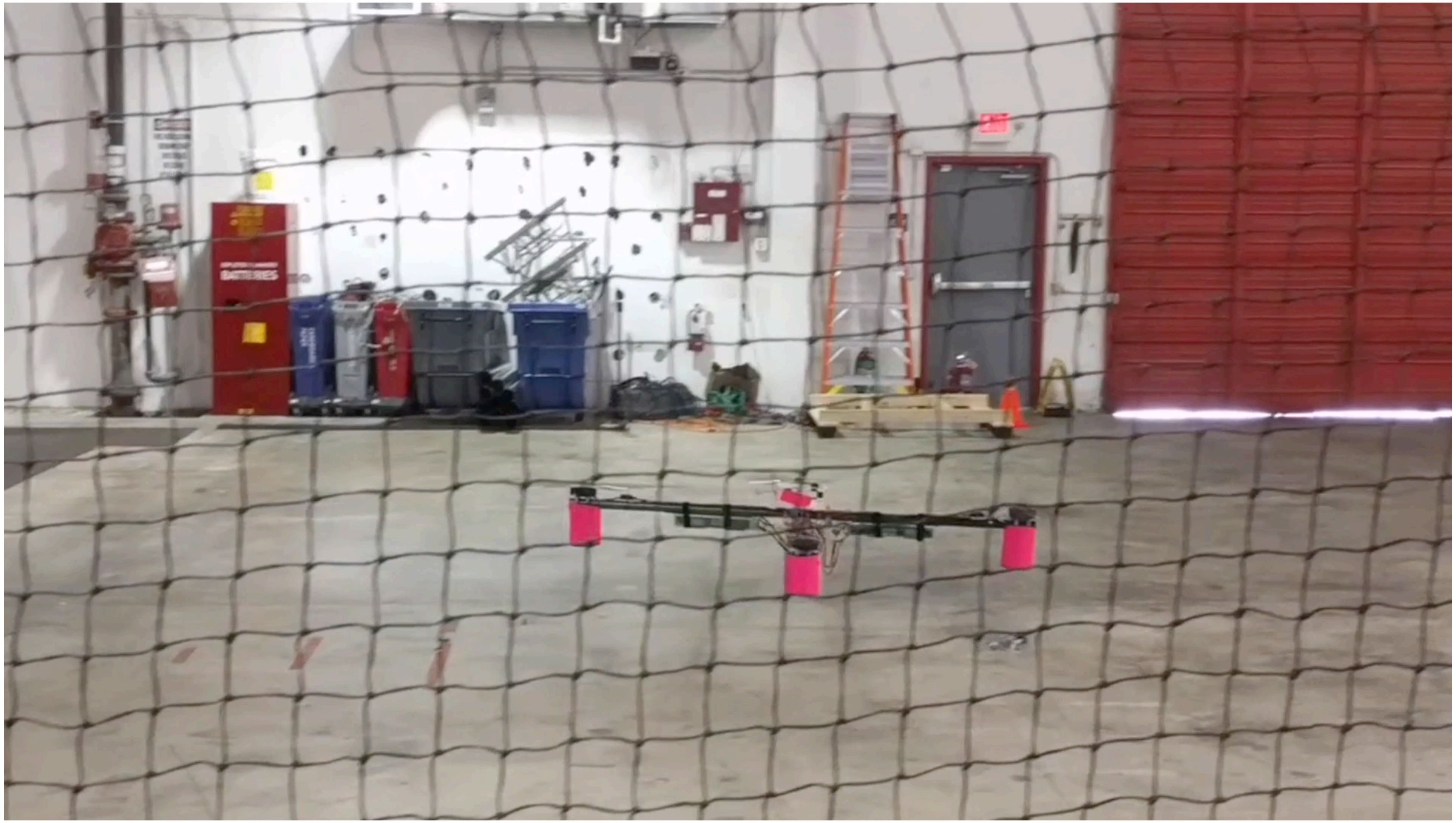


Experimental MPC Results from Sept 14-17, 2019



Every 2 minutes, a new optimization problem is solved, using a real-time algorithm based on CasADi, IPOPT [Wächter and Biegler 2006], and Pycombina [Bürger et al, 2019], an implementation of the combinatorial integral approximation (CIA) method [Sager 2009].

Human sized quadcopter control (Nonlinear MPC)
at Kitty Hawk, California, using acados



[Zanelli, Horn, Frison, D., 2018]

Electrical Compressor Control at ABB (Norway)

- work of Dr. Joachim Ferreau and Dr. Thomas Besselmann, ABB
- nonlinear MPC with qpOASES and ACADO, 1ms sampling time
- first tests at 48 MW Drive
- currently, 15% of Norwegian Gas Exports are controlled by Nonlinear MPC

Joachim Ferreau (email from 7.3.2016):

The NMPC installations in Norway (actually 5 compressors at two different sites) are doing fine since last autumn – roughly 80 billion NMPC instances solved by now. In addition, they have proven to work as expected when handling external voltage dips.

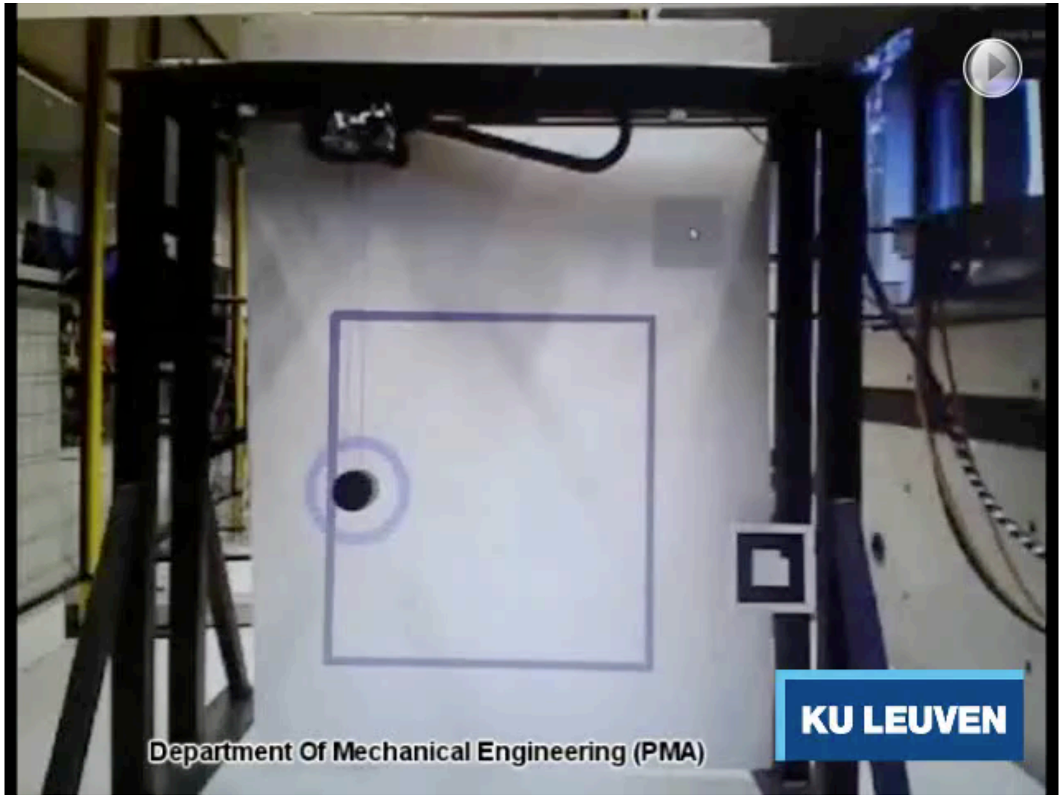
eco4wind: MPC for wind turbine control

Industrial partners: IAV, SENVION

Nonlinear MPC with about 40 states based on ACADO code generation with QP solver HPIPM running on industrial hardware at IAV

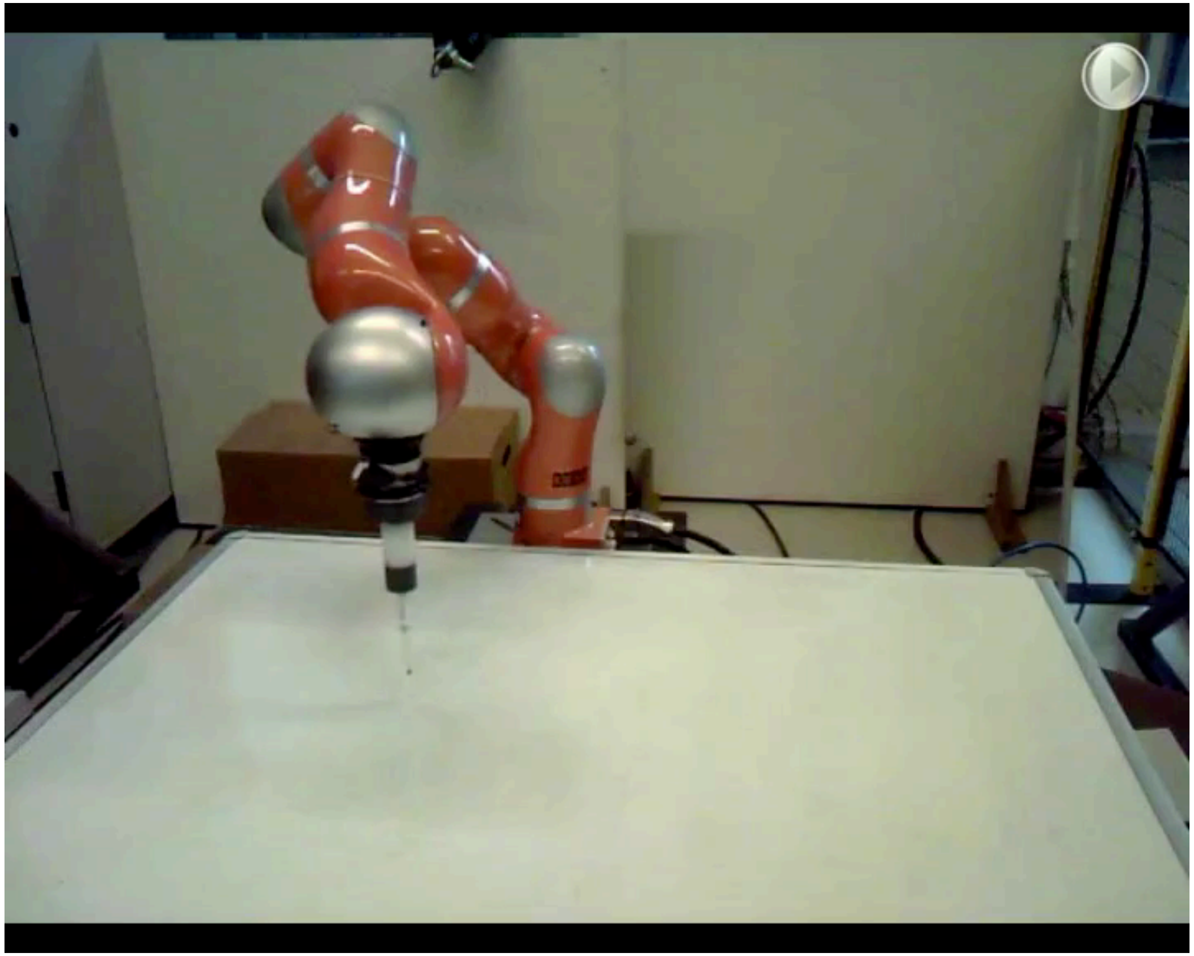
Time Optimal “drawing” by crane

Univ. Leuven [Wannes Van Loock et al.,] (CasADi)

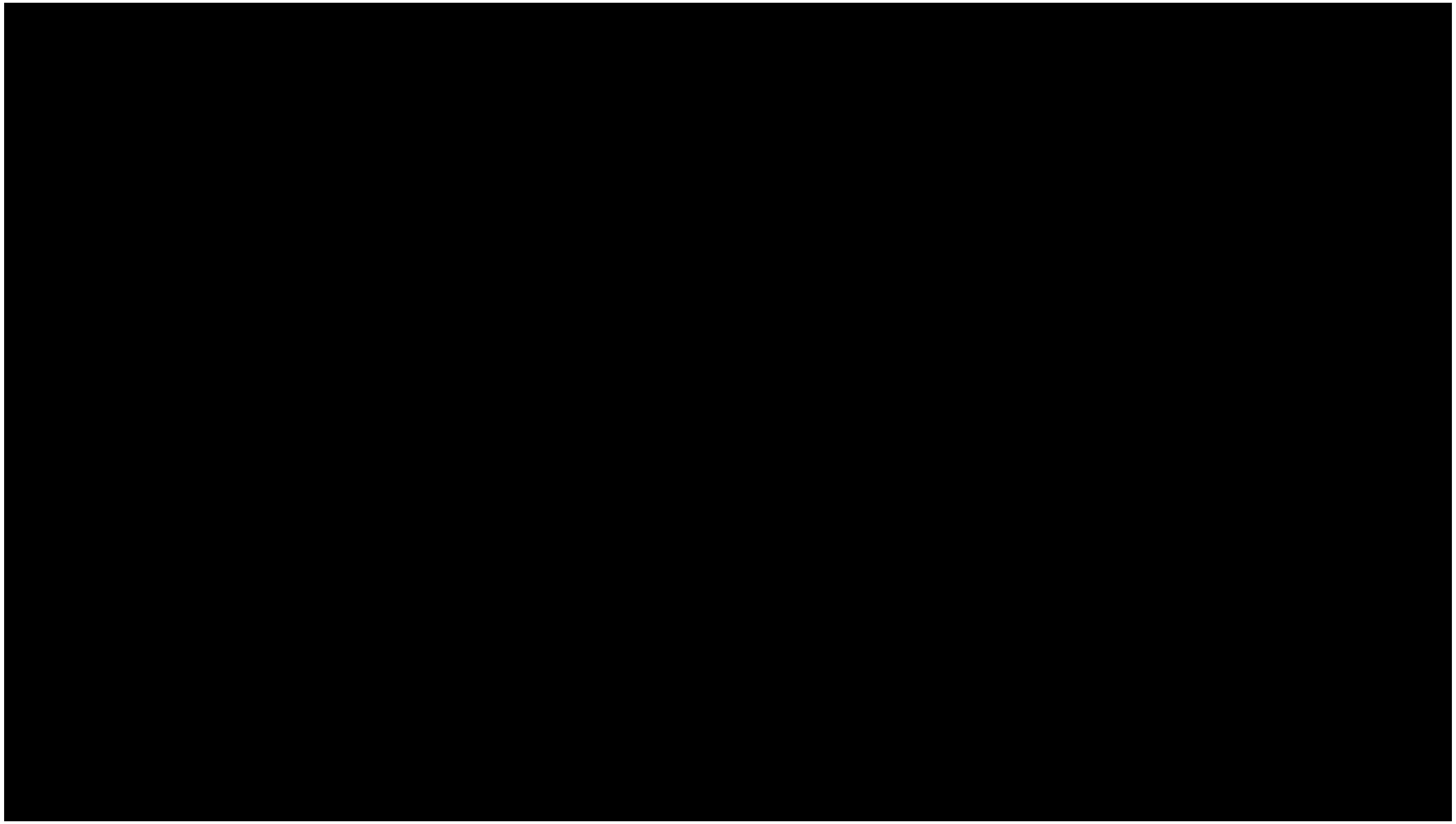


Time-optimal "hand writing" by robot

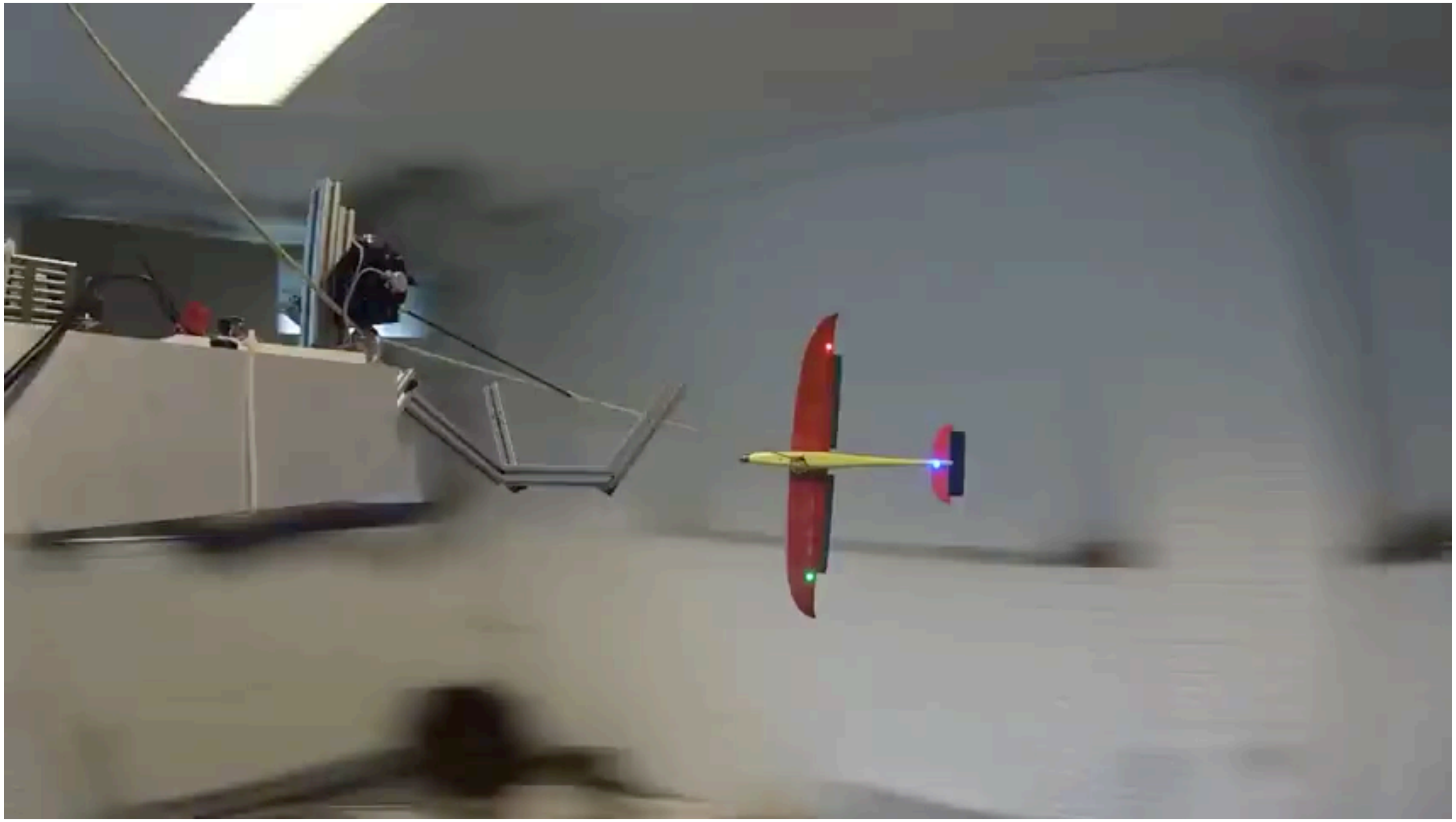
Univ. Leuven [Debrouwere, Swevers] using [Verscheure et al, IEEE TAC 2009]



Predictive control of flight carousel (in Freiburg)



Flight carousel (in Leuven, by M. Vukov)



Nonlinear MPC and Moving Horizon Estimation (MHE)

Closed loop experiments with NMPC & NMHE

KU LEUVEN

UNI FREIBURG

ERC HIGHWIND

4 kHz Nonlinear Model Predictive Control for RSM

IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, VOL -, NO. -, - 1

Continuous Control Set Nonlinear Model Predictive Control of Reluctance Synchronous Machines

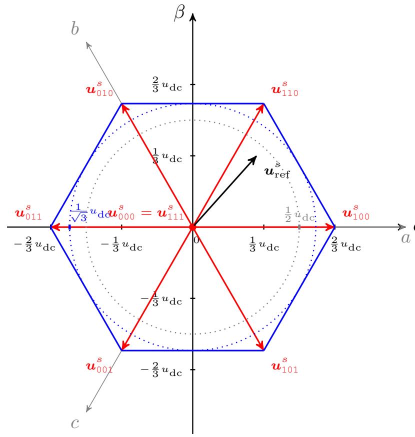
Andrea Zanelli, Julian Kullick, Hisham Eldeeb, Gianluca Frison, Christoph Hackl, Moritz Diehl

Aim:

- reliably control torque of reluctance synchronous machine (RSM) at all reachable speeds
- track flux setpoints corresponding to maximum-torque-per-ampere (MTPA)
- respect circular voltage constraints in (d,q)-frame (inscribing hexagon)

Model Predictive Control (MPC) setup:

- use two-stage voltage source inverter in order to convert from (d,q)-frame
- predict 3.2ms with nonlinear differential equation model
- penalise least squares tracking error
- use open-source software `acados` on dSPACE
- every 0.25ms, solve one MPC optimisation problem, i.e., sample at 4kHz



RSM: Control Oriented Differential Algebraic Equation Model

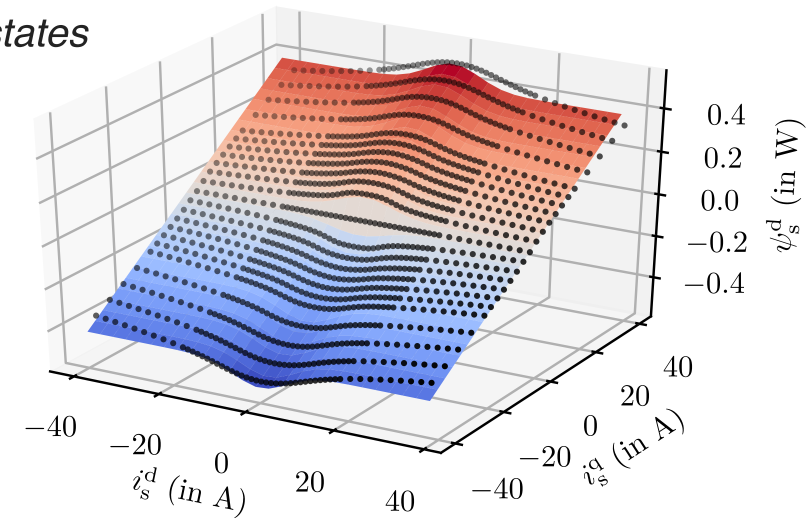
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\psi_{\mathrm{s}} &= u_{\mathrm{s}} - R_{\mathrm{s}} i_{\mathrm{s}} - \omega J \psi_{\mathrm{s}} + v, \\ 0 &= \psi_{\mathrm{s}} - \Psi_{\mathrm{s}}(i_{\mathrm{s}}), \end{aligned}$$

- differential algebraic equation (DAE)
- currents i_{s} as implicitly defined *algebraic states*
- analytical flux map approximations:

$$\Psi_{\mathrm{s}}^{\mathrm{q}}(i_{\mathrm{s}}^{\mathrm{d}}, i_{\mathrm{s}}^{\mathrm{q}}, \theta_{\mathrm{q}}) = \frac{c_0^{\mathrm{q}}}{\sqrt{2\pi\sigma_{\mathrm{d}}^2}} \exp\left(-\gamma\left(i_{\mathrm{s}}^{\mathrm{d}}, \sigma_{\mathrm{d}}\right)\right) \mathrm{atan}(c_1^{\mathrm{q}}\, i_{\mathrm{s}}^{\mathrm{q}}) + c_2^{\mathrm{q}}\, i_{\mathrm{s}}^{\mathrm{q}},$$

ι

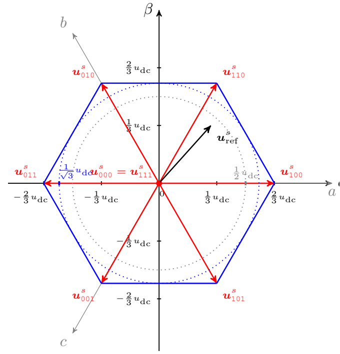
$$\gamma(x, y) := \frac{1}{2}\left(\frac{x}{y}\right)^2$$



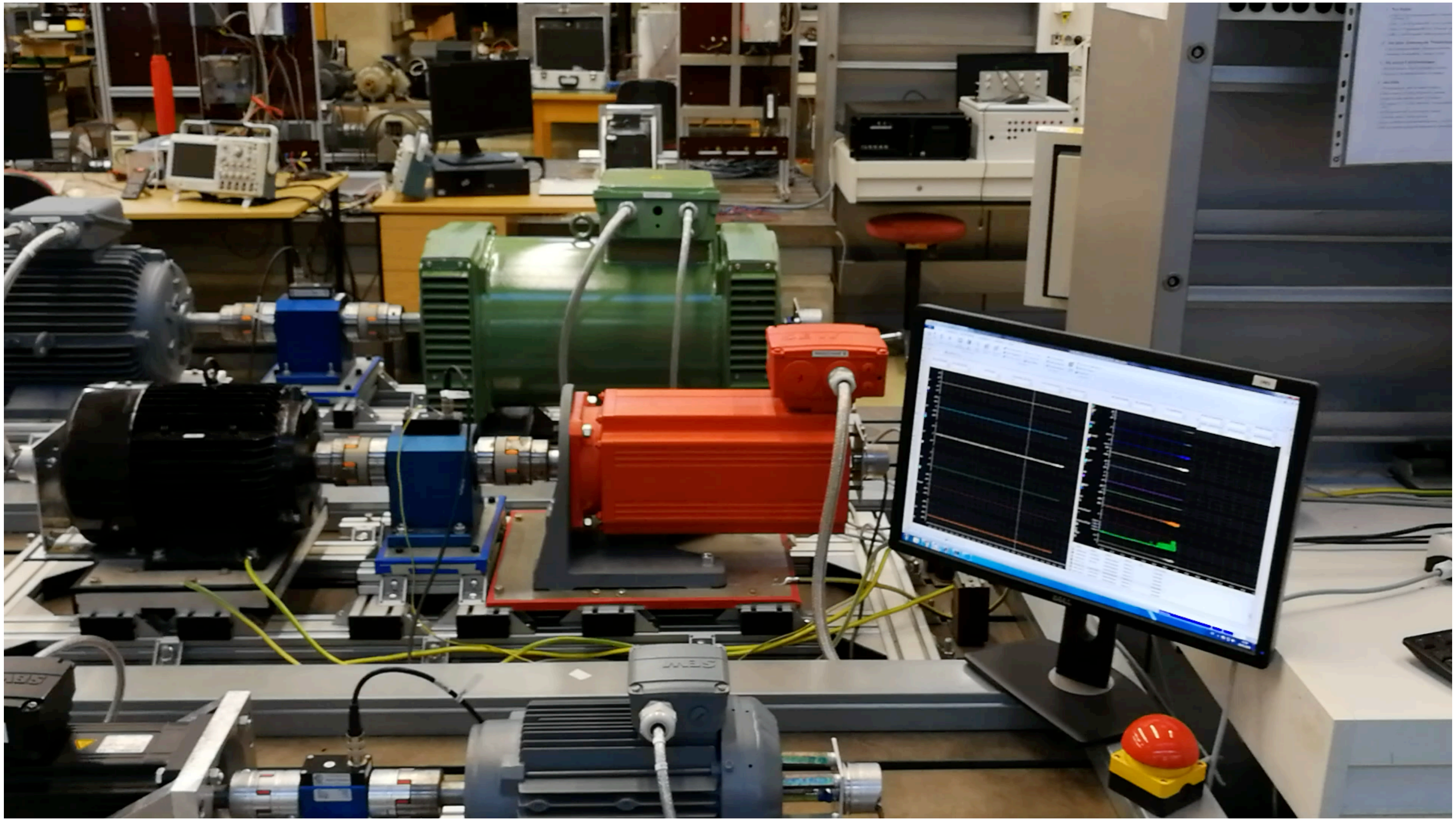
(b) fitted grey box flux model - q-component

Optimization Problem resulting from Direct Multiple Shooting:

$$
\begin{aligned}
\min_{\substack{\psi_0,\ldots,\psi_N \\ u_0,\ldots,u_{N-1}}} \quad & \frac{T_h}{2N} \sum_{i=0}^{N-1} \left\| \begin{matrix} \psi_i - \bar{\psi} \\ u_i - \bar{u} \end{matrix} \right\|_W^2 + \frac{1}{2} \|\psi_N - \bar{\psi}\|_{W_N}^2 \\
\text{s.t.} \quad & \psi_0 - \psi_\mathrm{e} = 0, \\
& g(\psi_i, u_i, \omega_\mathrm{e}, v_\mathrm{e}) - \psi_{i+1} = 0, \quad i = 0, \ldots, N-1, \\
& u_i^\top u_i \leq \left(\frac{u_\mathrm{dc}}{\sqrt{3}} \right)^2, \qquad i = 0, \ldots, N-1, \\
& \hat{C} u_i \leq \hat{c}, \qquad i = 0, \ldots, N-1,
\end{aligned}
$$



RSM: Video from NMPC Experiments at TU Munich



CS-NMPC significantly better than PI Controller

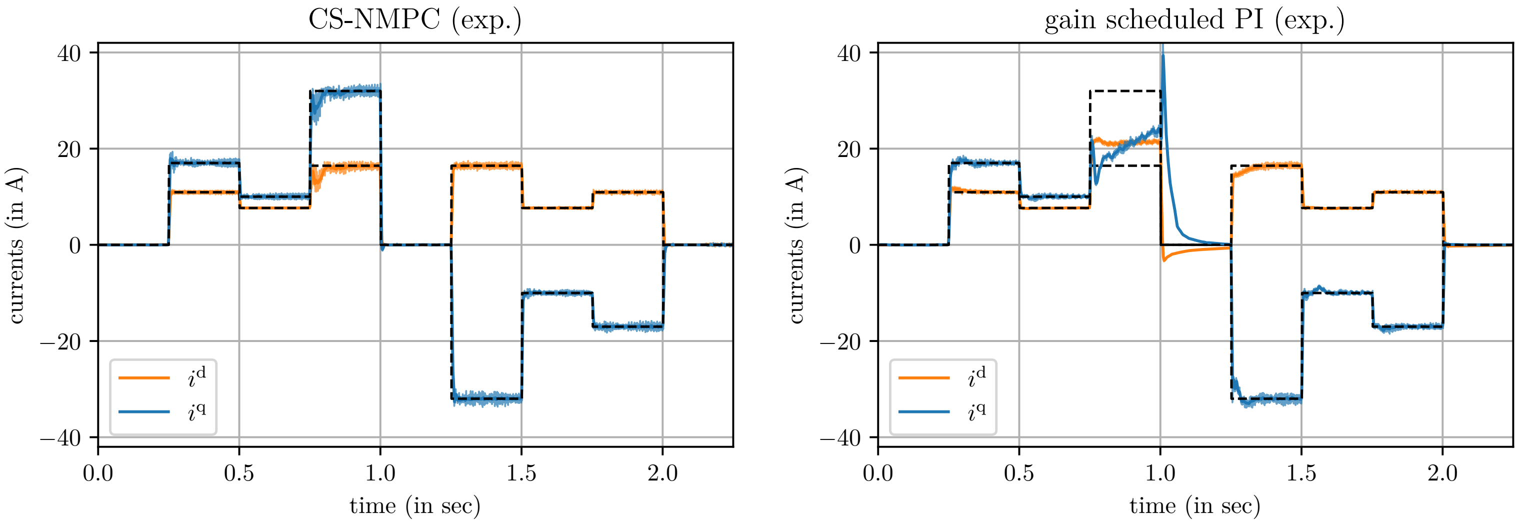


Figure 8: Current steps at 157 $\frac{\mathrm{rad}}{\mathrm{s}}$ (experiment): results obtained using the proposed CS-NMPC controller (left) and gain-scheduled PI controller (right). The CS-NMPC controller outperforms the PI controller, especially when the input constraints become active (e.g., between $t = 0.75\,\mathrm{s}$ and $t = 1.00\,\mathrm{s}$). At the same time, as it can be seen especially between $t = 1.25\,\mathrm{s}$ and $t = 1.50\,\mathrm{s}$, a faster transient can be achieved, even when the constraints are active only for a short time.

Agenda

Time Slots	Monday	Tuesday	Wednesday	Thursday	Friday
09:00-09:45	Lecture 1 - Introduction - Joschka Boedecker and Moritz Diehl	Lecture 4 - Dynamic Programming and LQR - Moritz Diehl	Lecture 7 - Numerical Optimal Control - Moritz Diehl	Lecture 10 - On-policy RL with Function Approximation - Joschka Boedecker	Microexam
10:00-10:45	Lecture 2: Dynamic Systems and Simulation (Moritz Diehl) (/files/2021ss/MPCRL/lecture-2-simulation.pdf)	Lecture 4 (continued) - Moritz Diehl	Lecture 7 (continued) - Moritz Diehl	Lecture 11 - Off-policy RL with Function Approximation - Joschka Boedecker	Lecture 14 - Robust and Stochastic MPC - Moritz Diehl
11:15-12:00	Exercise 1 - Dynamic System Simulation - Katrin Baumgärtner und Jasper Hoffmann	Exercise 3 - Dynamic Programming and LQR - Katrin Baumgärtner	Exercise 5 - Numerical Optimal Control - Katrin Baumgärtner	Exercise 7 - RL with Function Approximators - Jasper Hoffmann	Project Guidelines
14:00-14:45	Lecture 3: Numerical Optimization (Moritz Diehl) (/files/2021ss/MPCRL/lecture-3-optimization.pdf)	Lecture 5 - MDP Formalization and Monte Carlo Methods - Joschka Boedecker	Lecture 8 - MPC Stability Theory - Moritz Diehl	Lecture 12 - Policy Gradient Methods - Joschka Boedecker	Lecture 15 - Planning and Learning - Joschka Boedecker
15:00-15:45	Extended Coffee Break / Get-to-Know Each-Other-Session	Lecture 6 - Temporal Difference and Q-Learning - Joschka Boedecker	Lecture 9 - MPC Algorithms - Moritz Diehl	Lecture 13 - Advanced Value-based Methods - Joschka Boedecker	Lecture 16 - Differences and Similarities of MPC and RL - Joschka Boedecker and Moritz Diehl
16:15-17:00	Exercise 2 - Numerical Optimization - Katrin Baumgärtner	Exercise 4 - Q-Learning - Jasper Hoffmann	Exercise 6 - Model Predictive Control - Katrin Baumgärtner	Exercise 8 - Policy Gradient - Jasper Hoffmann	Project Pitch Presentations

Time Slots	Monday	Tuesday	Wednesday
09:00-09:45	Guest Lecture Sebastien Gros: Adaptation of MPC via RL: fundamental principles	Project Work	Project Work
10:00-10:45	Project Commitments	Project Status Updates	Project Presentations
11:15-12:00	Project Work Question session: Jasper Hoffmann & Katrin Baumgärtner	Project Work Question session: Jasper Hoffmann & Katrin Baumgärtner	Project Presentations
14:00-14:45	Guest Lecture Sebastien Gros: RL and MPC: safety, stability, and some more recent results	Project Work	Project Presentations
15:00-15:45	Project Work	Project Work Question Session: Jasper Hoffmann & Katrin Baumgärtner	Certificate Handout (A,B,C)
16:15-17:00	Project Work Question Session: Jasper & Katrin Baumgärtner	Guest Lecture - Sergey Levine: Model-Free and Model-Based Reinforcement Learning from Offline Data	End at 16:00