# Fusing Deep Reinforcement Learning and Model Predictive Control for Adaptive, Stochastic, and Robust Nonlinear Motion Control of Autonomous Mobile Robots

Baha Zarrouki[1,2] and Johannes Betz[2]

***Abstract*— We present a unified framework for high-speed autonomous vehicle control integrating four key advancements in Nonlinear Model Predictive Control (NMPC) for high-dimensional, real-time control systems: (1) Reduced Robustified NMPC enabling real-time computation through sensitivity-based uncertainty propagation, (2) Stochastic NMPC with adaptive Uncertainty Propagation Horizon (UPH) to prevent covariance explosion, (3) Deep Reinforcement Learning (DRL)-driven parameter adaptation for dynamic disturbance handling, and (4) Safe Reinforcement Learning (RL) with Pareto-constrained weight optimization. Validated through extensive simulation and real-world testing at speeds up to 37.5 m/s (135 km/h), our framework achieves real-time performance with 140 Hz solving frequencies, ensuring 98.77% constraint satisfaction and sub-0.1m tracking accuracy. All implementations are available open-source at `github.com/bzarr/TUM-CONTROL`.**

## I. Introduction

Autonomous vehicle motion control at operational limits demands simultaneous optimization of conflicting objectives: precise trajectory tracking, passenger comfort preservation, and robust constraint satisfaction under time-varying disturbances. While nonlinear model predictive control (NMPC) provides a principled framework for multi-objective optimization, three fundamental limitations hinder its deployment in high-speed scenarios:

- **Computational Complexity**: Traditional robust/ stochastic NMPC methods become intractable for high-dimensional vehicle models with long prediction horizons
- **Conservatism-Feasibility Tradeoff**: Stochastic NMPC approaches suffer from either covariance explosion through full-horizon propagation or excessive conservatism through fixed truncation
- **Context Sensitivity**: Static cost function weights optimized for nominal conditions degrade performance during transient maneuvers (e.g., obstacle avoidance) and across vehicle platforms

Our framework addresses these challenges through three interconnected novelties: (1) Reduced-order uncertainty propagation using sensitivity matrices from prior solutions, cutting computation time by 36x versus state-of-the-art Robustified NMPC (Fig.1); (2) Adaptive UPH management via DRL to balance constraint feasibility and controller aggressiveness (Fig.3); (3) Pareto-constrained RL adaptation ensuring safe weight selection while enabling 210% tracking improvement over manual tuning (Fig.4).

[1] Chair of Automotive Technology, Technical University of Munich
[2] Professorship of Autonomous Vehicle Systems, TUM School of Engineering and Design, Technical University of Munich, 85748 Garching, Germany; Munich Institute of Robotics and Machine Intelligence (MIRMI), {baha.zarrouki, johannes.betz}@tum.de

## II. Core Methodologies

### A. Reduced Robustified NMPC (R²NMPC)

Our R²NMPC algorithm simplifies robust NMPC by approximating ellipsoidal uncertainty dynamics using sensitivities from prior solutions. Traditional robust NMPC [1] propagates ellipsoidal uncertainty sets through the prediction horizon, requiring $O(n_x^2)$ additional optimization variables. Our reduced formulation instead uses sensitivity matrices $\mathbf{A}, \mathbf{B}$ from prior solutions:

$$\mathbf{\Sigma}_{k+1} = \mathbf{A}_k \mathbf{\Sigma}_k \mathbf{A}_k^\top + \mathbf{B}_k \mathbf{W} \mathbf{B}_k^\top \tag{1}$$

where $\mathbf{W}$ is the disturbance covariance. This allows constraint tightening [2]–[7] via:

$$h_{\text{rob}} = h_{\text{nom}} + \sqrt{\nabla h^\top \mathbf{\Sigma} \nabla h} \tag{2}$$

while maintaining the nominal NMPC's $O(n_x)$ complexity. Experimental validation shows the method achieves a mean solving frequency of 140Hz with 98.77% constraint satisfaction under disturbances, operating 4.67 times faster than traditional Robustified NMPC [7], [8] and 36 times faster than ZoRo NMPC [9] (Fig.1).

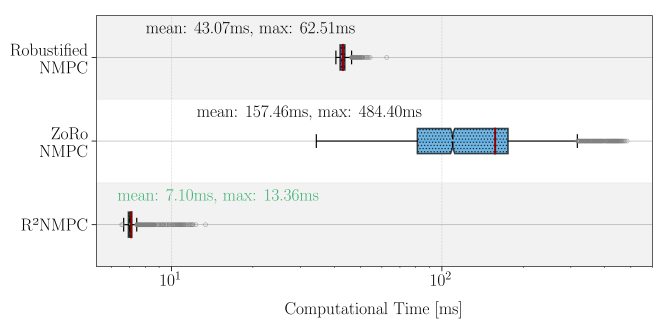



Fig. 1: Computational times needed for solving the OCP over 6000 simulation steps for different Robustified NMPC approaches [1], [8], [9].

### B. Stochastic NMPC with Uncertainty Propagation Horizon

To address the divergence of uncertain states in SNMPC, we introduced the concept of Uncertainty Propagation Horizon (UPH), which limits the time horizon for uncertainty propagation [10], [11]. For stochastic systems [12]–[16], we employ Polynomial Chaos Expansion (PCE) [17]–[19] with truncated propagation:

$$\mathbb{E}[h] = \begin{cases} c_0^{(h)} & t \leq T_u \\ h(\mathbb{E}[\mathbf{x}]) & t > T_u \end{cases} \tag{3}$$

where $T_u$ is the adaptive Uncertainty Propagation Horizon (UPH) and $c_0^{(h)}$ represents the zeroth-order PCE coefficient capturing the mean value of constraint $h$, calculated via:

$$c_0^{(h)} = \frac{1}{n_s} \sum_{i=1}^{n_s} h(\tilde{\mathbf{x}}^{(i)}) \tag{4}$$

with $\tilde{\mathbf{x}}^{(i)}$ being sampled uncertainty realizations.
The robustification factor $\kappa = \sqrt{(1-p)/p}$ transforms chance constraints [14], [20]:

$$\mathbb{P}(h \leq \bar{h}) \geq 1 - p \Rightarrow \mathbb{E}[h] + \kappa\sqrt{\mathrm{Var}[h]} \leq \bar{h} \tag{5}$$

This approach achieves a minimum solving frequency of 97Hz (Tab.I) while delivering robust performance. Experiments on autonomous vehicle motion control demonstrated a 61% improvement in maximum lateral deviation compared to nominal NMPC (2), with only 1.6% performance degradation under disturbances. The method reduces infeasibility by 63% compared to full-horizon propagation.

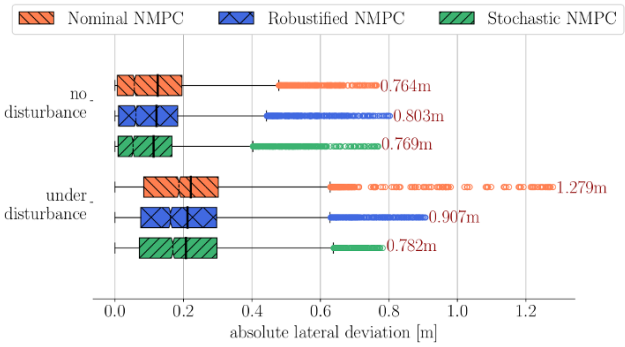



Fig. 2: Absolute lateral deviation of different NMPC designs after one lap on Monteblanco race track [1], [10].

| MPC solution frequency | Mean [Hz] | Min [Hz] |
|---|---|---|
| **R²NMPC [1]** | 144 | 82 |
| **SNMPC [10]** | 172 | 97 |
| **NMPC [10]** | 953 | 465 |

TABLE I: Benchmark of NMPC computational times across different approaches (Intel i7-11850H 2.50GHz CPU).

### C. Adaptive SNMPC with Deep Reinforcement Learning

We introduced an Adaptive SNMPC framework driven by Deep Reinforcement Learning (DRL) to optimize uncertainty handling and constraints robustification dynamically. Our adaptive framework [11] combines Polynomial Chaos Expansion (PCE) with DRL for real-time parameter adaptation:

$$\pi_\phi : (\mathbf{x}_t, \mathbf{r}_{t:t+H}) \mapsto (\kappa_t, T_{u,t}) \tag{6}$$

where $\kappa_t$ is the robustification factor and $T_{u,t}$ the Uncertainty Propagation Horizon (UPH). The DRL agent learns to adjust these parameters to enhance both feasibility and closed-loop performance. Results show this adaptive approach [21]–[23] achieves a minimum solving frequency of 110Hz while delivering 31% lateral deviation reduction compared to static controllers (Fig.3), maintaining safety throughout training under time-variant disturbances.

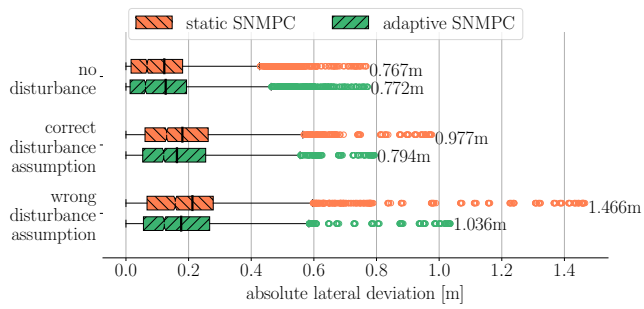



Fig. 3: Effect of the disturbance on different SNMPC designs w.r.t. the absolute lateral deviation [11].

### D. RL-Driven MPC Cost Function Adaptation (RL- Weights-varying MPC)

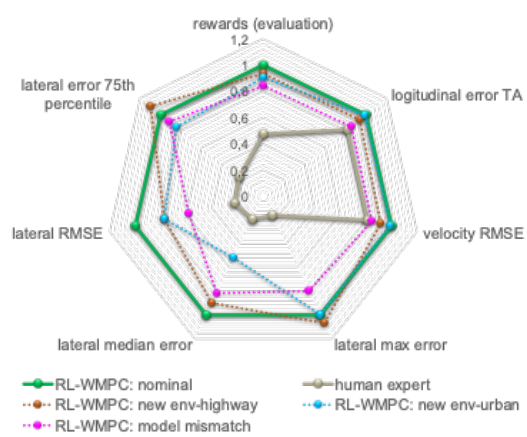



Fig. 4: DRL agent adapts the cost function weights of a Weights-varying MPC (WMPC): DRL-WMPC outperforms an MPC expert and is robust when faced with model-mismatch and new inexperienced operational environment.

Modern autonomous vehicles require adaptive control strategies that strike a trade-off between multiple objectives under varying operational conditions. MPC's effectiveness

depends critically on appropriate cost function weighting. Our research program addresses this challenge through two novel approaches:

*1) Continuous DRL Adaptation (2021) [24]:* our initial novel introduction of a Weights-varying MPC (WMPC) with neural network policies trained in continuous action spaces with DRL employing Twin Delayed Deep Deterministic Policy Gradient (TD3):

$$\pi_\phi^{\text{cont}} : (\mathbf{o}_t, \boldsymbol{\beta}_t) \mapsto \boldsymbol{\theta}_t \in \mathbb{R}_+^6 \tag{7}$$

where:

- $\mathbf{o}_t = [e_{\text{lat}}, e_{\text{vel}}, a_{\text{lat}}, j_{\text{long}}]$: Observation vector
- $\boldsymbol{\beta}_t$: Future reference trajectory features
- $\boldsymbol{\theta}_t = [q_{xy}, q_\psi, q_a, r_a, r_\omega, r_\vartheta]$: Adapted MPC weights

The Multi-Objective Cascaded Gaussian (MOCG) reward enables simultaneous optimization of tracking and comfort:

$$R_t = R_2\left(R_2(e_{\text{vel}}, e_{\text{lat}}) - 1, R_2(a_{\text{lat}}, j_{\text{long}}) - 1\right) \tag{8}$$

Key results demonstrated 210% tracking improvement over manually tuned MPC controllers by MPC experts and exhibited robustness in previously inexperienced scenarios (Fig.4). The framework also proved highly adaptable to modified optimization goals without requiring structural changes, highlighting its flexibility.

*2) Safe RL with Pareto Constraints (2024) [25]:* Safety-enhanced extension [26]–[28] using multi-objective Bayesian optimization (MOBO) [29]–[34] to constrain RL actions to pre-computed optimal weight sets.

Building on the foundational work, the enhanced framework introduces:

$$\pi_\phi^{\text{safe}} : (\mathbf{x}_t, \mathbf{r}_{t:t+H}) \mapsto a_t \in \{0, \ldots, n_{\mathcal{A}} - 1\} \tag{9}$$

where:

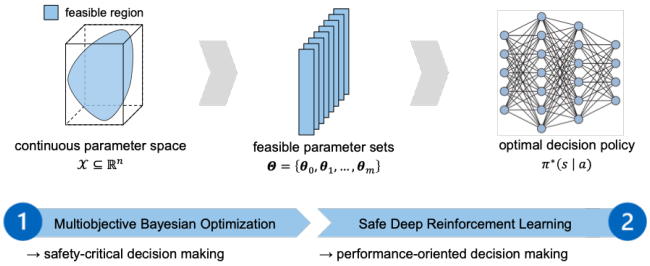
- $\mathcal{A} = \{\boldsymbol{\theta}^{(i)}\}_{i=0}^{n_{\mathcal{A}}-1}$: safe action-space (MOBO-optimized Pareto front)
- $a_t$: Discrete action selecting from $\mathcal{A}$
- $H$: Look-ahead horizon matching MPC prediction

The Pareto-optimal weight sets $\mathcal{A}$ are pre-optimized solving:

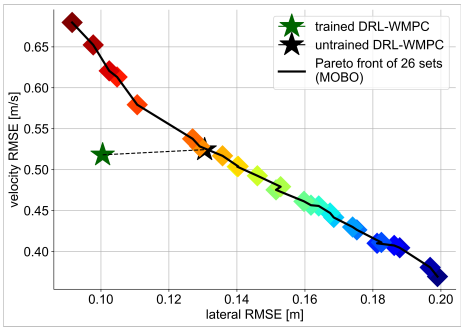
$$\min_{\boldsymbol{\theta}} [J_1(\boldsymbol{\theta}), J_2(\boldsymbol{\theta}), \ldots, J_n(\boldsymbol{\theta})]^T \quad \text{s.t.} \quad g(\boldsymbol{\theta}) \leq 0 \tag{10}$$

with $\boldsymbol{g}$ being safety constraints and $n$ the number of optimization objectives.

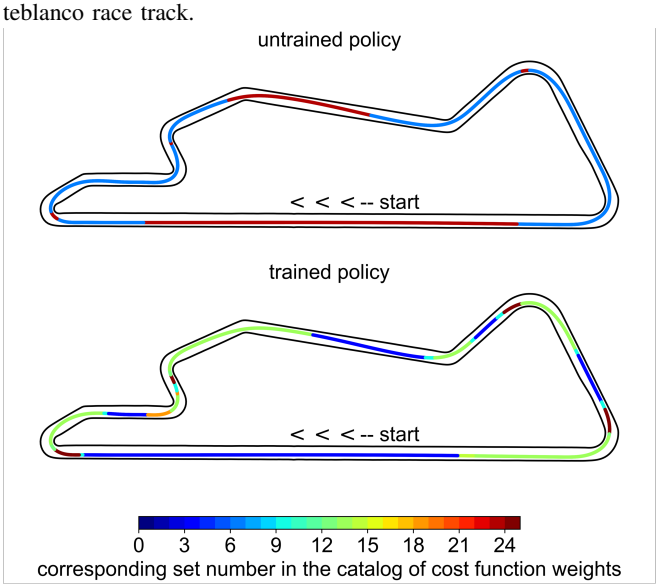
Experimental results demonstrate that the RL-driven selector achieves closed-loop performance surpassing individual Pareto-optimal configurations (Fig.5 and 6), improving multi-objective trade-offs by 18–24% compared to static weight choices. Crucially, the constrained action space ensures baseline Pareto-optimal behavior persists even with untrained agents, validating the inherent safety guarantees. While changes to MPC prediction model parametrization necessitate cost function re-optimization, the framework's automated retraining pipeline significantly reduces adaptation effort compared to manual tuning.




(a) Approach of the Safe Deep Reinforcement Learning (DRL) agent adapts the cost function weights of a Weights-varying MPC (WMPC).




(b) Untrained agent: Pareto optimal performance. Trained RL-WMPC: achieves performance beyond the Pareto-front on Monteblanco race track.

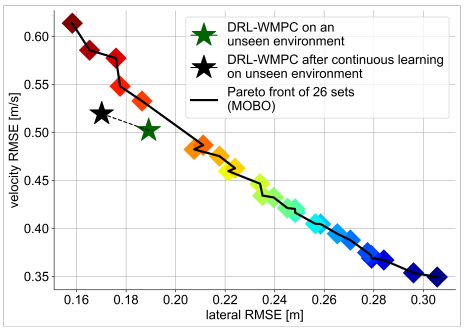



(c) RL actions heatmap.

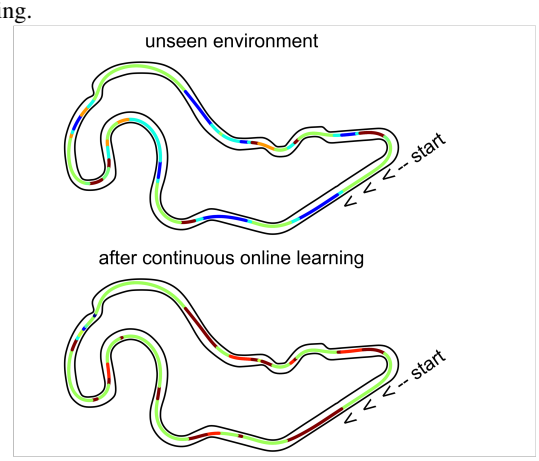
Fig. 5: Safe RL action space: choose a set of cost function weights from a Pareto-front of 26 sets obtained with Multi-objective Bayesian Optimization (MOBO).

## III. Conclusion and Future Directions

This work establishes advancements in high-speed autonomous motion control through three key advances: (1) Sensitivity-aware uncertainty propagation enabling real-time robust/stochasic NMPC, (2) DRL-driven Adaptive uncertainty horizons balance feasibility and conservatism, and (3) Safety-guaranteed RL adaptation through MOBO-constrained Pareto optimization provides context-aware




(a) Generalization to unseen environments and continuous online learning.




(b) Safe RL action space: choose a set of cost function weights from a Pareto-front of 26 sets obtained with Multi-objective Bayesian Optimization (MOBO).

Fig. 6: Safe continuous online learning enables the controller to autonomously expand its knowledge in previously inexperienced environments.

adaptation without stability compromises.

Future work will focus on two frontiers: (1) Meta-learning for rapid adaptation to new vehicle dynamics, and (2) Formal stability proofs for RL-augmented NMPC. Our implementations are open-source and available at `github.com/bzarr/TUM-CONTROL`. All experiments were conducted using the TUM-CONTROL simulation framework on a standard laptop with an Intel i7-11850H 2.50GHz CPU and 16GB RAM. We implemented our NMPC formulations using the ACADOS library [35] in Python 3.9, with prediction and simulation models built in CasADi [36].

## References


[1] Baha Zarrouki, Joao Nunes, and Johannes Betz. R2nmpc: A real-time reduced robustified nonlinear model predictive control with ellipsoidal uncertainty sets for autonomous vehicle motion control. *IFAC-PapersOnLine*, 58(18):309–316, 2024. 8th IFAC Conference on Nonlinear Model Predictive Control NMPC 2024.

[2] Alexander Kurzhanski and István Vályi. *Ellipsoidal calculus for estimation and control*. Springer, 1997.

[3] Mario E Villanueva, Rien Quirynen, Moritz Diehl, Benoît Chachuat, and Boris Houska. Robust mpc via min–max differential inequalities. *Automatica*, 77:311–321, 2017.

[4] Zhenyu Han, Shurong Li, and Shaowen Peng. Lyapunov equations approach for robust nonlinear optimal control problems. *Research Journal of Applied Sciences, Engineering and Technology*, 4(13):2017–2023, 2012.

[5] Makito Oku and Kazuyuki Aihara. On the covariance matrix of the stationary distribution of a noisy dynamical system. *Nonlinear Theory and Its Applications, IEICE*, 9:166–184, 04 2018.

[6] Maximilian Behr, Peter Benner, and Jan Heiland. Solution formulas for differential sylvester and lyapunov equations. *Calcolo*, 56(4):51, 2019.

[7] Boris Houska. *Robust Optimization of Dynamic Systems*. PhD thesis, 08 2011.

[8] B. Houska and M. Diehl. Robust nonlinear optimal control of dynamic systems with affine uncertainties. In *In Proceedings of the 48th Conference on Decision and Control, Shanghai, China*, page 2274–2279, 2009.

[9] Andrea Zanelli, Jonathan Frey, Florian Messerer, and Moritz Diehl. Zero-order robust nonlinear model predictive control with ellipsoidal uncertainty sets. *IFAC-PapersOnLine*, 54(6):50–57, 2021. 7th IFAC Conference on Nonlinear Model Predictive Control NMPC 2021.

[10] Baha Zarrouki, Chenyang Wang, and Johannes Betz. A stochastic nonlinear model predictive control with an uncertainty propagation horizon for autonomous vehicle motion control. In *2024 American Control Conference (ACC)*, pages 5466–5473, 2024.

[11] Baha Zarrouki, Chenyang Wang, and Johannes Betz. Adaptive stochastic nonlinear model predictive control with look-ahead deep reinforcement learning for autonomous vehicle motion control. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12726–12733, 2024.

[12] Lorenzo Fagiano and Mustafa Khammash. Nonlinear stochastic model predictive control via regularized polynomial chaos expansions. In *2012 IEEE 51st IEEE conference on decision and control (cdc)*. IEEE, 2012.

[13] Eric Bradford and Lars Imsland. Stochastic nonlinear model predictive control using gaussian processes. In *2018 european control conference (ECC)*, pages 1027–1034. IEEE, 2018.

[14] Eric Bradford, Lars Imsland, Dongda Zhang, and Ehecatl Antonio del Rio Chanona. Stochastic data-driven model predictive control using gaussian processes. *Computers & Chemical Engineering*, 2020.

[15] Eric Bradford and Lars Imsland. Combining gaussian processes and polynomial chaos expansions for stochastic nonlinear model predictive control. *arXiv preprint arXiv:2103.05441*, 2021.

[16] Ali Mesbah. Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control Systems Magazine*, 36(6):30–44, 2016.

[17] Robert H Cameron and William T Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, pages 385–392, 1947.

[18] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

[19] Frank WJ Olver. *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge university press, 2010.

[20] Ali Mesbah, Stefan Streif, Rolf Findeisen, and Richard D Braatz. Stochastic nonlinear model predictive control with probabilistic constraints. In *2014 American control conference*, pages 2413–2419. IEEE, 2014.

[21] Eivind Bøhn, Signe Moe, Sebastien Gros, and Tor Arne Johansen. Reinforcement learning of the prediction horizon in model predictive control. *IFAC-PapersOnLine*, 54(6):314–320, 2021.

[22] Eivind Bøhn, Sebastien Gros, Signe Moe, and Tor Arne Johansen. Optimization of the model predictive control meta-parameters through reinforcement learning. *Engineering Applications of Artificial Intelligence*, 123:106211, 2023.

[23] Mohit Mehndiratta, Efe Camci, and Erdal Kayacan. Automated tuning of nonlinear model predictive controller by reinforcement learning. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3016–3021. IEEE, 2018.

[24] Baha Zarrouki, Verena Klös, Nikolas Heppner, Simon Schwan, Robert Ritschel, and Rick Voßwinkel. Weights-varying mpc for autonomous vehicle guidance: a deep reinforcement learning approach. In *2021 European Control Conference (ECC)*, pages 119–125, 2021.

[25] Baha Zarrouki, Marios Spanakakis, and Johannes Betz. A safe reinforcement learning driven weights-varying model predictive control for autonomous vehicle motion control. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 1401–1408, 2024.

[26] Kim P Wabersich and Melanie N Zeilinger. Performance and safety of bayesian model predictive control: Scalable model-based rl with guarantees. *arXiv preprint arXiv:2006.03483*, 2020.

[27] Felix Berkenkamp, Angela P Schoellig, and Andreas Krause. Safe controller optimization for quadrotors with gaussian processes. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 491–496. IEEE, 2016.

[28] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. *Advances in neural information processing systems*, 30, 2017.

[29] Dario Piga, Marco Forgione, Simone Formentin, and Alberto Bemporad. Performance-oriented model learning for data-driven mpc design. *IEEE control systems letters*, 3(3):577–582, 2019.

[30] Ali Gharib, David Stenger, Robert Ritschel, and Rick Voßwinkel. Multi-objective optimization of a path-following mpc for vehicle guidance: A bayesian optimization approach. In *2021 European Control Conference (ECC)*, pages 2197–2204. IEEE, 2021.

[31] Farshud Sorourifar, Georgios Makrygirgos, Ali Mesbah, and Joel A Paulson. A data-driven automatic tuning method for mpc under uncertainty using constrained bayesian optimization. *IFAC-PapersOnLine*, 54(3):243–250, 2021.

[32] Joel A Paulson, Ketong Shao, and Ali Mesbah. Probabilistically robust bayesian optimization for data-driven design of arbitrary controllers with gaussian process emulators. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3633–3639. IEEE, 2021.

[33] Alisa Rupenyan, Mohammad Khosravi, and John Lygeros. Performance-based trajectory optimization for path following control using bayesian optimization. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 2116–2121. IEEE, 2021.

[34] Georgios Makrygiorgos, Angelo D Bonzanini, Victor Miller, and Ali Mesbah. Performance-oriented model learning for control via multi-objective bayesian optimization. *Computers & Chemical Engineering*, 162:107770, 2022.

[35] Robin Verschueren, Gianluca Frison, Dimitris Kouzoupis, Jonathan Frey, Niels van Duijkeren, Andrea Zanelli, Branimir Novoselnik, Thivaharan Albin, Rien Quirynen, and Moritz Diehl. acados – a modular open-source framework for fast embedded optimal control. *Mathematical Programming Computation*, Oct 2021.

[36] Joel A E Andersson, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 2019.