ANNUAL REVIEWS

Annual Review of Control, Robotics, and Autonomous Systems

Learning-Based Model Predictive Control: Toward Safe Learning in Control

Lukas Hewing, Kim P. Wabersich, Marcel Menner, and Melanie N. Zeilinger

Institute for Dynamic Systems and Control, ETH Zurich, Zurich 8092, Switzerland; email: lhewing@ethz.ch, wkim@ethz.ch, mmenner@ethz.ch, mzeilinger@ethz.ch

Annu. Rev. Control Robot. Auton. Syst. 2020. 3:269–96

First published as a Review in Advance on December 2, 2019

The *Annual Review of Control, Robotics, and Autonomous Systems* is online at control.annualreviews.org

https://doi.org/10.1146/annurev-control-090419-075625

Keywords

model predictive control, learning-based control, safe learning, adaptive control, autonomous systems

Abstract

Recent successes in the field of machine learning, as well as the availability of increased sensing and computational capabilities in modern control systems, have led to a growing interest in learning and data-driven control techniques. Model predictive control (MPC), as the prime methodology for constrained control, offers a significant opportunity to exploit the abundance of data in a reliable manner, particularly while taking safety constraints into account. This review aims at summarizing and categorizing previous research on learning-based MPC, i.e., the integration or combination of MPC with learning methods, for which we consider three main categories. Most of the research addresses learning for automatic improvement of the prediction model from recorded data. There is, however, also an increasing interest in techniques to infer the parameterization of the MPC controller, i.e., the cost and constraints, that lead to the best closed-loop performance. Finally, we discuss concepts that leverage MPC to augment learning-based controllers with constraint satisfaction properties.

1. INTRODUCTION

Model predictive control (MPC) has seen significant success in recent decades and has established itself as the primary control method for the systematic handling of system constraints (1), with wide adaptation in diverse fields (2), such as process control, automotive systems, and robotics. The MPC scheme relies on a sufficiently descriptive model of the system to optimize performance and ensure constraint satisfaction, rendering modeling critical for the success of the resulting control system. In practice, however, model descriptions can be subject to considerable uncertainty, originating, e.g., from insufficient data, restrictive model classes, or the presence of external disturbances. The fields of robust (3) and stochastic (4) MPC provide a systematic treatment of numerous sources of uncertainty affecting the MPC controller, ensuring constraint satisfaction with regard to certain disturbance or uncertainty classes. These controller paradigms, however, follow a strict separation of a design phase, which is carried out offline by a control engineer, and an application phase of closed-loop control, during which the formulation of the controller remains largely unchanged. The availability of increasing computational power and sensing and communication capabilities, as well as advances in the field of machine learning, has given rise to a renewed interest in automating controller design and adaptations based on data collected during operation, e.g., for improved performance, facilitated deployment, and a reduced need for manual controller tuning.

This review provides an overview of these research efforts in the context of learning-based MPC. Most research has focused on an automatic data-based adaptation of the prediction model or uncertainty description. The opportunities for learning- and data-based adaptation in MPC, however, go beyond a data-driven model improvement. The relationship between the specific parameterization of an MPC controller—e.g., the cost function, horizon length, or terminal components employed—and the resulting closed-loop behavior of the control system is often difficult and cannot be analytically described, motivating several research efforts toward the use of data-driven techniques for controller design and improvement. A conceptually different direction using MPC in the context of learning-based control leverages MPC solely to address constraint satisfaction and safety, while closed-loop performance is optimized, e.g., using a reinforcement learning algorithm.

The field of learning-based and data-driven control is vast, and we would like to point out some connections to related methods and viewpoints that we do not address in this review. There is a large body of research on iterative learning control, some of which has direct connections to MPC (5, 6). While there are efforts toward approximating explicit MPC control laws using learning techniques (e.g., 7–10), we focus on learning in online MPC, i.e., where the learning affects the MPC optimization problem to be solved at each sampling instance. A central question in the context of learning and control is the resulting exploration–exploitation trade-off and the problem of dual control (11), i.e., the optimal simultaneous identification and control. We include only a brief discussion of this aspect and refer readers to a recent survey by Mesbah (12) for an overview. A data-based optimization technique with several applications to MPC that has received substantial interest in recent years is scenario optimization. We briefly make the connection to these results in this review but refer readers to References 13–15 for a more detailed discussion of the approach and resulting MPC schemes. For an overview of reinforcement learning from the control perspective, we refer readers to Reference 16, which includes a brief discussion of MPC in this context.

The review is organized as follows. In Section 2 we give a technical introduction to the problem of stochastic optimal control and motivate (learning-based) MPC as an approximate solution strategy. Section 3 reviews learning methods focusing on an improvement of the dynamics model in MPC, whereas Section 4 focuses on learning-based parameterization of MPC controllers to improve closed-loop behavior. In Section 5 we finally discuss MPC as a technique to augment learning-based control methods in order to achieve safety during operation and learning. We close with a discussion in Section 6.

2. OPTIMAL CONTROL OF UNCERTAIN SYSTEMS

In order to motivate and classify the various types of learning-based MPC, we begin with the interpretation of MPC as an approximation of the ideal stochastic optimal control problem that is to be solved, but the exact solution of which is typically not tractable. By referring to this underlying original problem, we then highlight different opportunities in which adaptation and learning in the MPC formulation have been investigated with the goal of improving the approximation and the resulting controller.

2.1. The General Stochastic Optimal Control Problem

The methods discussed in this review deal with the problem of controlling dynamical systems that are subject to system constraints under uncertainty, which can affect numerous parts of the problem formulation. We formulate the system dynamics in discrete time as

$$x(k+1) = f_{\mathrm{t}}(x(k), u(k), k, w(k), \theta_{\mathrm{t}}), \qquad 1.$$

in which $x(k) \in \mathbb{R}^{n_x}$ is the system state and $u(k) \in \mathbb{R}^{n_u}$ is the applied input at time k. We use the subscript "t" to emphasize that these quantities represent the true system dynamics or true optimal control problem. The dynamics are subject to various sources of uncertainty, which we distinguish by using two categories: $\theta_{\mathrm{t}} \sim \mathcal{Q}^{\theta_{\mathrm{t}}}$ is a random variable describing the parametric uncertainty of the system, which is therefore constant over time, and $w(k)$ describes a sequence of random variables corresponding to disturbances or process noise in the system, which are often assumed to be independent and identically distributed (i.i.d.). The true problem therefore relates to the development of an optimal controller for a distribution of systems given by $\mathcal{Q}^{\theta_{\mathrm{t}}}$ under random disturbances $w(k)$. Throughout this review, we assume direct access to measurements of the system state $x(k)$ and neglect the problem of state estimation.

The optimality of the controller is defined with respect to a cost or objective function. In the presence of random model uncertainties, the cost is often defined as the expectation of a sum of potentially time-varying stage costs of the states and inputs over a possibly infinite horizon $\bar{N}$:

$$J_{\mathrm{t}} = \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_{\mathrm{t}}(x(k), u(k), k)\right), \qquad 2.$$

where the expected value is taken with respect to all random variables, i.e., $w(k)$ and θ_{t}.

The difficulty central to the methods discussed in this review is the requirement to satisfy constraints on states and inputs during closed-loop control. While input constraints typically represent a physical actuator limitation, such as the limited torque of an electric motor, constraints on the system state can arise in different capacities. They can similarly represent the physical limits of the system but also express constraints in terms of system safety or relate to a desired performance, e.g., by encoding a desired comfort range. Expressing the state and input trajectories over the horizon as $\bar{X} = [x(0), \ldots, x(\bar{N})]$ and $\bar{U} = [u(0), \ldots, u(\bar{N})]$, respectively, one can compactly express such system constraints as

$$\Pr(\bar{X} \in \bar{\mathcal{X}}_j) \geq p_j \text{ for all } j = 1, \ldots, n_{c_x}, \qquad 3a.$$

$$\Pr(\bar{U} \in \bar{\mathcal{U}}_j) \geq p_j \text{ for all } j = 1, \ldots, n_{c_u}. \qquad 3b.$$

Note that the probability $p_j = 1$ implies the important special case of hard constraints, i.e., constraints that are to be fulfilled for every possible realization of the random variables $w(k)$ and θ_{t}. In the form presented, the constraints express joint chance constraints in time, i.e., constraints that need to hold for the entire state or input trajectory with a given probability. This general form, however, includes the case of individual constraints at each time step, e.g., $\Pr(x(k) \in \mathcal{X}) \geq p$ through a suitable definition of $\bar{\mathcal{X}}_j$. The resulting constrained stochastic optimal control problem can be formulated as

$$J_{\mathrm{t}}^* = \underset{\{\pi_k\}}{\text{minimize}} \quad \mathbb{E}\left(\sum_{k=0}^{\bar{N}} l_{\mathrm{t}}(x(k), u(k), k)\right) \tag{4a.}$$

$$\text{subject to} \quad x(k+1) = f_{\mathrm{t}}(x(k), u(k), k, w(k), \theta_{\mathrm{t}}), \tag{4b.}$$

$$u(k) = \pi_k(x(0), \ldots, x(k)), \tag{4c.}$$

$$\bar{W} = [w(0), \ldots, w(\bar{N}-1)] \sim \mathcal{Q}^{\bar{W}},\ \theta_{\mathrm{t}} \sim \mathcal{Q}^{\theta_{\mathrm{t}}}, \tag{4d.}$$

$$\Pr(\bar{X} = [x(0), \ldots, x(\bar{N})] \in \bar{\mathcal{X}}_j) \geq p_j \text{ for all } j = 1, \ldots, n_{c_x}, \tag{4e.}$$

$$\Pr(\bar{U} = [u(0), \ldots, u(\bar{N}-1)] \in \bar{\mathcal{U}}_j) \geq p_j \text{ for all } j = 1, \ldots, n_{c_u}, \tag{4f.}$$

optimizing over a sequence of control laws $\{\pi_k\}$, which can make use of all information in the form of state measurements $x(k)$ up to time step k. Problems of the form in Equation 4 are in general very hard to solve, and direct efforts typically rely on some form of discretization in space and approximate dynamic programming or reinforcement learning (16, 17). A notable exception is linear systems under additive noise and quadratic stage costs in the unconstrained setting, for which an exact solution to Equation 4 exists (18), such as the standard linear quadratic regulator (LQR).

The challenges in a direct attempt to solve Equation 4 are manifold. In particular, a direct optimization over general feedback laws π_k is not possible. Other difficulties are related to the evaluation of the cost and the nonconvexity of the resulting problem. A particular challenge is due to the model uncertainty, where an optimal control policy should not only be robust, but also take into account that information about the model can be gained from state measurements. In the following, we motivate MPC as a tractable approximate solution strategy. This is followed by a discussion of efforts using learning and adaptation in MPC, e.g., by collecting information about the parametric uncertainty θ_{t} from measurement data (Section 3), with the goal of improving performance with respect to the general optimal control problem in Equation 4 while taking constraints into account.

2.2. Model Predictive Control

MPC approximates the problem in Equation 4 by repeatedly solving a simplified version of the problem initialized at the currently measured state $x(k)$ over a shorter horizon N in a receding-horizon fashion. To this end, we introduce the prediction model

$$x_{i+1|k} = f(x_{i|k}, u_{i|k}, i+k, w_{i|k}, \theta), \tag{5.}$$

as well as predicted states and inputs. We use the subscript $i|k$ to emphasize predictive quantities, where, e.g., $x_{i|k}$ is the i-step-ahead prediction of the state, initialized at $x_{0|k} = x(k)$. The prediction dynamics f typically aims at approximating the true dynamics in Equation 1 but often differs, e.g., for computational reasons or because a succinct description of the true dynamics is unavailable.

While the true system dynamics f_t can be very complex, control formulations based on, e.g., a linear approximation can often yield tractable and powerful controllers.

MPC formulations of varying complexity exist, the most widespread of which are nominal MPC schemes, which do not consider any uncertainties in the prediction model but instead rely exclusively on the compensation of uncertainties via feedback and by re-solving the problem at the next sampling instance. We exemplify the MPC problem in the following for such a nominal approach. Robust and stochastic variants can similarly be derived from the optimal control problem in Equation 4 by considering the cost and constraints over a shortened prediction horizon N. In nominal MPC, the optimization can be performed over control sequences $U = [u_{0|k}, \ldots, u_{N-1|k}]$ rather than policies, resulting in the constrained optimal control problem

$$J^* = \underset{U}{\text{minimize}} \quad l_f\,(x_{N|k}, u_{N|k}, k+N) + \sum_{i=0}^{N-1} l(x_{i|k}, u_{i|k}, i+k) \qquad \text{6a.}$$

$$\text{subject to} \quad x_{i+1|k} = f(x_{i|k}, u_{i|k}, i+k), \qquad \text{6b.}$$

$$U = [u_{0|k}, \ldots, u_{N|k}] \in \mathcal{U}_j \text{ for all } j = 1, \ldots, n_{c_u}, \qquad \text{6c.}$$

$$X = [x_{0|k}, \ldots, x_{N|k}] \in \mathcal{X}_j \text{ for all } j = 1, \ldots, n_{c_x}, \qquad \text{6d.}$$

$$x_{N|k} \in \mathcal{X}_f, \qquad \text{6e.}$$

$$x_{0|k} = x(k). \qquad \text{6f.}$$

Again, the stage cost function l does not necessarily coincide with the actual cost l_t, which may express complex objectives that can be sparse, nondifferentiable, or even unavailable as a precise mathematical expression, such that it is often ill suited for a solution using gradient-based numerical optimization. Instead, l is often chosen to facilitate the optimization problem, e.g., providing differentiability or convexity. The cost function most widely used in MPC is a weighted quadratic cost, which is particularly suitable for tracking tasks.

To mitigate the effect of the shortened horizon, a particular cost $l_f(x_{N|k}, u_{N|k}, k+N)$ and constraint $\mathcal{X}_f$ on the last predicted state are imposed to approximate the cost and the effect of the constraints over the remainder of the possibly infinite control horizon $\bar{N}$. In many MPC formulations, these terminal components l_f and $\mathcal{X}_f$ play a significant role in establishing properties of the closed-loop control system, such as asymptotic stability (19). The MPC control law is then implicitly defined through the optimization problem in Equation 6 as

$$\pi^{\text{MPC}}(x(k), k) = u^*_{0|k}, \qquad \text{7.}$$

where $u^*_{0|k}$ is the first element of the computed optimal control sequence U^*.

In particular, in the convex and time-invariant setting (i.e., for linear dynamics, convex cost function, and constraint sets), nominal MPC methods and robust variants are well studied, and the theory is comparatively well established (3, 19, 20). MPC theory usually involves two main properties: recursive or persistent feasibility and stability. Recursive feasibility states that if the MPC optimization problem in Equation 6 is feasible for a state $x(k)$, and the computed control input $u(k) = u^*_{0|k}$ is applied to the dynamical system, it can be ensured that the MPC problem is again feasible for (all possible) resulting states $x(k+1)$, and the problem is therefore feasible for all future time steps. The feasibility of the MPC problem is used to show the satisfaction of the constraints in Equation 3 during closed-loop control. While different notions of stability for the resulting closed-loop system exist, the stability analysis is often based on Lyapunov-type arguments, e.g., using the optimal cost function J^* of the MPC as a Lyapunov function.

As such, theoretical guarantees in MPC are typically ensured by design in the controller synthesis, as opposed to an a posteriori analysis of the resulting control system, either analytically or by large-scale simulation, as is commonly the case, e.g., in computer science and reinforcement learning. In the following, we give a brief overview of learning in MPC and an outline of the remainder of the article.

2.3. Overview of Learning-Based Model Predictive Control

Learning-based MPC addresses the automated and data-driven generation or adaptation of elements of the MPC formulation such that the control performance with respect to the desired closed-loop system behavior—i.e., the general optimal control problem (Equation 4)—is improved. The setup in which this learning takes place can be diverse. For instance, offline learning considers the adaptation of the controller between different trials or episodes of a control task, during which data are collected. In methods that learn online, on the other hand, the controller is adjusted during closed-loop operation (e.g., while performing repetitive tasks) or using the data collected during one task execution.

While much of the research in learning-based MPC is focusing on automatically improving the model quality, which is the most obvious component affecting MPC performance, several research efforts are addressing the formulation of the MPC problem directly or utilizing the MPC concept to satisfy constraints during learning-based control. In the remainder of the review, we discuss the research in the following three categories:

- Learning the system dynamics: MPC relies heavily on suitable and sufficiently accurate model representations of the system dynamics. One path of learning-based MPC considers the automatic adjustment of the system model, either during operation or between different operational instances. Section 3 provides an overview of this rather broad direction and related issues.
- Learning the controller design: A second interesting research direction focuses less on the prediction model and more on the remaining problem formulation, such as the employed cost function l, the constraints $\mathcal{X}$, or the terminal components l_f and $\mathcal{X}_f$, such that the resulting closed-loop MPC controller behaves favorably with respect to the underlying task, i.e., the stochastic optimal control problem. We discuss these approaches in Section 4.
- MPC for safe learning: A third direction is the use of MPC techniques to derive safety guarantees for learning-based controllers. The main idea is to decouple the optimization of the objective function l_t from the requirement of constraint satisfaction, which is addressed using MPC techniques. We discuss this research direction in Section 5.

The organization of this review, together with the corresponding referenced literature, is shown in **Table 1**.

3. LEARNING THE SYSTEM DYNAMICS

System modeling represents the first step of an MPC design, which is traditionally addressed by deriving a parametric prediction model offline (i.e., before closed-loop control) using physical principles and applying system identification techniques. Robust or stochastic MPC can then be used to systematically address predictive control of dynamical systems under remaining uncertainties and disturbances. In the robust case, for instance, model uncertainty as well as process noise is often considered to lie in compact sets $\theta_\mathrm{t} \in \mathcal{T}$, $w(k) \in \mathcal{W}$, while stochastic approaches make use of distributional information on the uncertainties. In both cases, however, model and

Table 1 Organization of the review and the referenced literature

Section	Specific subtopic	References
Learning the System Dynamics	Robust models	37–41
	Parametric	44–61
	Nonparametric	64–70
	Stochastic models/robust in probability	71–73
	Parametric	77–84
	Nonparametric	86, 87, 90–92, 95–100
Learning the Controller Design	Performance-driven learning	—
	Bayesian optimization	102–108
	Terminal components	110–113
	Inverse optimal control	114–126
Model Predictive Control for Safe Learning	—	72, 73, 127–138

uncertainty descriptions are typically fixed offline and assumed available before controller design. Learning-based approaches, on the other hand, take collected data, such as state measurements, explicitly into account—e.g., by constructing the function f or uncertainty sets $\mathcal{T}$ and $\mathcal{W}$ from data and possibly updating these sets over time.

Many learning-based MPC techniques make use of an explicit distinction between a nominal system model f_{n} and an additive learned term f_{l} accommodating uncertainty:

$$f(x, u, k, \theta, w) = f_{\mathrm{n}}(x, u, k) + f_{\mathrm{l}}(x, u, k, \theta, w). \qquad 8.$$

An implicit assumption is then that (rudimentary) control of the true system based on the nominal model is possible, such that data in terms of state measurements $x(k)$ resulting from the true system dynamics f_{t} under a specific true parameter realization θ_{t} can be safely collected. Most methods do not distinguish between the true system dynamics f_{t} and prediction dynamics f—i.e., a typical assumption is that the true dynamics lie within the class of considered prediction dynamics, and given knowledge of the true parameter realization θ_{t}, there is no model mismatch. For simplicity, we refer to f as both the system dynamics and the prediction model in the following and highlight the difference only where relevant.

Successful learning methods are often based on probabilistic formulations (21), leading to a nonlinear stochastic prediction model in Equation 5. Leveraging the full potential of such models within MPC, however, is very challenging and remains an active research field (4). Many model-learning MPC schemes have therefore evolved from the extensively studied field of robust MPC, offering a large body of available theoretical results (3, 20). We first discuss learning-based MPC methods making use of a robust paradigm in Section 3.1.

As a second category, we consider techniques that are stochastic in nature in Section 3.2. This category includes adaptations of robust control schemes to stochastic estimation techniques. For instance, the basic MPC mechanism can be adopted from robust MPC, while in particular the considered parametric uncertainty θ is stochastic. This results in schemes that are robust with respect to some (likely) set of parameter values, which we call robust in probability. Lastly, we also investigate fully stochastic approaches, in which typically both stochastic parameters θ and process noise $w(k)$ are considered, aiming at a further reduction of conservatism and integration of stochastic learning techniques.

In this review, we focus on methods allowing for a principled treatment of the resulting model uncertainty, which is critical in the context of constraint satisfaction. There is related research using model learning in MPC that neglects uncertainties of the learned model $f_{\mathrm{l}}(x, u, k, \theta, w)$ and results in nominal MPC formulations. Important examples include the use of artificial

MODEL PREDICTIVE CONTROL WITH ACTIVE LEARNING AND DUAL CONTROL

Dual control considers the challenge of simultaneous identification and control of dynamical systems (11) and was first studied by Feldbaum (26, 27). In the optimal control setting, such as the problem in Equation 4, the optimal policy needs to take the information gain over the horizon into account in order to address this trade-off. The problem can be studied using dynamic programming; however, the exact solution is computationally intractable for all but the smallest toy examples (18), which is why most learning-based MPC techniques use passive learning schemes. Mesbah (12) presented a recent survey of dual control in the context of MPC.

Dual control methods are typically divided into two categories: explicit and implicit. In explicit dual control, the exploratory aspect of the controller (i.e., the dual effect) is explicitly inserted into the control scheme—e.g., through a form of exploration bonus in the cost function (28, 29) or by ensuring persistent excitation (30, 31) to enhance adaptation speed and ensure the quality of the parameter estimates. The resulting techniques typically do not significantly increase computational complexity, but the exploration–exploitation trade-off is addressed heuristically.

In implicit dual control, the optimal stochastic control problem in Equation 4, including the information gain with respect to the parameters along the control horizon, is directly approximated. This can be achieved, e.g., through the use of sampling (32–34) or a specific system structure (35, 36). In contrast to explicit approaches, the exploration–exploitation trade-off is therefore naturally addressed, but the resulting methods tend to be limited to specific system classes or are computationally expensive.

neural networks (e.g., 22, 23) and recent results using data-based nonparametric linear system descriptions (as in 24, 25). MPC approaches explicitly addressing the dual control problem (i.e., the optimal simultaneous identification and control) are briefly discussed in the sidebar titled Model Predictive Control with Active Learning and Dual Control.

3.1. Robust Models

Robust MPC schemes guarantee the satisfaction of closed-loop constraints for all possible realizations of the uncertain elements θ_{t} and $w(k)$—i.e., the problem in Equation 4 is guaranteed to hold with probability one. Similarly to Equation 8, a split into a nominal model f_{n} and additive uncertainty is often employed, where the uncertainty is assumed to lie in a compact set, i.e., $f_{\mathrm{l}}(x, u, k, \theta_{\mathrm{t}}, w) \in \bar{\mathcal{W}}$, and the controller is designed to be robust against the additive uncertainty. While the predicted MPC cost is then typically optimized for the nominal system, some learning-based extensions consider optimization of the cost for a performance model that is generated or improved from data. The benefit of such an approach is that theoretical guarantees, particularly recursive feasibility and satisfaction of constraints in closed loop, carry over from robust MPC theory, while performance can be improved based on measurements received during operation.

Aswani et al. (37) proposed a learning-based MPC framework based on this concept. Their approach uses two prediction models. The first is a nominal linear model with an additive term comprising all uncertainty $w_i \in \bar{\mathcal{W}}$,

$$z_{i+1|k} = Az_{i|k} + Bv_{i|k} + w_{i|k},$$

on the basis of which the recursive feasibility of the approach is established using techniques from linear robust MPC. The second is a learned performance model,

$$x_{i+1|k} = Ax_{i|k} + Bu_{i|k} + \mathcal{O}(x_{i|k}, u_{i|k}, i, k),$$

where $\mathcal{O}$ is an oracle representing an arbitrary learning-based approximation of $f_{\mathrm{l}}(x, u, k, \theta_{\mathrm{t}}, w)$. By optimizing a cost with respect to the performance model, i.e., $x_{i|k}$, the MPC can therefore improve its performance while robustly enforcing all constraints. Practical demonstrations of this method have been presented in References 38–41. By using a decomposition in terms of a safety and performance model, the approach has direct connections to MPC for safe learning, which we discuss in further detail in Section 5.

While the method presented by Aswani et al. (37) therefore allows for some adaptation and improvements with respect to the performance achieved, it considers one fixed model uncertainty set $f_{\mathrm{l}}(x, u, k, \theta_{\mathrm{t}}, w) \in \bar{\mathcal{W}}$ assumed to be known beforehand. Various learning-based MPC schemes aim at estimating this model uncertainty set directly from data, potentially adjusting it over time to reduce conservatism. Many of the techniques are based on set-membership identification, estimating the function f from noisy measurements with bounded uncertainty $w(k) \in \mathcal{W}$. We distinguish parametric approaches, which aim at finding the set of parameter values θ consistent with the observations, and nonparametric approaches, which form their estimate of the function f directly from observed data points and derive bounds on the function, e.g., based on Lipschitz arguments. We provide a simplified conceptual overview of set-membership techniques here; for additional details, we refer readers to Reference 42 for the parametric case and Reference 43 for the nonparametric case.

3.1.1. Robust parametric models.

Given a measured state and input trajectory $X = [x(0), \ldots, x(k)]$, $U = [u(0), \ldots, u(k-1)]$, parametric set-membership estimation aims at finding the set of possible parameter values θ for which the observed trajectories are consistent. This is formulated as a feasible parameter set:

$$\mathcal{T}_k = \{\theta \mid \forall j = 0, \ldots, k-1\ \exists w(j) \in \mathcal{W},\ \text{such that}\ x(j+1) = f(x(j), u(j), j, \theta, w(j))\}. \qquad 9.$$

In a stochastic setting, Equation 9 can be interpreted as the set of parameter values with nonzero probability. An iterative update can be defined as

$$\mathcal{T}_{k+1} = \{\theta \in \mathcal{T}_k \mid \exists w(k) \in \mathcal{W},\ \text{such that}\ x(k+1) = f(x(k), u(k), k, \theta, w(k))\}. \qquad 10.$$

It can be directly seen that the estimates of the parameter set using set-membership techniques are subject to a nestedness property

$$\mathcal{T}_{k+1} \subseteq \mathcal{T}_k, \qquad 11.$$

such that the set of consistent parameters is nonincreasing over time. As the complexity of set iterates can grow unbounded as more state measurements are included in the estimation, measures to counteract this complexity growth are usually employed—e.g., Veres et al. (44) limited the number of half-spaces defining the resulting polytopes. In addition to the estimation of viable parameter values $\mathcal{T}_k$, many techniques use a nominal model estimate $\hat{\theta}_k$, which, as with a nominal model in robust MPC, is often used to evaluate the cost term J in the MPC and to optimize controller performance. Such a nominal estimate can, e.g., be found by maximizing the distance to the boundary of $\mathcal{T}_k$ or using a suitable projection of any point estimate onto the set $\mathcal{T}_k$.

Parametric set-membership estimation has been used in various results, particularly for different forms of linear time-invariant systems, expressed as

$$f(x, u, k, \theta, w) = A(\theta)x + B(\theta)u + w, \qquad 12.$$

which, depending on the type of selected model, may have a specific structure.

Tanaskovic et al. (45) presented a set-membership MPC approach for steady-state reference tracking of finite impulse response systems, which is therefore limited to open-loop stable systems and output noise. Set-membership identification is applied to the parameters of the impulse response using a conservative approximation of the set $\mathcal{T}_k$ for computational reasons, which ensures that the set containment property in Equation 11 is maintained. This results in a recursively feasible adaptive MPC scheme, in which the uncertainty set $\mathcal{T}_k$ is subsequently reduced, and convergence to a desired steady state can be shown under some assumptions. The approach has been implemented in a building control setting (46), extended to linear time-varying systems (47) and stochastic output noise (48), and combined with a learning approach for terminal components (49) (see Section 4.1). Fagiano et al. (14) reviewed the method and contrasted it with scenario MPC, which we briefly touch on in Section 3.2.1. Terzi et al. (50–52) discussed the use of models mapping an entire input sequence to a state sequence over the prediction horizon for autoregressive exogenous models using set-membership identification. For polytopic state-space linear time-invariant systems in the general form of Equation 12, Di Cairano (53) and Zhou et al. (54) presented an MPC approach in which, similarly to the method of Aswani et al. (37), the model is improved for performance, but the model uncertainty sets $\mathcal{T}$ are not updated over time. Lorenzen et al. (55, 56) presented an approach for a similar problem class, taking additive process noise and recursive updates of the uncertainty sets $\mathcal{T}$ into account using set-membership techniques, and adapting the technique such that slowly varying parameters can be tracked.

Related ideas have been proposed using parameter adaptation laws for single-input linear systems (57, 58) and nonlinear systems that are linear in their parameters (59–61), thereby including systems of the form in Equation 12, i.e.,

$$f(x, u, \theta, w) = \Phi(x, u)\theta + w. \tag{13}$$

Instead of making use of set-membership identification, point estimates $\hat{\theta}_k$ and uncertainty sets $\mathcal{T}_k$ are derived from parameter update laws. Under the assumption that the true system parameter lies in an initially known compact set $\theta \in \mathcal{T}_0$, containment of the true parameter and nestedness $\mathcal{T}_k \subseteq \mathcal{T}_{k-1}$ can be ensured by construction. For the resulting MPC, the methods described by Sugie and colleagues (57, 58) use a scalar comparison system to bound the growth of uncertainty in the state prediction, taking the knowledge gained over the parameter into account. In the nonlinear case, two resulting MPC formulations have been presented for continuous-time (59, 60) and discrete-time (61) systems. The first approach finds a policy considering the worst-case parameter in the uncertainty set using min-max optimization, which is in general computationally intractable. The second formulation presents a tractable approximation based on linearization and Lipschitz bounds and is able to maintain theoretical properties such as recursive feasibility and convergence.

3.1.2. Robust nonparametric models.

In nonparametric estimation, the function f is not specified explicitly using parameter values; instead, a function estimate is formed directly based on (noisy) observations of function values and regularity properties. The aim is to avoid the specification or iterative search for an adequate parametric representation (43). Similar to parametric set-membership methods, the nonparametric form aims at providing strict bounds on the possible function values, given some observations with bounded uncertainty $w(k) \in \mathcal{W}$. Similar and related approaches are known as Lipschitz interpolation (62) and kinky inference (63).

Considering again a given trajectory X and U and a time-invariant function f with bounds on its gradient $|\nabla f| \leq \epsilon$ and bounded additive noise, the feasible system set is defined as

$$\mathcal{F}_k = \{f \mid \|\nabla f\| \leq \epsilon, \forall 0 \leq j \leq k-1 \; \exists w(j) \in \mathcal{W}, \text{such that } x(j+1) = f(x(j), u(j)) + w(j)\}.$$

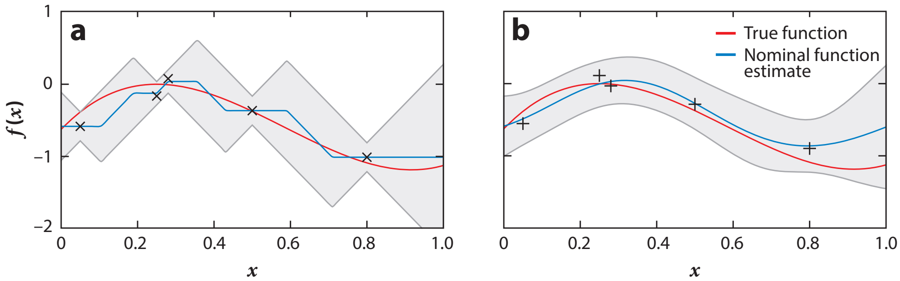

Figure 1

Robust and stochastic nonparametric estimation techniques. Noisy measurements of the true function (*red*) are shown as crosses; the nominal function estimate is shown in blue. (*a*) Illustration of the robust set-membership technique based on a bound of the noise $w(k)$ and the gradient of f, with the resulting bounds shown in gray. (*b*) Gaussian process regression estimate based on a squared exponential kernel function, with the 2-σ confidence bound shown in gray.

A recursive definition is given by

$$\mathcal{F}_{k+1} = \{f \in \mathcal{F}_k \mid \|\nabla f\| \leq \epsilon, \exists w(k) \in \mathcal{W}, \text{such that } x(k+1) = f(x(k), u(k)) + w(k)\}.$$

A nominal function estimate $\hat{f}$ is often chosen to maximize the distance to resulting feasible system bounds, i.e., as the average between the upper and lower bounds (illustrated in **Figure 1**).

To the best of our knowledge, the available results based on nonparametric set-membership identification do not consider an update of the model uncertainty during closed-loop control, but rather focus on robust control using models generated from data or an adaptation of a performance model. Canale et al. (64, 65) proposed an MPC approach for nonlinear autoregressive exogenous models using nonparametric set-membership identification. Limon et al. (66) presented a related approach that makes use of a data-based Lipschitz estimation technique (67). The methods generally suffer from the nonsmooth function estimate f (illustrated in **Figure 1**), which makes numerical optimization in an MPC scheme challenging. Ideas for smoothing the predictions have been proposed that use either a convex combination resulting from a number of data points surrounding the query point (68, 69) or a weighted sum of basis functions (70), both of which improve the computational properties while maintaining favorable theoretical guarantees.

3.2. Stochastic Models

While the robust treatment of uncertainty is an established technique that gives rise to a number of learning-based MPC methods with strong theoretical properties, their application typically requires a hard a priori bound covering all uncertainty, which can be conservative in practice. Stochastic MPC methods, on the other hand, rely on distributional information on the system uncertainty and do not necessarily make use of hard bounds, which, however, renders a theoretical analysis more difficult and often results in less rigorous results (4).

An approach that aims at combining the benefits of both methods converts the stochastic problem into a related robust one. The use of stochastic parameter estimation techniques leads to a stochastic belief about the parameter distribution $p(\theta|X, U)$, expressing the knowledge about the system dynamics given data X, U. Based on this distribution, in analogy to the feasible parameter sets in Equation 9, one can construct parameter confidence sets $\mathcal{T}_k$

$$\Pr(\theta \in \mathcal{T}_k | X, U) \geq p$$

with a specified probability p or, similarly, confidence sets directly on f in the case of nonparametric approaches. Given the confidence sets, established robust MPC methods can be used to design controllers that are robust in probability. If the true function lies in the estimated confidence set, robust guarantees carry over, and choosing a sufficiently large probability p ensures satisfaction of chance constraints (Equation 3). Along the same lines, robust MPC approaches have, e.g., been combined with stochastic model learning by Soloperto et al. (71) and Koller et al. (72).

One issue that arises in this approach is that, for the stochastic estimation procedure, $w(k)$ is often assumed to be Gaussian and can have large or unbounded support, making it incompatible with a robust control approach. Although presented in the context of MPC for safe learning (see Section 5), this issue was addressed by Wabersich et al. (73) by splitting the difference of the measured state from a nominal linear system into a robust-in-probability part (dealing with the parametric uncertainty) and a stochastic part (dealing with the unbounded noise). Paulson et al. (74) presented similar ideas in the context of stochastic MPC.

A second issue concerns the nestedness of parameter confidence sets along Equation 11, which, as opposed to robust set-membership techniques, is more challenging to guarantee in the stochastic case and makes automatic set updates during closed-loop control challenging. For instance, in each adaptation step there is a chance of up to $1 - p$ that θ does not lie in the current confidence set $\mathcal{T}_k$, which makes it difficult to provide feasibility guarantees.

In the following, we discuss learning-based MPC methods that are formulated directly in a stochastic setting, for which a theoretical analysis is generally challenging. We again address parametric uncertainty first and treat measurable disturbances $w(k)$ that are correlated in time as a special case, which is often approached using scenario optimization techniques (75, 76). For nonparametric approaches, Gaussian process (GP) regression has become a popular learning approach by directly providing an assessment of the residual model uncertainty after learning.

3.2.1. Stochastic parametric models. We take a Bayesian viewpoint of parametric estimation techniques, for which we again consider measured state and input trajectories X, U, as well as a noise distribution $p(w)$ and prior distribution over the parameter $p(\theta)$, expressing knowledge about the parameter that is assumed before observing data. The posterior distribution of the parameter, given state measurements, can then be expressed using Bayes' rule as

$$p(\theta|X, U) = \frac{p(X|U, \theta)p(\theta)}{\int p(X|U, \bar{\theta})p(\bar{\theta})\mathrm{d}\bar{\theta}}, \qquad 14.$$

where $p(X|U, \theta)$ can be obtained via the distribution of the noise $p(w)$ (i.e., the likelihood of the parameter θ is related to the likelihood of the particular noise sequence explaining the observed states X under this parameter). Using conditional independence due to the dynamics in Equation 1, we can define a recursive parameter update as

$$p(\theta|x(k+1), u(k), X, U) = \frac{p(x(k+1)|x(k), u(k), \theta))p(\theta|X, U)}{\int p(x(k+1)|x(k), u(k), \bar{\theta}))p(\bar{\theta}|X, U)\mathrm{d}\bar{\theta}}, \qquad 15.$$

such that one can iteratively update the posterior distribution $p(\theta|X, U)$ instead of tracking the entire state and input history. These equations can therefore be viewed as stochastic analogies generalizing the robust case in Equations 9 and 10. The computation of the distributions, however, is in general not tractable, with the notable exception of dynamic systems that are linear in their parameters (i.e., Equation 13) under Gaussian noise $w(k) \sim \mathcal{N}(\mu(k), \Sigma(k))$ and Gaussian prior parameter distribution $\theta \sim \mathcal{N}(\mu^\theta, \Sigma^\theta)$, for which Equations 14 and 15 can be solved by (recursive) Bayesian linear regression (21).

Due to many technical challenges in stochastic MPC and some of the issues discussed above, many stochastic learning-based MPC schemes come with little theoretical analysis, but they have nevertheless been very successful in practical implementations. For instance, Desaraju and colleagues (77, 78) presented parametric learning-based MPC approaches for robotic systems that enhance the model of a quadrotor over time while taking approximate uncertainty in the prediction into account through constraint tightening. Similarly, McKinnon & Schoellig (79) presented an implementation for trajectory tracking of an off-road robotic vehicle based on linear Bayesian regression for the actuator dynamics, with the particular focus of mitigating the conflict between fast adaptation and long term learning.

Several research efforts have focused on linear stochastic MPC with data-driven model adaptation based on measurable disturbance sequences w that are correlated in time, which allows for a more thorough analysis. Given measurements of $w(k)$ during operation, the conditional or predictive distribution $p(w(k), \ldots, w(k+N-1) | w(0), \ldots, w(k-1))$ can be used to adapt and improve the resulting expected cost J^* of the MPC. This idea has been presented in scenario-based MPC for additive disturbances (80), as well as for correlated multiplicative noise (81), which can be similarly interpreted as time-varying parametric uncertainty. Scenario-based MPC methods typically come with limitations regarding recursive feasibility and closed-loop constraint satisfaction, which Lorenzen et al. (82) overcame by generating fixed disturbance samples of the original distribution $p(W)$ instead of the conditional predictive one. Hewing et al. (83) presented an approach for systems under additive noise that similarly achieved recursive feasibility by considering the distribution of the disturbance sequence $p(W)$ while optimizing performance with respect to the conditional distribution. Di Cairano et al. (84) presented related techniques for vehicle control that modeled driver actions as an additive disturbance, estimated from data as a Markov chain.

3.2.2. Stochastic nonparametric approaches.

Because it provides a flexible stochastic nonparametric approach, GP regression is the most commonly employed technique in learning-based control. We give a brief account here and refer readers to Reference 85 for a detailed exposition. For notational simplicity, we consider here the scalar case; multiple outputs are typically considered by training a GP for each dimension independently.

GP regression for model learning usually assumes dynamics with i.i.d. zero mean additive Gaussian noise $w(k) \sim \mathcal{N}(0, \sigma_w^2)$ of the form $x(k+1) = f_{\mathrm{t}}(x(k), u(k)) + w(k)$ and builds on the assumption that values of the function f_{t} at different inputs x, u are jointly Gaussian distributed according to a kernel function k, expressing the covariance between function values. Using recorded state and input sequences X, U and considering for each state and applied input the resulting state measurement $X^+ = [x(1), \ldots, x(N+1)]$, the joint distribution of data points and function values at a test point (x, u) is

$$\begin{bmatrix} X^+ \\ f_{\mathrm{t}}(x,u) \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K + I\sigma_w^2 & K_{x,u} \\ K_{x,u}^T & k([x,u],[x,u]) \end{bmatrix}\right),$$

where K is the Gram matrix of recorded data X, U under the kernel k, and $K_{x,u}$ represents the corresponding entries for test point (x, u). The kernel k is essential for shaping the resulting prediction and must be chosen such that any resulting Gram matrix K is positive definite. Several functions that satisfy this condition are available, the most popular of which is the squared exponential kernel

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}(x - x')^T L (x - x')\right),$$

in which σ_f^2 is the signal variance and L is a positive diagonal length scale, constituting (together with the noise variance σ_w) the hyperparameters of the GP regression. Applying the rules of conditioning for Gaussian random variables then results in the fact that the function value $f_t(x, u)$ at a test point (x, u) given the recorded data is similarly normally distributed:

$$f_t(x, u)|X^+ \sim \mathcal{N}(\mu(x, u), \Sigma(x, u)),$$

where $\mu(x, u)$, and $\Sigma(x, u)$ are, respectively, the posterior mean and variance functions constituting the estimator, which depend on the collected data points X^+, X, and U. The variance function $\Sigma(x, u)$ can be used to assess residual model uncertainty due to insufficient data near the test point (x, u). **Figure 1** provides an illustration of a resulting function estimate.

A challenge in an MPC formulation based on a GP dynamics model is the propagation of the resulting stochastic state distributions over the prediction horizon. This is typically achieved by assuming that subsequent function evaluations of $f_t(x, u)|X^+$ are independent. Approximate Gaussian distributions of the predicted states $x_{i|k}$ over the prediction horizon are then derived using techniques related to extended Kalman filtering, e.g., by linearization (86), sigma-point transform (87), or exact moment matching (88)—an approach tailored to GP regression. Partly due to the approximate nature of the prediction, there is little theoretical analysis of the resulting MPC controller, with the exception of Reference 89, which presents nominal robustness properties of GP-based MPC, and Reference 72, which provides robust bounds on the estimation error that hold with a specified probability under the assumption that the true function f_t is an element of a reproducing kernel Hilbert space (RKHS) with a bounded RKHS norm.

There are, however, a number of promising practical results. The prediction model is typically split into a nominal and learning part, as in Equation 8, the latter of which is subsequently improved using GP regression to enhance the overall model quality. Due to its nonparametric nature, GP regression is particularly well suited to identifying discrepancies with a nominal system model, which are challenging to parameterize. Many techniques use the residual model uncertainty estimate to provide a heuristic constraint tightening and practical safety margins, as shown in **Figure 2** for an autonomous racing application (86, 90).

Computation times are a fundamental challenge in GP-based MPC, particularly since the model complexity grows with the number of recorded data points. Techniques to address this issue include data selection and maintenance of a dictionary of data points of limited size, computational approximations of the GP [e.g., via inducing points (93) or selected basis functions (94)], and neglecting or simplifying the variance dynamics (i.e., the evolution of the uncertainty). Building on these ideas, successful applications of GP-based MPC to robotic systems have been presented, e.g., for trajectory tracking with robotic manipulators (95), path following of off-road mobile robots (87, 96), autonomous racing (86, 90, 97), process control (98), and telescope systems in the context of periodic additive disturbances (99). Using autoregressive models, Jain et al. (100) presented a simulation study for building control and demand response, which additionally addresses the problem of suitable exploration by maximizing the information gain. Additionally, Kamthe & Deisenroth (88) showed that learning-based MPC is a highly competitive model-based reinforcement learning technique, particularly with regard to data efficiency.

4. LEARNING THE CONTROLLER DESIGN

While the system model is a core element of a predictive controller, other elements in the MPC formulation, such as the cost function and constraints, also have a major influence on the resulting closed-loop performance. In this section, we discuss learning-based methods for designing the

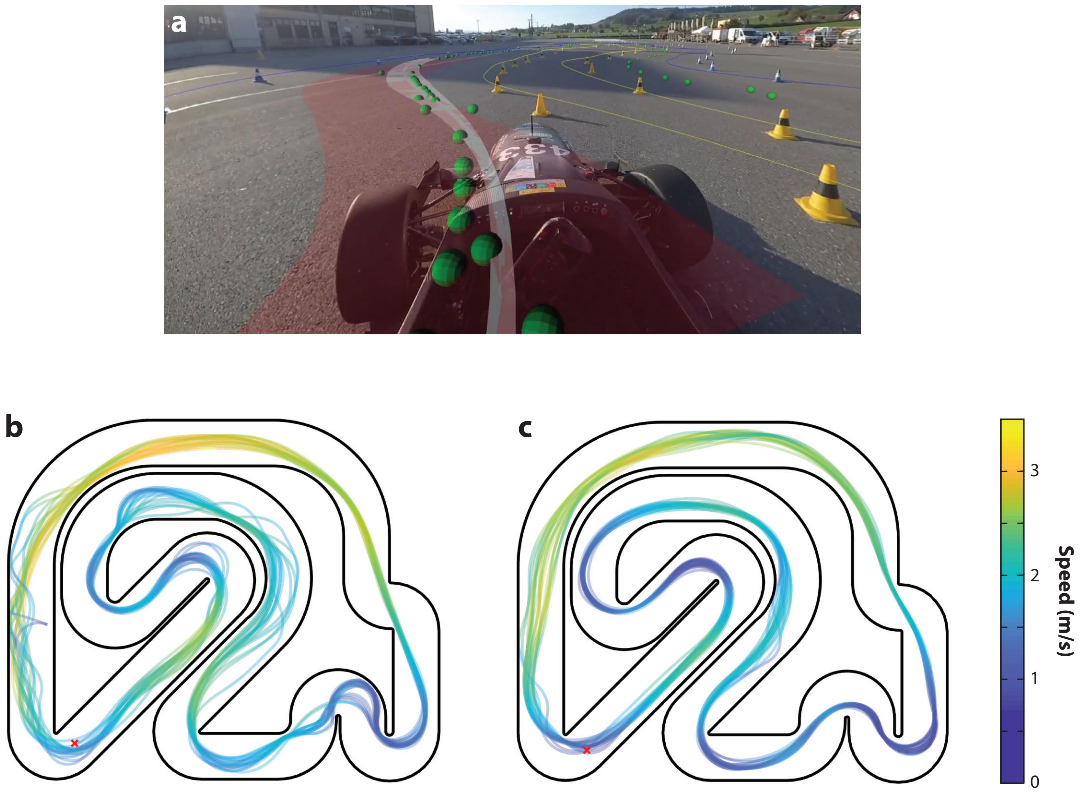

Figure 2

Gaussian process–based model predictive control for autonomous racing. (*a*) Application to a full-sized race car (90), showing the position constraint set (*red*); the predicted trajectory including uncertainty (*gray*), which is used for constraint tightening; and selected data points from previous rounds (*green spheres*). (*b*,*c*) The resulting trajectories of a similar approach applied to miniature radio-controlled cars (86), with the initial nominal controller shown in panel *b* and the improved trajectories after learning shown in panel *c*. For videos of these experiments, see References 91 and 92. Panel *a* is a still from Reference 92. Panels *b* and *c* adapted from Reference 86 with permission.

MPC problem according to a desired controller behavior. For this purpose, we consider a parameterized version of an MPC problem, including a parameterization of the cost function $l(x, u, \theta_l)$, and constraints $\mathcal{X}(\theta_{\mathcal{X}}), \mathcal{U}(\theta_{\mathcal{U}})$, i.e.,

$$U^* = \arg\min_{U} \sum_{i=0}^{\bar{N}} l(x_i, u_i, \theta_l) \qquad \text{16a.}$$

$$\text{subject to} \quad x_{i+1} = f(x_i, u_i, \theta_f), \qquad \text{16b.}$$

$$U = [u_0, \dots, u_N] \in \mathcal{U}(\theta_{\mathcal{U}}), \qquad \text{16c.}$$

$$X = [x_0, \dots, x_N] \in \mathcal{X}(\theta_{\mathcal{X}}), \qquad \text{16d.}$$

$$x_0 = x(k). \qquad \text{16e.}$$

We group the techniques into two main areas. Performance-driven controller learning, discussed in Section 4.1, aims at improving the closed-loop performance by successively adjusting the parameterization in Equation 16 in an episodic setting. Learning from demonstrations via inverse optimal control, on the other hand, addresses the fact that we can often observe a desired behavior of a closed-loop control system and would like to design an automatic controller according to these specifications. The goal is to translate the recorded data into a corresponding parameterization of the MPC problem, which we discuss in Section 4.2.

4.1. Performance-Driven Controller Learning

Receding-horizon controllers are often a rough approximation of the true stochastic optimal control problem, e.g., due to the finite prediction horizon, simplified models, and neglected stochastic disturbances. Performance-driven controller learning focuses on finding a parameterization of an MPC as in Equation 16 that optimizes closed-loop performance with respect to the optimal control problem in Equation 4. Below, we provide a brief overview of two main directions for addressing this problem: black-box optimization in the form of Bayesian optimization, and learning terminal components to counteract the finite-horizon nature of the controller.

4.1.1. Bayesian optimization for controller tuning. In the context of controller learning, Bayesian optimization aims at optimizing the true closed-loop cost $J_{\mathrm{t}}(\theta)$ as a mapping from the controller parameters $\theta = [\theta_l^T, \theta_f^T, \theta_{\mathcal{X}}^T, \theta_{\mathcal{U}}^T]^T$ to the cost achieved when performing a task with the MPC controller. Most commonly, Bayesian optimization considers the function $J_{\mathrm{t}}(\theta)$ using GP regression as a regularity assumption. (For a discussion of GPs in the context of system dynamics learning, see Section 3.2.2; for a general introduction to Bayesian optimization, see Reference 101.) In this episodic learning task, past trials—e.g., with parameter θ_i and resulting closed-loop cost $J_{\mathrm{t},i}$—are used as data points to infer information about the function $J_{\mathrm{t}}(\theta)$. This is achieved by relating different trials via a kernel function $k(\theta_i, \theta_j)$ that specifies the covariance between the trials at θ_i and θ_j. Based on this estimate of the function $J_{\mathrm{t}}(\theta)$, an acquisition function $\alpha(\theta)$ is used to trade off exploration and exploitation and to determine $\theta^\star = \arg\min_\theta \alpha(\theta)$ as the next parameter to be evaluated.

While this scheme can be applied to most parameterized controllers (102, 103), only a few techniques explicitly consider the parameterization of optimal control formulations. In the context of unconstrained optimal control, Marco et al. (104) introduced a parametric cost function $l(x, u, \theta_l) = x^T Q(\theta_l) x + u^T R(\theta_l) u$ and optimized the parameters θ_l in order to compensate for deviations of the true dynamics f_{t} from linear prediction dynamics f, such that the performance metric in closed loop is improved. In a similar setup, Fröhlich et al. (105) investigated learning in higher-dimensional spaces by automatic domain adaptation. For constrained optimal control, the approach of Bansal et al. (106) instead uses Bayesian optimization to learn the parameters θ_f of a linear prediction model $f(x, u, \theta_f) = A(\theta_f)x + B(\theta_f)u$ in an MPC, similarly with the goal of improving the closed-loop performance of a specific task for unknown, potentially nonlinear and stochastic true system dynamics f_{t}. In addition to optimizing over a linear parametric prediction model, Piga et al. (107) considered the prediction horizon of the MPC formulation as a parameter and optimized with Bayesian optimization. Gros & Zanon (108) presented related ideas in the context of economic MPC, making use of reinforcement learning rather than Bayesian optimization for parameter adaptation.

4.1.2. Learning terminal components. A potential limitation of MPC is related to its shortsightedness due to the limited prediction horizon, which is often imposed by computational constraints. A terminal cost function and constraint can be used to mitigate this effect; however, their design is challenging for complex systems and tasks. Typically, the goal is to find a large terminal invariant set to ensure recursive feasibility for a large set of initial states, and a terminal cost approximating the infinite-horizon cost and/or guaranteeing theoretical properties such as asymptotic stability of the closed-loop system. Learning in this context makes use of collected data to construct or improve the terminal cost and constraint.

The main mechanism for improving the terminal constraint relies on the fact that, for nominal time-invariant systems, a trajectory $X = [x(0), \dots, x(k)]$ terminating in a control-invariant set $\mathcal{X}_f$ yields an enlarged invariant set

$$\mathcal{X}_f^+ = \mathcal{X}_f \cup \left(\bigcup_{j=0}^{k} \{x(j)\} \right).$$

For linear time-invariant systems, Blanchini & Pellegrino (109) showed that the convex hull of this set is similarly control invariant. Brunner et al. (110) employed this mechanism in a linear MPC scheme that additionally uses scaling and translations of the terminal set around the trajectory points to further enlarge the data-based terminal set. The terminal cost function is then adapted as an interpolation of cost functions defined on the translated sets. Rosolia & Borrelli (111) presented a method for iterative tasks that uses a similar notion for a sampled safe set as the terminal constraint and combines it with terminal cost learning, assigning the cost-to-go from previous iterations to the points in the sampled safe set; they showed that the resulting overall cost of the trajectory of an episode is nonincreasing over the successive episodes. For nonlinear systems, this approach results in mixed-integer optimization, whereas for linear systems, relaxations based on the convex hull of the terminal set maintain theoretical guarantees. Rosolia et al. (112) presented a robust variant of the approach, and Rosolia & Borrelli (113) later extended the method to a repetitive learning setting and applied it to autonomous racing. A recent review by Rosolia et al. (15) also discusses the approach.

4.2. Learning from Demonstration with Inverse Optimal Control

Designing a cost function or defining an objective mathematically in order to achieve a desired complex behavior can be tedious and require extensive parameter tuning and development time. Inverse optimal control addresses the problem of defining an objective function l or constraints $\mathcal{X}, \mathcal{U}$ systematically by inferring them from demonstrations. To do so, a prediction model f of the system dynamics is usually assumed to be given. By learning the components of an optimal controller rather than imitating an observed policy, inverse optimal control offers favorable generalization properties, such that a resulting MPC control law can be derived for the entire state space even though observations are sparse. Another motivation arises in the context of human–controller interaction, where adjusting the controller based on human demonstrations makes its behavior more natural for the human.

The hypothesis underlying inverse optimal control is that the observed demonstrations are the solution of a corresponding optimal control problem. Specifically, we consider here the case where a demonstrated input trajectory $U_\mathrm{t} = [u(0), \dots, u(N)]$ starting at $x(0)$ is the optimal solution to the unknown deterministic optimal control problem in Equation 16 for specific values of the unknown parameters θ_l, $\theta_{\mathcal{X}}$, and $\theta_{\mathcal{U}}$, i.e., $U_\mathrm{t} = U^*$. Most of the research in inverse optimal control has assumed knowledge of potential constraints and focused on the cost function instead, following three main steps:

1. Define the optimal control problem with a parametric cost function $l(x, u, \theta_l)$, e.g., quadratic costs with unknown weights $l(x, u, \theta_l) = x^T Q(\theta_l) x + u^T R(\theta_l) u$.
2. Derive optimality conditions for the parametric optimal control problem.
3. Solve optimality conditions for the parameters θ_l given the demonstration.

Since a deterministic optimization problem with perfect model knowledge is typically assumed, an exact solution to the optimality conditions is generally feasible only if the demonstrations are indeed optimal and the modeling assumptions met. To account for suboptimal execution and noisy data, the optimality conditions are therefore relaxed, and an approximate solution is found that explains the demonstrated behavior as closely as possible.

The proposed methods differ in the chosen optimal control model and resulting optimality conditions, as well as in the relaxation of the optimality conditions. For the case of unconstrained linear system dynamics, quadratic cost function, and infinite horizon $N \to \infty$, the optimality conditions are given by the Riccati equation, corresponding to LQR control, which was investigated by Priess et al. (114) and Menner & Zeilinger (115). For more general cases, optimality conditions are given by the Karush–Kuhn–Tucker (KKT) conditions of the form

$$\nabla_U \mathcal{L}(U, \theta_l)|_{U=U_\mathrm{t}} = 0, \tag{17a}$$

$$\lambda^T g(U_\mathrm{t}) = 0, \tag{17b}$$

$$\lambda \geq 0, \tag{17c}$$

where $\mathcal{L}(U, \theta_l)$ is the Lagrangian, λ is the Lagrange multiplier, and $g(U) \leq 0$ defines the constraint set of the optimal control problem. The approach described by Englert et al. (116) addresses nonlinear system dynamics and complex manipulation tasks such as the interaction with objects, which is incorporated using additional contact constraints. The method presented by Menner et al. (117) similarly considers nonlinear system dynamics in the infinite-horizon case, including manipulation experiments, with the goal of predicting human movements. Both methods use a relaxation of the optimality conditions and minimize $\|\nabla_U \mathcal{L}(U, \theta_l)|_{U=U_\mathrm{t}}\|$ instead of enforcing Equation 17a, yielding a convex optimization problem if the cost function is linear in its parameters θ_l. Aswani et al. (118) proposed an inverse formulation resulting in a bilevel optimization problem for convex forward problems in Equation 16, taking noise in the demonstrations explicitly into consideration. Instead of relaxing the KKT conditions, the approach computes an input sequence U that exactly satisfies Equation 17 and minimizes the distance to the demonstration, i.e., $\min_U |U_\mathrm{t} - U|$ subject to Equation 17. The resulting parameter estimates were shown to be statistically consistent (118); however, the approach is computationally more demanding.

A related field of research is inverse reinforcement learning. This technique similarly addresses the problem of identifying a cost or reward function, typically in the context of probabilistic decision-making, which is often expressed in terms of Markov decision processes (see, e.g., 119–122). While such frameworks facilitate the interpretation of model-dependent optimality conditions in a probabilistic sense, where the unknown parameters are, e.g., obtained using likelihood maximization, they are usually formulated for discrete state and action spaces and do not take system constraints explicitly into account. Menner et al. (123) achieved nondeterministic decision-making with a continuous state space by introducing a probabilistic control objective, which allows for formulating the forward control problem in a Kalman filter framework and the corresponding inverse problem as maximum likelihood estimation.

Few results address the problem of inferring constraints (simultaneously with the cost) from demonstrations. Conceptually, the idea is again to leverage the optimality conditions. Chou et al. (124) presented a sampling-based method that learns constraints from demonstrations for a given objective function $l(x, u, \theta_l)$ and known parameters θ_l. The approach uses sampled trajectories of the unconstrained forward optimal control problem and compares them with demonstrations (which by definition satisfy the constraints) to infer the presence of constraints from the deviation of both trajectories. Menner et al. (117) identified constraints and cost simultaneously in a two-step procedure. First, candidate constraints are constructed using the convex hull of the state and the input trajectories. Second, the Lagrange multipliers λ of the KKT conditions (Equation 17) are utilized to identify constraints $\theta_{\mathcal{U}}, \theta_{\mathcal{X}}$ from the candidate constraint set, which were active during a demonstration.

5. MODEL PREDICTIVE CONTROL FOR SAFE LEARNING

While general learning-based control, particularly reinforcement learning, has shown great success in solving complex and high-dimensional control tasks (see, e.g., 125, 126), most techniques cannot ensure that safety constraints under physical limitations are met, particularly during learning iterations. To address this limitation, safety frameworks emerged from control theory (127), handling safety using a model-based controller when necessary, while otherwise a learning-based control method is used to optimize the overall cost J_{t} (Equation 2). MPC techniques can be used for such safety filters (128) to turn a safety-critical dynamical system into an inherently safe system to which any learning-based controller without safety certificates can be applied out of the box (see also **Figure 3**).

The separation of constraint satisfaction provided by an MPC approach and optimization of performance via direct use of learning-based methods, such as reinforcement learning, is motivated by the complexity of the stochastic optimal control problem, as described in Section 2.1. A direct optimization of the true cost J_{t} in an MPC often cannot be performed, and the cost would have to be approximated (see also Section 4.1), e.g., if J_{t} describes complex tasks requiring extensive planning horizons, is discontinuous, or is given as a black-box function (i.e., accessible via numerical function evaluations only).

To alleviate these restrictions, the idea is to address the solution to the stochastic optimal control problem through learning-based control methods, such as stochastic policy search or approximate dynamic programming (16). The proposed learning-based control input $u_{\mathcal{L}}(k)$ at time k is then verified in terms of safety by computing a safe backup trajectory from the one-step predicted state $x_{1|k}$ to a safe terminal set $\mathcal{X}_f$ or by modifying $u_{\mathcal{L}}(k)$ as little as possible while still providing a

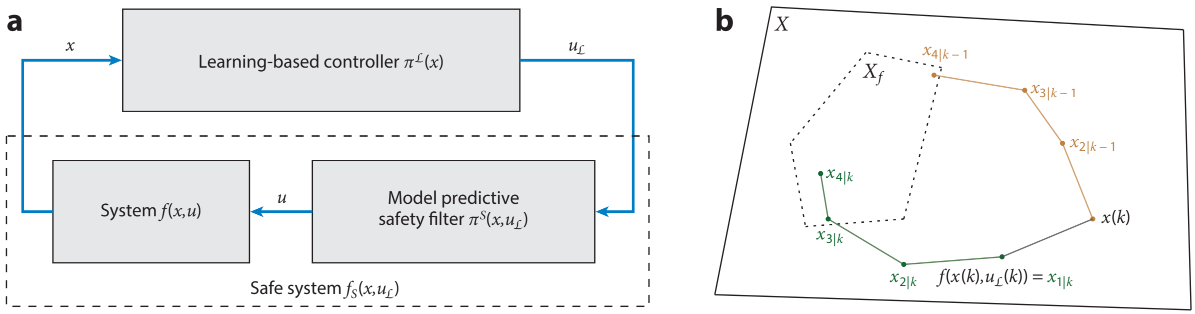

Figure 3

(*a*) The concept of a model predictive safety filter. Based on the current state x, a learning-based controller provides an input $u_{\mathcal{L}} = \pi^{\mathcal{L}}(x) \in \mathbb{R}^m$, which is processed by the safety filter $u = \pi^{\mathcal{S}}(x, u_{\mathcal{S}})$ and applied to the real system. (*b*) Inner working of a model predictive safety filter. The current system state $x(k)$ is shown with a safe backup plan (*brown*) from the solution at time $k-1$. An arbitrary learning input $u_{\mathcal{L}}(k)$ is passed through unfiltered if a feasible backup trajectory (*green*) can be obtained from the resulting $f(x(k), u_{\mathcal{L}}(k))$ via the optimization problem represented by Equation 18. Panel *a* adapted from Reference 128; panel *b* adapted from Reference 73.

safe backup trajectory. The optimization problem necessary for validating safety of the input $u_{\mathcal{L}}(k)$ is computationally cheaper than a direct optimization of the task J_t and can often be carried out over a reasonably short horizon.

More precisely, the model predictive safety filter $\pi^{\mathcal{S}}$ in **Figure 3*a*** is realized through an MPC-like optimization problem of the form

$$\begin{aligned} \underset{U}{\text{minimize}} \quad & \|u_{0|k} - u_{\mathcal{L}}(k)\| \\ \text{subject to} \quad & \text{Equations 6b–f} \end{aligned} \tag{18.}$$

with the goal of obtaining a safe backup state and input sequence with a small deviation between the first predicted input $u_{0|k}$ and the currently requested learning-based control input $u_{\mathcal{L}}(k)$. Choosing $\pi^{\mathcal{S}}(k) = u^*_{0|k}$ implements the desired filtering property—i.e., in the case that, for a given state $x(k)$ and proposed learning-based input $u_{\mathcal{L}}(k)$, all safety-relevant constraints from Equation 6 can be satisfied for all future times, we have $u^*_{0|k} = u_{\mathcal{L}}(k)$, and the learning-based control input $u_{\mathcal{L}}$ is safely passed through the filter (compare with **Figure 3*b***). If this solution does not exist, then the input $u_{\mathcal{L}}$ is deemed unsafe, and the filter will output a modified control input $\pi^{\mathcal{S}} = u^*_{0|k}$, which differs from $u_{\mathcal{L}}(k)$ as little as possible while still providing a backup trajectory from the resulting $x(k+1)$ that ensures safety. Importantly, by forcing the terminal predicted state to be contained inside an appropriately chosen terminal set $\mathcal{X}_f$ via Equation 6e, one can directly use basic MPC arguments to prove safety for all future times. In contrast to standard MPC, the terminal set $\mathcal{X}_f$ is thereby not coupled to a terminal cost function mimicking the tail of the underlying optimal control problem. Instead, it can consist of any explicit safe set together with a safety controller, which guarantees infinite constraint satisfaction once the system enters the safe set, as proposed in References 129–131. The data-driven computation of such safe sets together with safety controllers has been studied for different classes of systems (see, e.g., 132–135).

The safety filter in Equation 18 neglects model uncertainties, as similarly considered by Gurriet et al. (136) and Bastani (137). MPC techniques with model learning, as discussed in Section 3.2, can be used to facilitate the application of model predictive safety filters in the common case where the model is uncertain and needs to be learned from data. This problem was investigated by Wabersich and colleagues (73, 138) for approximately linear stochastic systems, along similar lines as the methods presented in Section 3.2. To support safe exploration of nonlinear system dynamics beyond available data, Koller et al. (72) and Wabersich & Zeilinger (128) developed methods that are robust in probability (compare with Section 3.2) and are capable of treating state- and input-dependent model uncertainties at a prespecified probability level.

While virtually any (learning-based) controller can be enhanced with rigorous safety guarantees using model predictive safety filters, the resulting performance of the overall system—i.e., the combination of the safety filter with different types of learning-based control approaches—remains to be investigated.

6. DISCUSSION

Opportunities for using learning and data-driven techniques in MPC are abundant, and while this review presents some major directions, we do not claim to provide an exhaustive discussion. This review has focused on highlighting mechanisms that are utilized when integrating learning in MPC and discussing their benefits and potential limitations as well as the resulting theoretical properties. Before analyzing common questions, we provide brief conclusions about the three categories of learning in MPC:

- Learning the system dynamics: Most research efforts have focused on an automated improvement of the prediction model in MPC, addressing challenges such as costly system identification, fine-tuning for performance during operation, or adjustment to changing or aging plants. The theoretical research regarding model adaptation is founded mostly in established results from robust MPC. Considering the success of stochastic methods in system identification and regression, this leads to a gap that appears to be rooted largely in the technical difficulties of deriving theoretical properties for stochastic models. The focus in integrating stochastic prediction models has therefore been on practical or performance considerations, leaving large potential for theoretical results in a general stochastic MPC context. While most learning-based MPC methods implement passive learning, an important open question is the realization of active learning and dual control in constrained systems.
- Learning the controller design: Data-based design or improvement of MPC formulations based on observed closed-loop behavior emerges as an important domain where learning concepts can contribute to the field of predictive control. While the design of terminal components is comparably well studied and has been successfully demonstrated, there are many open problems related to more general automatic tuning of the cost and constraints. Techniques such as Bayesian optimization have been applied out of the box, and there is great potential in the development of concepts that are tailored to MPC, i.e., that exploit the problem structure and properties. First results in inverse optimal control motivate this relatively unexplored direction as an alternative to inverse reinforcement learning, with the ability to address constraint satisfaction. Open challenges include the systematic consideration of suboptimality and probabilistic concepts.
- MPC for safe learning: Safety is recognized as a central issue of learning-based control throughout domains. Safety filters offer a powerful concept but have suffered from scalability limitations. The combination of MPC to ensure constraint satisfaction with any learning-based controller to optimize performance enables this framework for complex systems and can therefore be a stepping stone toward an increased practical impact. All developments in learning-based MPC for model or constraint learning can be similarly leveraged in an MPC filter. An important issue to be addressed is the investigation and theoretical analysis of the resulting closed-loop system under such a safety filter.

Throughout the learning-based MPC techniques discussed in this review, general concepts and challenges are emerging in different areas. There is a tendency to disconnect performance optimization from constraint satisfaction and safety, which is most evident in MPC for safe learning but also common in model-learning MPC methods. While this split can negatively affect performance, it simplifies the theory and is often computationally beneficial by avoiding costly min-max or expected value optimization. Similarly, there is a general tension between stochastic methods from machine learning and deterministic or robust approaches in control. Many rigorous results in control and optimization struggle with encountered noise and modeling uncertainty. Initial results, however, motivate further research toward stochastic learning-based MPC. Considering developments thus far, the real-time capability of learning-based MPC has not been the center of attention but will become critical for its ultimate success. Stochastic formulations in particular, but also techniques such as set-membership methods, can result in prohibitive computation or require sophisticated hardware. While there has been some success in combating growing model complexity through suitable approximations, there is great potential in the development of control methods that are computationally elegant, efficient, and reliable.

Learning-based MPC has been demonstrated to result in competitive high-performance control systems and has the potential to reduce modeling and engineering effort in controller design. Recent successes and growing interest should motivate the further development of a systematic body of theory, advanced methodologies, and computational methods that can establish learning-based MPC as a powerful tool, bringing the large potential of learning into safety-critical control applications.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to acknowledge support from the Swiss National Science Foundation under grant number PP00P2_157601/1.

LITERATURE CITED

1. Morari M, Lee JH. 1999. Model predictive control: past, present and future. *Comput. Chem. Eng.* 23:667–82
2. Darby ML, Nikolaou M. 2012. MPC: current practice and challenges. *Control Eng. Pract.* 20:328–42
3. Bemporad A, Morari M. 1999. Robust model predictive control: a survey. In *Robustness in Identification and Control*, ed. A Garulli, A Tesi, pp. 207–26. London: Springer
4. Mesbah A. 2016. Stochastic model predictive control: an overview and perspectives for future research. *IEEE Control Syst. Mag.* 36(6):30–44
5. Lee JH, Lee KS. 2007. Iterative learning control applied to batch processes: an overview. *Control Eng. Pract.* 15:1306–18
6. Cueli JR, Bordons C. 2008. Iterative nonlinear model predictive control. Stability, robustness and applications. *Control Eng. Pract.* 16:1023–34
7. Parisini T, Zoppoli R. 1995. A receding-horizon regulator for nonlinear systems and a neural approximation. *Automatica* 31:1443–51
8. Domahidi A, Zeilinger MN, Morari M, Jones CN. 2011. Learning a feasible and stabilizing explicit model predictive control law by robust optimization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 513–19. Piscataway, NJ: IEEE
9. Lucia S, Karg B. 2018. A deep learning-based approach to robust nonlinear model predictive control. In *6th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2018*, pp. 511–16. IFAC-PapersOnLine 51(20). Amsterdam: Elsevier
10. Hertneck M, Köhler J, Trimpe S, Allgöwer F. 2018. Learning an approximate model predictive controller with guarantees. *IEEE Control Syst. Lett.* 2:543–48
11. Wittenmark B. 1995. Adaptive dual control methods: an overview. In *5th IFAC Symposium on Adaptive Systems in Control and Signal Processing*, pp. 67–72. IFAC Proc. Vol. 28(13). Amsterdam: Elsevier
12. Mesbah A. 2018. Stochastic model predictive control with active uncertainty learning: a survey on dual control. *Annu. Rev. Control* 45:107–17
13. Campi MC, Garatti S. 2018. *Introduction to the Scenario Approach*. Philadelphia: Soc. Ind. Appl. Math.
14. Fagiano L, Schildbach G, Tanaskovic M, Morari M. 2015. Scenario and adaptive model predictive control of uncertain systems. *5th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2015*, pp. 352–59. IFAC-PapersOnLine 48(23). Amsterdam: Elsevier
15. Rosolia U, Zhang X, Borrelli F. 2018. Data-driven predictive control for autonomous systems. *Annu. Rev. Control Robot. Auton. Syst.* 1:259–86

16. Recht B. 2019. A tour of reinforcement learning: the view from continuous control. *Annu. Rev. Control Robot. Auton. Syst.* 2:253–79
17. Bertsekas DP. 2012. *Dynamic Programming and Optimal Control*, Vol. 2: *Approximate Dynamic Programming*. Belmont, MA: Athena Sci. 4th ed.
18. Bertsekas DP. 2017. *Dynamic Programming and Optimal Control*, Vol. 1. Belmont, MA: Athena Sci. 4th ed.
19. Mayne DQ, Rawlings JB, Rao CV, Scokaert POM. 2000. Constrained model predictive control: stability and optimality. *Automatica* 36:789–814
20. Rawlings JB, Mayne DQ, Diehl MM. 2017. *Model Predictive Control: Theory, Computation, and Design*. Madison, WI: Nob Hill. 2nd ed.
21. Bishop C. 2006. *Pattern Recognition and Machine Learning*. New York: Springer
22. Piché S, Sayyar-Rodsari B, Johnson D, Gerules M. 2000. Nonlinear model predictive control using neural networks. *IEEE Control Syst. Mag.* 20(3):53–62
23. Nagabandi A, Clavera I, Liu S, Fearing RS, Abbeel P, et al. 2019. *Learning to adapt in dynamic, real-world environments through meta-reinforcement learning*. Paper presented at the Seventh International Conference on Learning Representations, New Orleans, LA, May 6–9. **https://openreview.net/forum?id=HyztsoC5Y7**
24. Yang H, Li S. 2015. A data-driven predictive controller design based on reduced Hankel matrix. In *2015 10th Asian Control Conference*. Piscataway, NJ: IEEE. **https://doi.org/10.1109/ASCC.2015.7244723**
25. Coulson J, Lygeros J, Dörfler F. 2019. Data-enabled predictive control: in the shallows of the DeePC. In *2019 18th European Control Conference*, pp. 307–12. Piscataway, NJ: IEEE
26. Feldbaum AA. 1960. Dual control theory. I. *Avtom. Telemekhanika* 21:1240–49
27. Feldbaum AA. 1960. Dual control theory. II. *Avtom. Telemekhanika* 21:1453–64
28. Heirung TAN, Foss B, Ydstie BE. 2015. MPC-based dual control with online experiment design. *J. Process Control* 32:64–76
29. Alpcan T, Shames I. 2015. An information-based learning approach to dual control. *IEEE Trans. Neural Netw. Learn. Syst.* 26:2736–48
30. Genceli H, Nikolaou M. 1996. New approach to constrained predictive control with simultaneous model identification. *AIChE J.* 42:2857–68
31. Marafioti G, Bitmead RR, Hovd M. 2014. Persistently exciting model predictive control. *Int. J. Adapt. Control Signal Process.* 28:536–52
32. Hanssen KG, Foss B. 2015. Scenario based implicit dual model predictive control. In *5th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2015*, pp. 416–21. IFAC-PapersOnLine 48(23). Amsterdam: Elsevier
33. Thangavel S, Lucia S, Paulen R, Engell S. 2018. Dual robust nonlinear model predictive control: a multi-stage approach. *J. Process Control* 72:39–51
34. Lee JM, Lee JH. 2009. An approximate dynamic programming based approach to dual adaptive control. *J. Process Control* 19:859–64
35. Heirung TAN, Ydstie BE, Foss B. 2017. Dual adaptive model predictive control. *Automatica* 80:340–48
36. Klenske ED, Hennig P. 2016. Dual control for approximate Bayesian reinforcement learning. *J. Mach. Learn. Res.* 17:1–30
37. Aswani A, Gonzalez H, Sastry SS, Tomlin CJ. 2013. Provably safe and robust learning-based model predictive control. *Automatica* 49:1216–26
38. Bouffard P, Aswani A, Tomlin CJ. 2012. Learning-based model predictive control on a quadrotor: onboard implementation and experimental results. In *2012 IEEE International Conference on Robotics and Automation*, pp. 279–84. Piscataway, NJ: IEEE
39. Aswani A, Bouffard P, Tomlin CJ. 2012. Extensions of learning-based model predictive control for real-time application to a quadrotor helicopter. In *2012 American Control Conference*, pp. 4661–66. Piscataway, NJ: IEEE
40. Aswani A, Master N, Taneja J, Culler D, Tomlin CJ. 2012. Reducing transient and steady state electricity consumption in HVAC using learning-based model-predictive control. *Proc. IEEE* 100:240–53

41. Aswani A, Master N, Taneja J, Krioukov A, Culler D, Tomlin CJ. 2012. Energy-efficient building HVAC control using hybrid system LBMPC. In *4th IFAC Conference on Nonlinear Model Predictive Control*, pp. 496–501. IFAC Proc. Vol. 45(17). Amsterdam: Elsevier
42. Milanese M, Vicino A. 1991. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica* 27:997–1009
43. Milanese M, Novara C. 2004. Set membership identification of nonlinear systems. *Automatica* 40:957–75
44. Veres SM, Messaoud H, Norton JP. 1999. Limited-complexity model-unfalsifying adaptive tracking-control. *Int. J. Control* 72:1417–26
45. Tanaskovic M, Fagiano L, Smith R, Morari M. 2014. Adaptive receding horizon control for constrained MIMO systems. *Automatica* 50:3019–29
46. Tanaskovic M, Sturzenegger D, Smith R, Morari M. 2017. Robust adaptive model predictive building climate control. In *20th IFAC World Congress*, pp. 1871–76. IFAC-PapersOnLine 50(1). Amsterdam: Elsevier
47. Tanaskovic M, Fagiano L, Gligorovski V. 2019. Adaptive model predictive control for linear time varying MIMO systems. *Automatica* 105:237–45
48. Bujarbaruah M, Zhang X, Borrelli F. 2018. Adaptive MPC with chance constraints for FIR systems. arXiv:1804.09790 [cs.SY]
49. Bujarbaruah M, Zhang X, Rosolia U, Borrelli F. 2018. Adaptive MPC for iterative tasks. In *2018 IEEE Conference on Decision and Control*, pp. 6322–27. Piscataway, NJ: IEEE
50. Terzi E, Fagiano L, Farina M, Scattolini R. 2018. Learning multi-step prediction models for receding horizon control. In *2018 European Control Conference*, pp. 1335–40. Piscataway, NJ: IEEE
51. Terzi E, Farina M, Fagiano L, Scattolini R. 2018. Robust predictive control with data-based multi-step prediction models. In *2018 European Control Conference*, pp. 1710–15. Piscataway, NJ: IEEE
52. Terzi E, Fagiano L, Farina M, Scattolini R. 2019. Learning-based predictive control for linear systems: a unitary approach. *Automatica* 108:108473
53. Di Cairano S. 2016. Indirect adaptive model predictive control for linear systems with polytopic uncertainty. In *2016 American Control Conference*, pp. 3570–75. Piscataway, NJ: IEEE
54. Zhou J, Di Cairano S, Danielson C. 2017. Indirect adaptive MPC for output tracking of uncertain linear polytopic systems. In *2017 American Control Conference*, pp. 3054–59. Piscataway, NJ: IEEE
55. Lorenzen M, Allgöwer F, Cannon M. 2017. Adaptive model predictive control with robust constraint satisfaction. In *20th IFAC World Congress*, pp. 3313–18. IFAC-PapersOnLine 50(1). Amsterdam: Elsevier
56. Lorenzen M, Cannon M, Allgöwer F. 2019. Robust MPC with recursive model update. *Automatica* 103:461–71
57. Fukushima H, Kim TH, Sugie T. 2007. Adaptive model predictive control for a class of constrained linear systems based on the comparison model. *Automatica* 43:301–8
58. Kim TH, Sugie T. 2008. Adaptive receding horizon predictive control for constrained discrete-time linear systems with parameter uncertainties. *Int. J. Control* 81:62–73
59. Adetola V, DeHaan D, Guay M. 2009. Adaptive model predictive control for constrained nonlinear systems. *Syst. Control Lett.* 58:320–26
60. Adetola V, Guay M. 2011. Robust adaptive MPC for constrained uncertain nonlinear systems. *Int. J. Adapt. Control Signal Process.* 25:155–67
61. Gonçalves GAA, Guay M. 2016. Robust discrete-time set-based adaptive predictive control for nonlinear systems. *J. Process Control* 39:111–22
62. Beliakov G. 2006. Interpolation of Lipschitz functions. *J. Comput. Appl. Math.* 196:20–44
63. Calliess JP. 2014. *Conservative decision-making and inference in uncertain dynamical systems*. PhD Thesis, Univ. Oxford, Oxford, UK
64. Canale M, Fagiano L, Signorile MC. 2011. Robust design of predictive controllers using set membership identified models. In *18th IFAC World Congress*, pp. 13414–19. IFAC Proc. Vol. 44(1). Amsterdam: Elsevier

65. Canale M, Fagiano L, Signorile MC. 2014. Nonlinear model predictive control from data: a set membership approach. *Int. J. Robust Nonlinear Control* 24:123–39
66. Limon D, Calliess J, Maciejowski JM. 2017. Learning-based nonlinear model predictive control. In *20th IFAC World Congress*, pp. 7769–76. IFAC-PapersOnLine 50(1). Amsterdam: Elsevier
67. Calliess JP, Roberts S, Rasmussen C, Maciejowski J. 2018. Nonlinear set membership regression with adaptive hyper-parameter estimation for online learning and control. In *2018 European Control Conference*. Piscataway, NJ: IEEE. **https://doi.org/10.23919/ECC.2018.8550439**
68. Manzano JM, Limon D, de la Peña DM, Calliess J. 2018. Robust data-based model predictive control for nonlinear constrained systems. In *6th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2018*, pp. 505–10. IFAC-PapersOnLine 51(20). Amsterdam: Elsevier
69. Manzano JM, Limón D, de la Peña DM, Calliess J. 2019. Output feedback MPC based on smoothed projected kinky inference. *IET Control Theory Appl.* 13:795–805
70. Maddalena ET, Jones CN. 2019. *Learning non-parametric models with guarantees: a smooth Lipschitz interpolation approach*. Tech. Rep., École Polytech. Féd. Lausanne, Lausanne, Switz.
71. Soloperto R, Müller MA, Trimpe S, Allgöwer F. 2018. Learning-based robust model predictive control with state-dependent uncertainty. In *6th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2018*, pp. 442–47. IFAC-PapersOnLine 51(20). Amsterdam: Elsevier
72. Koller T, Berkenkamp F, Turchetta M, Krause A. 2018. Learning-based model predictive control for safe exploration. In *2018 IEEE Conference on Decision and Control*, pp. 6059–66. Piscataway, NJ: IEEE
73. Wabersich KP, Hewing L, Carron A, Zeilinger MN. 2019. Probabilistic model predictive safety certification for learning-based control. arXiv:1906.10417 [eess.SY]
74. Paulson JA, Santos TLM, Mesbah A. 2019. Mixed stochastic-deterministic tube MPC for offset-free tracking in the presence of plant-model mismatch. *J. Process Control* 83:102–20
75. Calafiore G, Campi MC. 2005. Uncertain convex programs: randomized solutions and confidence levels. *Math. Program.* 102:25–46
76. Campi MC, Garatti S. 2011. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *J. Optim. Theory Appl.* 148:257–80
77. Desaraju VR, Michael N. 2016. Experience-driven predictive control. In *Proceedings of the 1st International Workshop on Robot Learning and Planning*, pp. 29–37. N.p.: RLP
78. Desaraju VR, Spitzer AE, Michael N. 2017. Experience-driven predictive control with robust constraint satisfaction under time-varying state uncertainty. In *Robotics: Science and Systems XIII*, ed. N Amato, S Srinivasa, N Ayanian, S Kuindersma, pap. 67. N.p.: Robot. Sci. Syst. Found.
79. McKinnon CD, Schoellig AP. 2019. Learn fast, forget slow: safe predictive control for systems with locally linear actuator dynamics performing repetitive tasks. *IEEE Robot. Autom. Lett.* 4:2180–87
80. Schildbach G, Fagiano L, Frei C, Morari M. 2014. The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations. *Automatica* 50:3009–18
81. Calafiore GC, Fagiano L. 2013. Stochastic model predictive control of LPV systems via scenario optimization. *Automatica* 49:1861–66
82. Lorenzen M, Dabbene F, Tempo R, Allgöwer F. 2017. Stochastic MPC with offline uncertainty sampling. *Automatica* 81:176–83
83. Hewing L, Zeilinger MN. 2019. Scenario-based probabilistic reachable sets for recursively feasible stochastic model predictive control. *IEEE Control Syst. Lett.* 4:450–55
84. Di Cairano S, Bernardini D, Bemporad A, Kolmanovsky IV. 2014. Stochastic MPC with learning for driver-predictive vehicle control and its application to HEV energy management. *IEEE Trans. Control Syst. Technol.* 22:1018–31
85. Rasmussen CE, Williams CKI. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press
86. Hewing L, Kabzan J, Zeilinger MN. 2019. Cautious model predictive control using Gaussian process regression. *IEEE Trans. Control Syst. Technol.* In press. **https://doi.org/10.1109/TCST.2019.2949757**
87. Ostafew CJ, Schoellig AP, Barfoot TD. 2016. Robust constrained learning-based NMPC enabling reliable mobile robot path tracking. *Int. J. Robot. Res.* 35:1547–63

88. Kamthe S, Deisenroth MP. 2018. Data-efficient reinforcement learning with probabilistic model predictive control. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, ed. A Storkey, F Perez-Cruz, pp. 1701–10. Proc. Mach. Learn. Res. 84. N.p.: PMLR
89. Maiworm M, Limon D, Manzano JM, Findeisen R. 2018. Stability of Gaussian process learning based output feedback model predictive control. In *6th IFAC Conference on Nonlinear Model Predictive Control: NMPC 2018*, pp. 455–61. IFAC-PapersOnLine 51(20). Amsterdam: Elsevier
90. Kabzan J, Hewing L, Liniger A, Zeilinger MN. 2019. Learning-based model predictive control for autonomous racing. *IEEE Robot. Autom. Lett.* 4:3363–70
91. Kabzan J, Hewing L, Liniger A, Zeilinger MN. 2019. MPCC with Gaussian process based model learning for autonomous racing. *YouTube*, Sept. 3. **https://www.youtube.com/watch?v=-cdXw1MyTUA**
92. Kabzan J, Hewing L, Liniger A, Zeilinger MN. 2019. Learning-based model predictive control for autonomous racing. *YouTube*, July 11. **https://www.youtube.com/watch?v=bjlT-6KVQ7U**
93. Quiñonero-Candela J, Rasmussen CE. 2005. A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* 6:1939–59
94. Lázaro-Gredilla M, Quiñonero-Candela J, Rasmussen CE, Figueiras-Vidal AR. 2010. Sparse spectrum Gaussian process regression. *J. Mach. Learn. Res.* 11:1865–81
95. Carron A, Arcari E, Wermelinger M, Hewing L, Hutter M, Zeilinger MN. 2019. Data-driven model predictive control for trajectory tracking with a robotic arm. *IEEE Robot. Autom. Lett.* 4:3758–65
96. Ostafew CJ, Collier J, Schoellig AP, Barfoot TD. 2016. Learning-based nonlinear model predictive control to improve vision-based mobile robot path tracking. *J. Field Robot.* 33:133–52
97. Pan Y, Yan X, Theodorou EA, Boots B. 2017. Prediction under uncertainty in sparse spectrum Gaussian processes with applications to filtering and control. In *Proceedings of the 34th International Conference on Machine Learning*, ed. D Precup, YW Teh, pp. 2760–68. Proc. Mach. Learn. Res. 70. N.p.: PMLR
98. Kocijan J, Murray-Smith R, Rasmussen CE, Girard A. 2004. Gaussian process model based predictive control. In *Proceedings of the 2004 American Control Conference*, pp. 2214–19. Piscataway, NJ: IEEE
99. Klenske ED, Zeilinger MN, Schölkopf B, Hennig P. 2016. Gaussian process-based predictive control for periodic error correction. *IEEE Trans. Control Syst. Technol.* 24:110–21
100. Jain A, Nghiem TX, Morari M, Mangharam R. 2018. Learning and control using Gaussian processes: towards bridging machine learning and controls for physical systems. In *Proceedings of the 9th ACM/IEEE International Conference on Cyber-Physical Systems*, pp. 140–49. Piscataway, NJ: IEEE
101. Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. 2016. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* 104:148–75
102. Berkenkamp F, Schoellig AP, Krause A. 2016. Safe controller optimization for quadrotors with Gaussian processes. In *2016 IEEE International Conference on Robotics and Automation*, pp. 491–96. Piscataway, NJ: IEEE
103. Neumann-Brosig M, Marco A, Schwarzmann D, Trimpe S. 2019. Data-efficient auto-tuning with Bayesian optimization: an industrial control study. *IEEE Trans. Control Syst. Technol.* In press
104. Marco A, Hennig P, Bohg J, Schaal S, Trimpe S. 2016. Automatic LQR tuning based on Gaussian process global optimization. In *2016 IEEE International Conference on Robotics and Automation*, pp. 270–77. Piscataway, NJ: IEEE
105. Fröhlich LP, Klenske ED, Daniel CG, Zeilinger MN. 2019. Bayesian optimization for policy search in high-dimensional systems via automatic domain selection. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 757–64. Piscataway, NJ: IEEE
106. Bansal S, Calandra R, Xiao T, Levine S, Tomlin CJ. 2017. Goal-driven dynamics learning via Bayesian optimization. In *2017 IEEE 56th Annual Conference on Decision and Control*, pp. 5168–73. Piscataway, NJ: IEEE
107. Piga D, Forgione M, Formentin S, Bemporad A. 2019. Performance-oriented model learning for data-driven MPC design. *IEEE Control Syst. Lett.* 3:577–82
108. Gros S, Zanon M. 2020. Data-driven economic NMPC using reinforcement learning. *IEEE Trans. Autom. Control* 65:636–48
109. Blanchini F, Pellegrino FA. 2005. Relatively optimal control: the static solution. In *16th IFAC World Congress*, pp. 676–81. IFAC Proc. Vol. 38(1). Amsterdam: Elsevier

110. Brunner FD, Lazar M, Allgöwer F. 2015. Stabilizing model predictive control: on the enlargement of the terminal set. *Int. J. Robust Nonlinear Control* 25:2646–70
111. Rosolia U, Borrelli F. 2018. Learning model predictive control for iterative tasks. A data-driven control framework. *IEEE Trans. Autom. Control* 63:1883–96
112. Rosolia U, Zhang X, Borrelli F. 2017. Robust learning model predictive control for uncertain iterative tasks: learning from experience. In *2017 IEEE 56th Annual Conference on Decision and Control*, pp. 1157–62. Piscataway, NJ: IEEE
113. Rosolia U, Borrelli F. 2019. Learning how to autonomously race a car: a predictive control approach. arXiv:1901.08184 [cs.SY]
114. Priess MC, Conway R, Choi J, Popovich JM, Radcliffe C. 2014. Solutions to the inverse LQR problem with application to biological systems analysis. *IEEE Trans. Control Syst. Technol.* 23:770–77
115. Menner M, Zeilinger MN. 2018. Convex formulations and algebraic solutions for linear quadratic inverse optimal control problems. In *2018 European Control Conference*, pp. 2107–12. Piscataway, NJ: IEEE
116. Englert P, Vien NA, Toussaint M. 2017. Inverse KKT: learning cost functions of manipulation tasks from demonstrations. *Int. J. Robot. Res.* 36:1474–88
117. Menner M, Worsnop P, Zeilinger MN. 2019. Constrained inverse optimal control with application to a human manipulation task. *IEEE Trans. Control Syst. Technol.* In press
118. Aswani A, Shen ZJ, Siddiq A. 2018. Inverse optimization with noisy data. *Oper. Res.* 66:870–92
119. Abbeel P, Ng AY. 2004. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, pap. 1. New York: ACM
120. Ziebart BD, Maas A, Bagnell JA, Dey AK. 2008. Maximum entropy inverse reinforcement learning. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pp. 1433–38. Palo Alto, CA: AAAI Press
121. Argall BD, Chernova S, Veloso M, Browning B. 2009. A survey of robot learning from demonstration. *Robot. Auton. Syst.* 57:469–83
122. Finn C, Levine S, Abbeel P. 2016. Guided cost learning: deep inverse optimal control via policy optimization. In *Proceedings of the 33rd International Conference on Machine Learning*, ed, MF Balcan, KQ Weinberger, pp. 49–58. Proc. Mach. Learn. Res. 48. N.p.: PMLR
123. Menner M, Berntorp K, Zeilinger MN, Di Cairano S. 2019. Inverse learning for human-adaptive motion planning. In *2019 IEEE Conference on Decision and Control*. Piscataway, NJ: IEEE. Forthcoming
124. Chou G, Berenson D, Ozay N. 2018. Learning constraints from demonstrations. arXiv:1812.07084 [cs.RO]
125. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518:529–33
126. Levine S, Finn C, Darrell T, Abbeel P. 2016. End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.* 17:1334–73
127. Gillulay JH, Tomlin CJ. 2011. Guaranteed safe online learning of a bounded system. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2979–84. Piscataway, NJ: IEEE
128. Wabersich KP, Zeilinger MN. 2018. Safe exploration of nonlinear dynamical systems: a predictive safety filter for reinforcement learning. arXiv:1812.05506 [cs.SY]
129. Seto D, Krogh B, Sha L, Chutinan A. 1998. The simplex architecture for safe online control system upgrades. In *Proceedings of the 1998 American Control Conference*, pp. 3504–8. Piscataway, NJ: IEEE
130. Wieland P, Allgöwer F. 2007. Constructive safety using control barrier functions. In *7th IFAC Symposium on Nonlinear Control Systems*, pp. 462–67. IFAC Proc. Vol. 40(12). Amsterdam: Elsevier
131. Wieland P, Ebenbauer C, Allgöwer F. 2007. Ensuring task-independent safety for multi-agent systems by feedback. In *2007 American Control Conference*, pp. 3880–85. Piscataway, NJ: IEEE
132. Berkenkamp F, Turchetta M, Schoellig AP, Krause A. 2017. Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems 30*, ed. I Guyon, UV Luxburg, S Bengio, H Wallach, R Fergus, et al., pp. 908–18. Red Hook, NY: Curran
133. Fisac JF, Akametalu AK, Zeilinger MN, Kaynama S, Gillula J, Tomlin CJ. 2019. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Trans. Autom. Control* 64:2737–52

134. Wabersich KP, Zeilinger MN. 2018. Scalable synthesis of safety certificates from data with application to learning-based control. In *2018 European Control Conference*, pp. 1691–97. Piscataway, NJ: IEEE
135. Larsen RB, Carron A, Zeilinger MN. 2017. Safe learning for distributed systems with bounded uncertainties. In *20th IFAC World Congress*, pp. 2536–42. IFAC-PapersOnLine 50(1). Amsterdam: Elsevier
136. Gurriet T, Mote M, Ames AD, Feron E. 2018. An online approach to active set invariance. In *2018 IEEE Conference on Decision and Control*, pp. 3592–99. Piscataway, NJ: IEEE
137. Bastani O. 2019. Safe reinforcement learning via online shielding. arXiv:1905.10691 [cs.LG]
138. Wabersich KP, Zeilinger MN. 2018. Linear model predictive safety certification for learning-based control. In *2018 IEEE Conference on Decision and Control*, pp. 7130–35. Piscataway, NJ: IEEE