Performance-oriented model learning for data-driven MPC design

Dario Piga, Marco Forgione, Simone Formentin, Alberto Bemporad

Abstract— **Model Predictive Control (MPC) is an enabling technology in applications requiring controlling physical processes in an optimized way under constraints on inputs and outputs. However, in MPC closed-loop performance is pushed to the limits only if the plant under control is accurately modeled; otherwise, robust architectures need to be employed, at the price of reduced performance due to worst-case conservative assumptions. In this paper, instead of adapting the controller to handle uncertainty, we adapt the learning procedure so that the prediction model is selected to provide the best closed-loop performance. More specifically, we apply for the first time the above "identification for control" rationale to hierarchical MPC using data-driven methods and Bayesian optimization.**

I. Introduction

Nowadays, *Model Predictive Control* (MPC) has become the most popular advanced control technology for several complex engineering applications [1]. Apart from computational aspects, it is widely recognized that one key practical challenge in MPC arises when dealing with uncertainty, especially when the prediction model is identified using open-loop data taken from a specific operation of the plant [2].

In case of partially known systems, traditional MPC approaches exhibit some degree of robustness, so that marginal robust performance can be guaranteed. When intrinsic robustness of deterministic MPC is not enough, robust MPC [3] and stochastic MPC [2] approaches have been developed to take into account uncertainties. However, regardless of the specific technique, increasing robustness of the MPC controller usually leads to conservative performance [4].

While there is usually a separation between model identification and control design, an alternative approach to managing uncertainty in designing control systems is to revisit the identification process as a procedure *to be designed* by bearing the final control application in mind. Such a rationale is known as *Identification for Control* (I4C) and has been widely studied for fixed-order (oftentimes, PID) control of *Linear Time-Invariant* (LTI) systems [5]. According to I4C, the best model for control may *not* be the one providing the least output prediction errors, but the one providing the best performance on the true system when in closed loop with the associated model-based controller.

As far as we are aware of, the I4C modeling approach has *never* been applied to MPC control. Learning techniques have instead been shown to be useful for iterative MPC tasks in [6] and in reinforcement learning applied to MPC [7]. Furthermore, data-driven approaches have been proposed for direct MPC optimization using open- and closed-loop data, see, e.g., [8], [9]. Although the above approaches are powerful tools for control design in case of unknown systems, they fail to provide a mathematical (albeit control-oriented) description of the plant. Indeed, the latter can often be useful for physical interpretation, performance monitoring, and diagnosis [10].

In this work, we propose an *Identification for (Model-Predictive) Control* approach aimed at finding the best prediction model for MPC from experimental data, by considering the control objective directly in the model learning phase. We propose a hierarchical architecture, typically employed in several industrial applications, in which the inner controller is a parametric filter (e.g., a PID controller) aimed to stabilize the system at a fast pace, whereas the outer loop plays the role of a reference governor (RG) [11], [12] with a twofold goal: (*i*) boosting the performance of the inner loop and (*ii*) handling the signals constraints due, e.g., to actuator bounds or system limitations. Within this framework, the RG is typically an MPC law based on a model of the inner loop. According to the I4C philosophy, we propose a change of perspective and treat such a model as a *design parameter* instead. Such a parameter will be iteratively optimized, together with the inner controller, using closed-loop data collected on the plant and Bayesian optimization. Finally, we show that, using the same rationale and tools, also the prediction horizon, a critical parameter to tune in MPC, can be optimized from data.

For the sake of completeness, the first use of Bayesian optimization in control-oriented identification was proposed in [13], based on a simpler control scheme. The same hierarchical architecture was instead addressed in [9] to design the controller from data, but without providing an MPC-oriented model of the plant.

The remainder of the paper is as follows. In Section II the control problem of interest is formally stated. The hierarchical architecture is introduced in Section III, where also the parameterization of each block is described (and motivated) in detail. The proposed strategy is described in Section IV, where a discussion on how to practically restrict the parameter space is also provided. Section V illustrates the performance of the method on a benchmark example.

II. PROBLEM FORMULATION

Consider a *multi-input multi-output* (MIMO) plant $\mathcal{S}$, with input $u \in \mathbb{R}^{n_u}$ and output $y \in \mathbb{R}^{n_y}$ signals sampled at a regular time interval T_s. We aim at synthesizing a controller $\mathcal{C}$ for $\mathcal{S}$ such that the controlled closed-loop system achieves a desired engineering objective defined in terms of minimization of a cost $J(y_{1:T}, u_{1:T})$, where $y_{1:T}$ (resp. $u_{1:T}$) denotes the sequence of output (resp. input) signals measured at time steps $t = 1, \ldots, T$, and T is the length (measured in number of samples) of the experiment where the closed-loop performance is measured. Besides minimizing the cost $J(y_{1:T}, u_{1:T})$, the following constraints on inputs and outputs should be satisfied:

$$u_{\min} \leq u(t) \leq u_{\max}, \tag{1a}$$

$$\Delta u_{\min} \leq u(t) - u(t-1) \leq \Delta u_{\max}, \tag{1b}$$

$$y_{\min} \leq y(t) \leq y_{\max}, \quad t = 1, \ldots, T. \tag{1c}$$

Constraints (1) are generally imposed by actuator limitations or might reflect safety conditions. The control design problem is formulated as the following optimization problem:

$$\min_{\mathcal{C} \in \mathbf{C}} \quad J(y_{1:T}, u_{1:T}) \quad \text{s.t.} \quad (1), \tag{2}$$

with $\mathbf{C}$ denoting the set of controller candidates.

We rewrite constraints (1) as $h_i(t) \geq 0$, $i = 1, \ldots, 6$, with

$$h_1(t) = u(t) - u_{\min}, \qquad h_2(t) = u_{\max} - u(t), \tag{3a}$$

$$h_3(t) = u(t) - u(t-1) - \Delta u_{\min} \geq 0, \tag{3b}$$

$$h_4(t) = \Delta u_{\max} - u(t) + u(t-1) \geq 0, \tag{3c}$$

$$h_5(t) = y(t) - y_{\min}, \quad h_6(t) = y_{\max} - y(t), \tag{3d}$$

and treat them with penalty functions

$$\min_{\mathcal{C} \in \mathbf{C}} \quad \tilde{J}(y_{1:T}, u_{1:T}, \mathcal{C}) \tag{4}$$

where

$$\tilde{J}(y_{1:T}, u_{1:T}) = J(y_{1:T}, u_{1:T}) + \sum_{t=1}^{T} \sum_{i=1}^{6} b_t(h_i(t)) \tag{5}$$

and $b_t : \mathbb{R} \to \mathbb{R}$ are (possibly time-varying) barrier functions.

Assuming zero initial conditions, clearly $y_{1:T}$, $u_{1:T}$ in (5) are only functions of the controller $\mathcal{C}$ and of the process model $\mathcal{S}$. Rather than first fixing a model for $\mathcal{S}$ (either from first-principle physical laws or using system identification techniques), we follow a *performance-driven* control design paradigm and leave the model of $\mathcal{S}$ as a degree of freedom, used to minimize the closed-loop cost $\tilde{J}(y_{1:T}, u_{1:T})$.

III. CONTROL ARCHITECTURE

We adopt the hierarchical, multi-rate, reference-governor control architecture in Fig. 1, consisting of:

- an inner low-level controller $\mathcal{K}$ which operates at sampling time T_s and it is mainly used to handle fast dynamics of the system. This controller introduces a degree-of-freedom in the control design and, in case of unstable plants $\mathcal{S}$, it might also stabilize the inner

closed-loop system $\mathcal{M}$. Nevertheless, the latter is not a required condition in our design approach.
- an outer MPC to enhance performance of the inner loop $\mathcal{M}$ an to enforce constraints (1c). The MPC operates at a sampling time T_{MPC} that is an integer multiple of T_s, *i.e.*, $T_{\mathrm{MPC}} = N T_s$ with $N \in \mathbb{N}$. Setting T_{MPC} larger than T_s (thus, $N > 1$) may be needed to solve the constrained optimization problem on line, *i.e.*, within the MPC sampling time T_{MPC}.

In standard RG approaches, the outer MPC requires a prediction model of the inner loop $\mathcal{M}$. In accordance with the *performance-driven* approach proposed in this paper, we treat such a model as a *design parameter* and look for the model providing the best closed-loop performance according to the performance index $\tilde{J}(y_{1:T}, u_{1:T})$. In particular, as detailed in the following, a model of the plant $\mathcal{S}$ will be used neither to design the controllers nor to evaluate the performance index $\tilde{J}(y_{1:T}, u_{1:T})$, which will be instead measured directly from closed-loop experiments performed on the actual plant.

A. Inner controller parameterization

The inner controller $\mathcal{K}$ is parameterized by a vector $\theta \in \mathbb{R}^{n_\theta}$. For instance, $\mathcal{K}$ can be a simple discrete-time *proportional-integral-derivative* (PID) controller, with sampling time T_s and discrete-time transfer function

$$\mathcal{K}(z, \theta) = \theta_P + \theta_I T_s \frac{1}{z-1} + \theta_D \frac{N_d}{1 + N_d T_s \frac{1}{z-1}}, \tag{6}$$

where $\theta = [\theta_P \ \theta_I \ \theta_D]'$ is the design parameter vector and $N_d \gg 1$ limits the high-frequency gain of the PID controller. Although N_d may be treated as a design parameter, its tuning is generally not critical and thus not included in θ.

B. Outer MPC parameterization

The most important component of the outer MPC is the model used to predict the output y and input u as a function of the MPC command g. Let M be the dynamical model from g to $\begin{bmatrix} y \\ u \end{bmatrix}$, described in the state-space representation

$$\begin{cases} \xi(t+1) &= A_M \xi(t) + B_M g(t) \\ \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} &= C_M \xi(t) + D_M g(t), \end{cases} \tag{7}$$

where $\xi \in \mathcal{R}^{n_\xi}$ is the state of the closed-loop model. For instance, in the case of a single-input-single-output plant, the 2×1 transfer matrix M can be modelled as a pair of transfer functions with the same poles. Let $\mu \in \mathbb{R}^{n_\mu}$ be the vector obtained by stacking the entries of A_M, B_M, C_M and D_M.
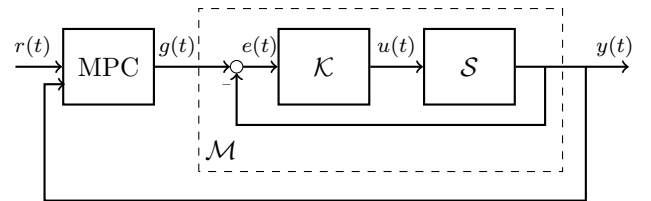

Fig. 1. Hierarchical control architecture. $\mathcal{S}$: plant to be controlled; $\mathcal{K}$: inner controller; $\mathcal{M}$: inner closed-loop system; $r(t)$: reference to be tracked.

At each time instant t integer multiple of the MPC sampling time T_{MPC} (*i.e.*, $t = hT_{\text{MPC}}$ with $h \in \mathbb{N}$), the outer MPC solves the minimization problem

$$
\min_{\{g(t+k|t)\}_{k=1}^{N_u}, \epsilon} Q_y \sum_{k=1}^{N_p} \left(y(t+k|t) - r(t+k)\right)^2 +
$$

$$
+ Q_u \sum_{k=1}^{N_p} \left(u(t+k|t) - u_{\text{ref}}(t+k)\right)^2 +
$$

$$
+ Q_{\Delta u} \sum_{k=1}^{N_p} (u(t+k|k) - u(t+k-1|t))^2 + Q_\epsilon \epsilon^2 \quad \text{(8a)}
$$

$$
\text{s.t.} \begin{bmatrix} y(t+k|t) \\ u(t+k|t) \end{bmatrix} = M\left(\mu, g(t+k|t)\right), \quad k = 1, \ldots, N_p
$$
$$\tag{8b}$$

$$
y_{\min} - V_y \epsilon \leq y(t+k|t) \leq y_{\max} + V_y \epsilon, \quad k = 1, \ldots, N_p
$$
$$\tag{8c}$$

$$
u_{\min} - V_u \epsilon \leq u(t+k|t) \leq u_{\max} + V_u \epsilon, \quad k = 1, \ldots, N_p
$$
$$\tag{8d}$$

$$
\Delta u_{\min} - V_{\Delta u} \epsilon \leq \Delta u(t+k|t), \quad k = 1, \ldots, N_p \quad \text{(8e)}
$$

$$
\Delta u(t+k|t) \leq \Delta u_{\max} + V_{\Delta u} \epsilon, \quad k = 1, \ldots, N_p \quad \text{(8f)}
$$

$$
g(t + N_u + j|t) = g(t + N_u|t), \quad j = 1, \ldots, N_p - N_u
$$
$$\tag{8g}$$

where $\Delta u(t+k|t) = u(t+k|t) - u(t+k-1|t)$, N_p and N_u are the prediction and control horizon, respectively, Q_y, Q_u, $Q_{\Delta u}$, Q_ϵ are nonnegative weights, u_{ref} and r are the input and output references, respectively, V_y, V_u, $V_{\Delta u}$ are positive vectors that are used to soften the constraints on plant's input and output, so that (8) always admits a solution. According to standard MPC design, in case $N_u < N_p$, the constraint (8g) enforces a constant value of g from time N_u to N_p. The reader is referred to [1] for an overview on MPC design.

We can also treat the prediction horizon N_p as a design parameter, and denote by $\nu = [\mu' \ N_p]'$, $\nu \in \mathbb{R}^{n_\mu} \times \mathbb{N}$ the overall vector of tuning parameters. The control horizon N_u determines, together with the number of constraints in (8), the computational complexity of the outer MPC controller. Therefore, it is usually fixed by the available online throughput. Alternatively, we can set $N_u = N_p$.

The remaining MPC parameters (N_u, Q_y, Q_u, $Q_{\Delta u}$, Q_ϵ, V_y, V_u and $V_{\Delta u}$) are treated as a specification of the desired closed-loop performance, and therefore not optimized. More generally, we could decouple the MPC quadratic cost in (8) from the closed-loop performance index $\tilde{J}(y_{1:T}, u_{1:T})$. For instance, $\tilde{J}(y_{1:T}, u_{1:T})$ can be a general, possibly non-convex function reflecting engineering or economic goals, while the cost of the MPC (8) is quadratic to facilitate online optimization. Indeed, in case the augmented model M is LTI as in (7), problem (8) reduces to a *quadratic programming* (QP) problem whose solution can be computed both *offline* using multiparametric quadratic programming [14] or online using dedicated QP solvers based, *e.g.*, on interior-point algorithms [15], fast gradient projection [16], or active set methods [17].

IV. PERFORMANCE-DRIVEN PARAMETER TUNING

Based on the controller parametrization introduced in the previous section, the closed-loop performance cost $\tilde{J}(y_{1:T}, u_{1:T})$ is a function of vectors θ and ν parametrizing the inner controller $\mathcal{K}$ and the outer MPC, respectively. Thus, under the hierarchical architecture of Fig. 1, the original control design problem (4) is equivalent to

$$
\min_{\theta, \nu} \quad \tilde{J}(y_{1:T}, u_{1:T}; \theta, \nu). \quad \text{(9)}
$$

A. Bayesian optimization for parameter selection

Algorithm 1 Bayesian optimization for controller design
---
1. **initialize** parameters *vs* performance set

$$\mathcal{D} \leftarrow \{(\theta_{1:N_{\text{in}}}, \nu_{1:N_{\text{in}}}), \tilde{J}_{1:N_{\text{in}}}\};$$

2. **for** $i = N_{\text{in}}, \ldots, i_{\max} - 1$ **do**
2.1. based on the data $\mathcal{D}$, **train** a GP approximating $\tilde{J}$;
2.2. **define** acquisition function $\alpha(\theta, \nu | \mathcal{D})$;
2.3. **compute** next controller parameters θ_i, ν_i as

$$\theta_{i+1}, \nu_{i+1} \leftarrow \arg\max_{\theta, \nu} \alpha(\theta, \nu | \mathcal{D});$$

2.4. **perform** closed-loop experiment and **measure** performance index $\tilde{J}_{i+1}$;
2.5. **augment** the training set $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\theta_{i+1}, \nu_{i+1}), \tilde{J}_{i+1}\}$;
2.6. **exit for loop** if a termination criterion is met;
3. **end for**
4. **compute** optimal parameters $\theta_{i^\star}, \nu_{i^\star}$, with $i^\star = \arg\min_i \tilde{J}_i$;

Output: Optimal controller parameters $\theta_{i^\star}$ and $\nu_{i^\star}$.

The design problem (9) is solved through the *Bayesian optimization* (BO) strategy [18] outlined in Algorithm 1. The algorithm is initialized (step 1) by performing $N_{\text{in}} \geq 1$ closed-loop experiments for N_{in} different (e.g., randomly chosen) values of controller parameters θ_i and ν_i (with $i = 1, \ldots, N_{\text{in}}$). For each pair (θ_i, ν_i), a closed-loop experiment is performed and the performance index $\tilde{J}_i$ is measured. In this way, an initial set $\mathcal{D} = \{(\theta_{1:N_{\text{in}}}, \nu_{1:N_{\text{in}}}), \tilde{J}_{1:N_{\text{in}}}\}$ of parameters and corresponding performance $\tilde{J}$ is constructed, with $\theta_{1:N_{\text{in}}}$ denoting the sequence θ_i for $i = 1, \ldots, N_{\text{in}}$. In practice, the experiment can be interrupted and large cost assigned to $\tilde{J}_i$ in case of safety constraint violations.

The algorithm is then iterated until a stopping criterion is met (e.g., maximum number of iterations reached). At each iteration $i \geq N_{\text{in}}$, the following two steps are performed:

- *Learning phase* (Step 2.1). In this step, a Gaussian Process (GP) describing our "best guess" of the cost $\tilde{J}(y_{1:T}, u_{1:T}; \theta, \nu)$ corresponding to the design parameters θ and ν is fitted on the available data $\mathcal{D}$.
  Under the prior assumption that the cost $\tilde{J}$ is generated by a GP with zero mean and covariance function $\kappa((\theta, \nu), (\tilde{\theta}, \tilde{\nu}))$, the posterior distribution of $\tilde{J}(y_{1:T}, u_{1:T}; \theta^\star, \nu^\star)$ for generic controller parameters $(\theta^\star, \nu^\star)$ can be computed analytically. Specifically, $\tilde{J}(y_{1:T}, u_{1:T}; \theta^\star, \nu^\star)$ is a Gaussian variable with mean

$$
m_i(\theta^\star, \nu^\star) = \mathbf{k}_i' \left(\mathbf{K}_i + \sigma_e^2 I\right)^{-1} \tilde{J}_{1:i}, \quad \text{(10a)}
$$

and variance

$$\sigma_i^2(\theta^\star, \nu^\star) = \kappa((\theta^\star, \nu^\star), (\theta^\star, \nu^\star)) + \tag{10b}$$

$$-\mathbf{k}_i' \left(\mathbf{K}_i + \sigma_e^2 I\right)^{-1} \mathbf{k}_i + \sigma_e^2, \tag{10c}$$

where the j-th element of the vector $\mathbf{k}_i \in \mathbb{R}^i$ is $\kappa((\theta^\star,\nu^\star), (\theta_j,\nu_j))$; the $[j,h]$-th entry of the Kernel matrix $\mathbf{K}_i \in \mathbb{R}^{i \times i}$ is $\kappa((\theta_j,\nu_j),(\theta_h,\nu_h))$; σ_e^2 represents the variance of an additive (Gaussian) noise possibly affecting the observations of the cost $\tilde{J}$; and I denotes the identity matrix of proper dimension.

The covariance function $\kappa((\theta,\nu),(\tilde{\theta},\tilde{\nu}))$ for the GP can be defined, for instance, in terms of the so-called *Squared Exponential* (SE) covariance kernel, defined as

$$\kappa((\theta,\nu),(\tilde{\theta},\tilde{\nu})) = \sigma_0^2 e^{-\frac{1}{2\lambda^2}\left[\theta'-\tilde{\theta}' \ \mu'-\tilde{\mu}'\right]\left[\theta'-\tilde{\theta}' \ \mu'-\tilde{\mu}'\right]'}.$$

The hyper-parameters σ_0 and λ characterizing the SE kernel, as well as the noise variance σ_e^2, can be chosen by maximizing the log marginal likelihood [19]

$$\log p(\tilde{J}_{1:i}|\theta_{1:i},\nu_{1:i},\sigma_0,\lambda,\sigma_e) \propto$$
$$\propto -\frac{1}{2}\log\det\left(\mathbf{K}_i + \sigma_e^2 I\right) - \frac{1}{2}\tilde{J}_{1:i}'\left(\mathbf{K}_i + \sigma_e^2 I\right)^{-1}\tilde{J}_{1:i}.$$

- *Optimization phase* (Steps 2.2-2.4). In this phase, the next design parameters θ_{i+1} and ν_{i+1} to test are chosen by maximizing the so-called *acquisition function* $\alpha(\theta,\nu|\mathcal{D})$ (Step 2.2). The acquisition function $\alpha(\theta,\nu|\mathcal{D})$ is constructed based on the mean and co-variance (eq. (10)) of the GP estimated in the learning step. The acquisition function balances exploration (*i.e.*, learning more about the objective $\tilde{J}$ in regions of the parameter space with high variance) and exploitation (*i.e.*, search over regions with high mean to optimize the expected performance based on past collected data). Different acquisition functions have been proposed in the literature (see [20] and the references therein for a deep overview). In the example reported in Section V, we use the *Expected Improvement* (EI) acquisition function, defined as

$$\alpha(\theta,\nu|\mathcal{D}) = \mathbf{EI}(\theta,\nu) = \mathbb{E}\left[\max\{0, \tilde{J}^- - \tilde{J}(\theta,\nu)\}\right], \tag{11}$$

where $\tilde{J}^-$ represents the best value of objective function at the i-th iteration, *i.e.*, $\tilde{J}^- = \min_{j=1,\dots,i} \tilde{J}(y_{1:T}, u_{1:T}; \theta_j, \nu_j)$. Under the GP framework previously discussed, the EI in (11) can be evaluated analytically and it is equal to:

$$\mathbf{EI}(\theta,\nu) = \left(\tilde{J}^- - m_i(\theta,\nu)\right)\Psi(Z) + \sigma_i(\theta,\nu)\psi(Z) \tag{12}$$

if $\sigma_i(\theta,\nu) > 0$, 0 otherwise. In (12), Z is defined as $Z = \frac{\tilde{J}^- - m_i(\theta,\nu)}{\sigma_i(\theta,\nu)}$, and ψ and Ψ are the *probability density function* and the *cumulative density function* of the standard normal distribution, respectively.

The advantages of using BO for tackling this design problem are twofold. First, it is a derivative-free optimization algorithm, which is useful since a closed-form expression of

the performance $\tilde{J}$ as a function of the design parameters θ and ν is not available. Second, it allows us to tune the controller parameters with as few evaluations of $\tilde{J}$ as possible. The latter point is crucial, since each evaluation can be costly and time-consuming, as it requires a closed-loop experiment.

B. Restricting the parameter space

Bayesian optimization allows setting bounds on the search space of the parameters θ and ν. These bounds can be included in the maximization of the acquisition function at Step 2.3 of Algorithm 1. Restricting the search space generally speeds up the algorithm's convergence, thus requiring fewer evaluations of the functional $\tilde{J}$. Suitable bounds may be defined exploiting prior system knowledge and design choices. Some applicable restrictions of the parameter space are discussed next.

It may be reasonable to assume that the optimal solution is achieved using an inner controller $\mathcal{K}$ that stabilizes the inner loop $\mathcal{M}$. Therefore, one may constrain μ so that the prediction model M used by MPC is asymptotically stable.

Some basic control design rules may be also used to restrict the search space of θ defining the inner-loop controller $\mathcal{K}$. For instance, if $\mathcal{K}$ is a PID controller parametrized as in (6), its static gain should have the same sign of the static gain of the (stable) system $\mathcal{S}$.

Another significant reduction of the parameter space may be achieved under the assumption that the prediction sub-model $M_y(\mu_y)$ used by the MPC accurately describes the system dynamics $\mathcal{M}$ from g to y. In this case, one can simply derive the augmented model M providing the relationship from g to the plant input u and output y as

$$\begin{bmatrix} u \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} K(\theta)(I - M_y(\mu_y)) \\ M_y(\mu_y) \end{bmatrix}}_{M(\mu_y, \theta)} g. \tag{13}$$

Note that in this case $\mu = \begin{bmatrix} \mu_y \\ \theta \end{bmatrix}$, that is the prediction model and controller share some parameters.

Other restrictions may be introduced according to the particular problem at hand and prior knowledge available to the user, *e.g.*, diagonal models assuming decoupled dynamics, grey-box models with known intervals for physical parameters, etc.

V. NUMERICAL EXAMPLE

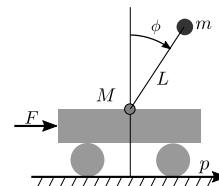As a case study, we consider the control problem of the inverted pendulum on a cart depicted in Fig. 2.

Fig. 2. Schematics of the inverted pendulum on a cart.

A. System description

The dynamics of the process are governed by the equations

$$(M + m)\ddot{p} + mL\ddot{\phi}\cos\phi - mL\dot{\phi}^2\sin\phi + b\dot{p} = F, \quad (14a)$$

$$L\ddot{\phi} + \ddot{p}\cos\phi - g\sin\phi + f_\phi\dot{\phi} = 0, \quad (14b)$$

where p is the cart position, ϕ is the angle of the pendulum with respect to the upright vertical position, and F is an input force acting on the cart. The following values of the physical parameters are used: $M = 0.5$ Kg (cart mass), $m = 0.2$ Kg (pendulum mass), $L = 0.3$ m (rod length), $g = 9.81$ m/s^2 (gravitational acceleration), $b = 0.1$ N/m/s, and $f_\phi = 0.1$ m/s (friction terms). According to the approach proposed in the paper, no knowledge of the physical model of the process is used in designing the controller, and (14) are only used for data generation and performance evaluation.

The output signals p and ϕ are measured every $T_s = 5$ ms and measurements are corrupted by an additive zero-mean white Gaussian noise with standard deviation 0.01 m and 0.01 rad, respectively. The input force F is also perturbed by an additive zero-mean random disturbance with standard deviation 1 N and bandwidth 10 rad/s.

In performing closed-loop experiments, the system is initialized at $[p(0)\ \dot{p}(0)\ \phi(0)\ \dot{\phi}(0)] = [0\ 0\ \frac{\pi}{20}\ 0]$. The objective is to move the pendulum to the vertical position $\phi = 0$, while limiting the cart displacement. The force F is constrained to belong to the interval $I_F = [-20\ 20]$ N, while the cart position p should stay within the range $I_p = [-1\ 1]$ m (representing, *e.g.*, finite length of the track where the cart moves).

B. Control design

The hierarchical controller in Fig. 1 is designed, with $y = [p\ \phi]$ and $u = F$. The inner-loop controller $\mathcal{K}$ is

$$u = [0 \quad \mathcal{K}_{PI}(z, \theta)]\,(g - y), \quad (15)$$

where $\mathcal{K}_{PI}(z, \theta)$ is a discrete-time transfer function of a PID controller parametrized as in (6), with $\theta = [\theta_P\ \theta_I\ \theta_D] \in \mathbb{R}^3$ and $N_d = 100$. Note that only the angle ϕ is actually fed back in the inner loop, thus the task of the inner controller $\mathcal{K}$ is only to stabilize the dynamics of the angle ϕ.

Besides taking care of the control objectives, the outer MPC shall also enforce constraints on the cart position p and on the input force F. The model used by the MPC to predict the dynamics of the inner loop $\mathcal{M}$ from the MPC command g to the plant output $y = [p\ \phi]$ (see Fig. 1) is parameterized as the continuous-time state-space model

$$\dot{\xi}_M = A_M(\mu)\xi_M + B_M(\mu)g, \quad y_M = \xi_M, \quad (16)$$

where $\xi_M \in \mathbb{R}^2$ is the state vector and $\mu \in \mathbb{R}^6$ contains the entries of $A_M \in \mathbb{R}^{2\times2}$ and the second column of $B_M \in \mathbb{R}^{2\times2}$. Because of the structure of the inner controller $\mathcal{K}$ in (15), the position p is not fed back to the inner loop. Thus, the first column of B_M is set to zero and not included in the design parameter vector μ. The overall MPC prediction model M is constructed using (13).

The MPC control law is computed solving (8) and applied in a receding-horizon fashion, using a sampled version of (16) with sampling time $T_{\mathrm{MPC}} = 10T_s = 50$ ms, reference $r = [r_p\ r_\phi] = [0\ 0]$ and real-time constraints on F and p based on the admissible intervals I_F and I_p, respectively.

Regarding the MPC design parameters, the weight matrices are not optimized and set to $Q_y = \mathrm{diag}(0.1, 0.1)$, $Q_u = 0$, $Q_{\Delta u} = 0.1$ and $Q_\epsilon = 10^5$. The prediction horizon N_p is considered as a free parameter to be adjusted in the Bayesian optimization, while N_u is set equal to N_p. The real-value design parameters θ and μ are constrained to belong to the interval $[-500 \quad 500]$, while the prediction horizon N_p can take integer values between 10 and 20. The MPC control law is computed using the MATLAB *Model Predictive Control Toolbox*. All the computations are carried out on an i5 2.60-GHz Intel core processor with 32 GB of RAM running MATLAB R2018a. The maximum computational time required to evaluate the MPC law over all the performed closed-loop experiments is 21 ms, thus lower than the sampling time $T_{\mathrm{MPC}} = 50$ ms.

Overall, there are 10 parameters to be designed, namely $\theta \in \mathbb{R}^3$, $\mu \in \mathbb{R}^6$, and $N_p \in \mathbb{N}$. The closed-loop performance cost $\tilde{J}$ to be minimized is defined as

$$\tilde{J} = \log\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{10}|r_p - p(t)| + \frac{9}{10}|r_\phi - \phi(t)|\right)\right] + $$
$$+ \log\left[\frac{1}{T}\sum_{t=1}^{T}b(p(t)) + 1\right], \quad (17)$$

where

$$b(p) = \begin{cases} 10(|p| - 1) & \text{for } |p| > 1, \\ 0 & \text{for } |p| \le 1. \end{cases} \quad (18)$$

is the barrier function taking into account violation on the physical constraints on the cart position p. The cost $\tilde{J}$ is evaluated over closed-loop experiments of length 10 s on the discrete-time samples collected at rate T_s. This objective function reflects the engineering objective of controlling the angle ϕ to 0, limiting the horizontal displacement and keeping the cart position in the admissible range I_p. The constraint on the force F is enforced by a saturation block at the system input, and thus it is not penalized in $\tilde{J}$.

The design problem (9) is solved using the MATLAB *Statistics and Machine Learning Toolbox*, setting the **EI** in (11) as acquisition function. $N_{\mathrm{in}} = 10$ random values of the design parameters θ, μ and N_p are generated to initialize Algorithm 1, which is then executed for 310 iterations. The complete test code of this paper is available for download at `http://www.marcoforgione.it/data/code/CSL2019_perf.zip`.

C. Simulation results

The performance cost $\tilde{J}$ *vs* the iteration index i of Algorithm 1 is shown in Fig. 3. For each iteration i, the performance of the current test point (black asterisk) and of the current best point up to iteration i (red line) are shown. From Fig. 3, it can be noticed that the optimal controller parameters are found at iteration 123 (green square). Furthermore, as

the iteration index i increases, more and more test points are concentrated in an area of low cost $\tilde{J}$.

A closed-loop experiment is repeated over a longer period of 20 s using the designed controller. The time trajectories of the cart's position p and the pendulum's angle ϕ are plotted in Fig. 4, which shows that the designed controller is able to stabilize the pendulum' angle in the upright vertical position, respecting the constraints on the cart's position p.

For the sake of comparison, the following two non-hierarchical model-based controllers are designed based on the physical model of the system (Eq. (14)) linearized around $[p(0)\ \dot{p}(0)\ \phi(0)\ \dot{\phi}(0)] = [0\ 0\ 0\ 0]$:

- an MPC, with the same sampling rate $T_{\text{MPC}} = 50$ ms considered before, which reflects real-time constraints. At this sampling rate, the MPC is not able to reject the disturbance and thus fails to stabilize the pendulum around the upright vertical position. This shows the advantages of the hierarchical multi-rate controller structure.
- a *Linear-Quadratic-Gaussian* (LQG) controller, with sampling rate $T_s = 5$ ms. This controller stabilizes the pendulum. However, besides requiring a knowledge of the plant, it achieves a performance cost $\tilde{J}$ (eq. (17)) equal to -2.41, which is worse than the cost $\tilde{J} = -3.66$ obtained using the proposed performance-oriented approach.

VI. CONCLUSIONS AND FOLLOW-UP

In this work, we described a method to learn *MPC-oriented models* for hierarchical control schemes via iterative closed-loop experiments. We showed that such experiments can be suitably designed using Bayesian optimization. In the proposed learning framework, the model does not necessarily

provide the highest input/output data fit, which is the typical objective of system identification, but is the one yielding the model-based controller corresponding to the best closed-loop performance. We also argued that the prediction horizon can be optimized using the same tools and experiments. Numerical simulations on a benchmark example showed that data can lead to satisfactory controllers with no knowledge of the system dynamics and no constraints on modeling accuracy. Future research will be devoted to the theoretical analysis of the proposed learning strategy as well as to its experimental validation on a real-world setup.

REFERENCES

[1] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[2] A. Mesbah, "Stochastic model predictive control: An overview and perspectives for future research," *IEEE Control Systems Magazine*, vol. 36, no. 6, pp. 30–44, 2016.

[3] P. Falugi and D. Q. Mayne, "Getting robustness against unstructured uncertainty: a tube-based mpc approach," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1290–1295, 2014.

[4] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in identification and control*. Springer, 1999, pp. 207–226.

[5] M. Gevers, "Identification for control: From the early achievements to the revival of experiment design," *European journal of control*, vol. 11, no. 4-5, pp. 335–352, 2005.

[6] U. Rosolia and F. Borrelli, "Learning model predictive control for iterative tasks. a data-driven control framework," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1883–1896, 2018.

[7] M. Zanon, S. Gros, and A. Bemporad, "Practical reinforcement learning of stabilizing economic MPC," in *European Control Conference*, 2019, to appear.

[8] R. Kadali, B. Huang, and A. Rossiter, "A data driven subspace approach to predictive controller design," *Control engineering practice*, vol. 11, no. 3, pp. 261–278, 2003.

[9] D. Piga, S. Formentin, and A. Bemporad, "Direct data-driven control of constrained systems," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 4, pp. 331–351, 2018.

[10] J. Schäfer and A. Cinar, "Multivariable mpc system performance assessment, monitoring, and diagnosis," *Journal of process control*, vol. 14, no. 2, pp. 113–129, 2004.

[11] E. Garone, S. Di Cairano, and I. Kolmanovsky, "Reference and command governors for systems with constraints: A survey on theory and applications," *Automatica*, vol. 75, pp. 306–328, 2017.

[12] A. Bemporad, A. Casavola, and E. Mosca, "Nonlinear control of constrained linear systems via predictive reference management," *IEEE Trans. on Automatic Control*, vol. 42, no. 3, pp. 340–349, 1997.

[13] S. Bansal, R. Calandra, T. Xiao, S. Levine, and C. J. Tomiin, "Goal-driven dynamics learning via bayesian optimization," in *IEEE 56th Annual Conference on Decision and Control*, 2017, pp. 5168–5173.

[14] A. Bemporad, M. Morari, V. Dua, and E. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, p. 3.20, 2002.

[15] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 267–278, March 2010.

[16] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 18–33, Jan 2014.

[17] H. J. Ferreau, H. G. Bock, and M. Diehl, "An online active set strategy to overcome the limitations of explicit MPC," *International Journal of Robust and Nonlinear Control*, vol. 18, no. 8, pp. 816–830, 2008.

[18] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.

[19] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT Press Cambridge, MA, 2006, vol. 2, no. 3.

[20] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.
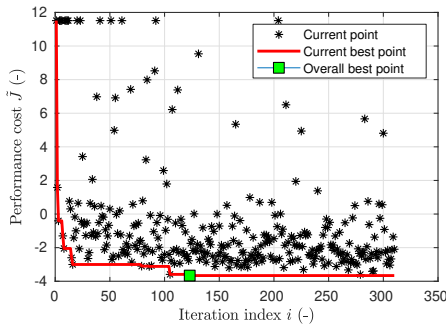
Fig. 3. Performance cost $\tilde{J}$ *vs* iteration i of Algorithm 1.

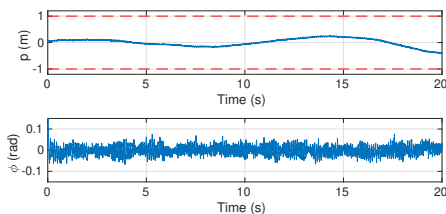Fig. 4. Closed-loop experiment: position p of the cart with the considered bounds (top plot) and pendulum's angle ϕ (bottom plot).