Data-based Robust MPC with Componentwise Hölder Kinky Inference

J.M. Manzano[1], D. Limon[1], D. Muñoz de la Peña[1], J.P. Calliess[2]

***Abstract*— The authors have recently developed predictive controllers based on prediction models derived from experimental data, by means of a class of Hölder interpolation called *kinky inference*. This paper provides a step forward by proposing a novel estimation method based on componentwise Hölder interpolation. This allows to explicitly consider the contribution of each component on each output, yielding better estimations. Following the procedure used in previous works, this estimation method is used to provide a predictor for a nonlinear robust data-based predictive controller, whose performance and robustness is enhanced by the new setting. The properties of the proposed controller are demonstrated in a case study.**

I. INTRODUCTION

Model predictive control (MPC) [1] is a well-known high level control technique. Its popularity arises from its ability to explicitly optimize the performance, while being able to ensure robust constraint satisfaction and closed-loop stability by design. As its name indicates, a model of the system to be controlled is required in order to predict future behaviour of the plant. Often, accurate models are not available, or they are too difficult to obtain or to manage (consider, for example, a large and highly nonlinear dynamical plant).

For this reason, predictive controllers which use models that are based on data have gained an increasing attention in the last decade. Within this setting, a machine learning technique is used to learn a dynamical function to predict the expected value of some unknown query point. This technique is then applied to model a dynamical system [2] in the context of control [3].

With respect to the different machine learning techniques used, some research consider *historian data bases* [4], others Gaussian processes [5], [6], neural networks [7], among many others.

In this paper, we consider a class of learning rules named *kinky inference* (KI) [8], [9]. This class of predictors encompasses Lipschitz interpolation [10], [11] and nonlinear set membership methods [12]. In [13] a data-based robust MPC is presented using the *smooth projected kinky inference* version of the kinky inference class for unconstrained systems. Later, in [14] a stabilizing robust version for systems subject to input and output constraints was developed.

The objective of this paper is to use an extended version of continuity of functions, namely *componentwise Hölder continuity* in order to derive a novel prediction method, named *componentwise Hölder kinky inference* (CHOKI), to be used within the framework proposed in [13]. This approach consists in extending the Lipschitz constant L to a matrix, and the same for the Hölder exponent p. With this setting, a new scope on Lipschitz interpolation arises, considering different gradients of the function to be learnt along its input dimensions, as well as different coefficients for each output. This approach will be applied to control output feedback nonlinear systems subject to hard constraints using the stabilizing formulation proposed in [14].

The advantage of using this matrix approach is two-fold: first, it allows the method to decrease the prediction error when estimating a function. Second, in the context of control, to counteract the effect of the mismatches between the predicted and the real value (due to model errors or disturbances) while satisfying hard constraints in the outputs, we will make use of the estimation of the reachable sets to be considered as back-off of tightened constraints, whose limits are improved by this matrix consideration.

Notation

The set of integers in the interval $[a, b]$ is denoted as $\mathbb{I}_a^b$. Given two column vectors, v and w, the notation (v, w) implies $[v^T, w^T]^T$. Given two sets A and B, $A \oplus B$ denotes the Minkowski sum, and $A \ominus B$ the Pontryagin difference. A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a $\mathcal{K}$-function if $\alpha(0) = 0$ and it is strictly increasing. Given a vector $v \in \mathbb{R}^{n_v}$ the ball $\mathcal{B}(v) \subset \mathbb{R}^{n_v}$ is defined as $\mathcal{B}(v) = \{y : |y_i| \leq v_i,\ i \in \mathbb{I}_1^{n_v}\}$. If M is a matrix, $\min M$ is a column vector containing the minimum element from each row.

II. COMPONENTWISE HÖLDER CONTINUITY

Consider two compact metric spaces $(\mathcal{W} \in \mathbb{R}^{n_w}, \mathfrak{d}_{\mathcal{W}})$ and $(\mathcal{Y} \in \mathbb{R}^{n_y}, \mathfrak{d}_{\mathcal{Y}})$ referred to as *input* and *output* space, respectively. A function $f : \mathcal{W} \to \mathcal{Y}$ is said to be Hölder continuous if

$$\mathfrak{d}_{\mathcal{Y}}(y_1, y_2) \leq L \mathfrak{d}_{\mathcal{W}}(w_1, w_2)^p,\ \forall w_1, w_2.$$

Here, the exponent p is a scalar such that $0 < p \leq 1$. In particular, if $p = 1$ this property is called Lipschitz continuity, with constant L, which is also a scalar.

Based on this property, several methods which learn the ground truth function f and infer its value for an unseen query point have been proposed. They use a data set of (possibly noisy) samples, as well as the constants L and p, e.g. [10], [11] and [12]. Similarly, a class of learning rules has been proposed under the name of *kinky inference* [8], [13].

This work was supported by the MINECO-Spain and FEDER Funds under project DPI2016-76493-C3-1-R and the University of Seville under contract VI-PPITUS.

[1]Department of Systems Engineering and Automation, Universidad de Sevilla, Spain. {manzano,dlm,dmunoz}@us.es

[2] Department of Engineering Science, University of Oxford, United Kingdom.jan-peter.calliess@oxford-man.ox.ac.uk

In this paper, a new approach on the Hölder property is considered. We exploit the fact that a function may have sharp gradients along one dimension of the input, while varying smoothly with respect to other input dimensions. Besides, if the function maps the inputs to several outputs, it seems intuitive to distinguish different parameters of the Hölder continuity for each output, as if they were different functions. Hence, we propose to use matrices as the parameters of the Hölder continuity: $\mathcal{L}$, $\mathcal{P} \in \mathbb{R}^{n_y \times n_w}$.

To this end, new notation is introduced. Given a vector $w \in \mathbb{R}^{n_w}$ and two matrices $\mathcal{L}$, $\mathcal{P} \in \mathbb{R}^{n_y \times n_w}$, we define

$$\mathfrak{d}(w) \doteq (|w_1|, \dots, |w_{n_w}|), \tag{1}$$

$$\mathfrak{d}_{\mathcal{L}}^{\mathcal{P}}(w) \doteq \left\{ a : a_i = \sum_{j=1}^{n_w} \mathcal{L}_{i,j} |w_j|^{\mathcal{P}_{i,j}} \right\}, \tag{2}$$

where $a \in \mathbb{R}^{n_y}$, and the notation $\mathcal{L}_{i,j}$ refers to the *i,j-th* component of the matrix $\mathcal{L}$.[1,2]

Definition 1 (Componentwise Hölder continuity): Given the matrices $\mathcal{L}$ and $\mathcal{P}$, a function $f : \mathcal{W} \to \mathcal{Y}$ is componentwise Hölder continuous if

$$\mathfrak{d}(y_1 - y_2) \leq \mathfrak{d}_{\mathcal{L}}^{\mathcal{P}}(w_1 - w_2), \ \forall w_1, w_2. \tag{3}$$

The following theorem states under which conditions Hölder continuity and componentwise Hölder continuity are equivalent.

Theorem 1:
1) Hölder continuity implies componentwise Hölder continuity.
2) Within a compact set, componentwise Hölder continuity implies Hölder continuity.

Proof of statement 1): Given two metrics $\mathfrak{d}_{\mathcal{Y}}$ and $\mathfrak{d}_{\mathcal{W}}$ and two parameters L and p such that $f : \mathcal{W} \to \mathcal{Y}$ is Hölder continuous, then

$$\begin{aligned} \mathfrak{d}_{\mathcal{Y}}(y_1, y_2) &\leq L \mathfrak{d}_{\mathcal{W}}(w_1, w_2)^p && \text{(4a)} \\ &\leq L_\infty \|w_1 - w_2\|_\infty^p && \text{(4b)} \\ &= L_\infty |w_{1,j} - w_{2,j}|^p && \text{(4c)} \\ &\leq L_\infty \sum_{j=1}^{n_w} |w_{1,j} - w_{2,j}|^p && \text{(4d)} \end{aligned}$$

where inequality (4b) arises from the equivalence of the norms, and inequality (4d) provided that $|w_{1,j} - w_{2,j}|^p \geq 0$, $\forall j$.

Then, defining $\mathcal{L} = L_\infty \mathbb{1}_{n_y \times n_w}$ and $\mathcal{P} = p \mathbb{1}_{n_y \times n_w}$ (where $\mathbb{1}$ denotes a matrix of ones), the function f is componentwise Hölder continuous. ∎

Proof of 2): Given that $|w_{1,j} - w_{2,j}| \leq \|w_1 - w_2\|_\infty$,

$$\begin{aligned} \mathfrak{d}(y_{i,1} - y_{i,2}) &\leq \sum_{j=1}^{n_w} \mathcal{L}_{i,j} |w_{1,j} - w_{2,j}|^{\mathcal{P}_{i,j}} \\ &\leq \sum_{j=1}^{n_w} \mathcal{L}_{i,j} \|w_1 - w_2\|_\infty^{\mathcal{P}_{i,j}}. \end{aligned}$$

[1]Note that, with a slight abuse of notation, $\mathfrak{d}(w)$ in equation (1) stands for a norm, rather than a metric.

[2]While w_j in equation (2) stands for the j-th component of w, in equation (3) the subindex w_1 indicates a data point.

If $\mathcal{W}$ is a compact space, there exists a c such that $\|w\|_\infty \leq c$, $\forall w \in \mathcal{W}$, so $\|w_1 - w_2\| \leq 2c$, $\forall w_1, w_2 \in \mathcal{W}$. Besides, if $x \in [0,\, 1]$ and $p_1 \geq p_2$, with $p_1, p_2 \in (0, 1]$, then $x^{p_1} \leq x^{p_2}$.

Let $\mathcal{P} = \min_{i,j} p_{i,j}$ and let $\mathcal{L} = \max_i \sum_{j=1}^{n_y} \mathcal{L}_{i,j} (2c)^{\mathcal{P}_{i,j} - p}$. Then

$$\begin{aligned} \mathcal{L}_{i,j} \|w_1 - w_2\|_\infty^{\mathcal{P}_{i,j}} \leq \\ &\leq \mathcal{L}_{i,j} (2c)^{\mathcal{P}_{i,j}} \left(\frac{\|w_1 - w_2\|}{2c} \right)^{\mathcal{P}_{i,j}} \\ &\leq \mathcal{L}_{i,j} (2c)^{\mathcal{P}_{i,j}} \left(\frac{\|w_1 - w_2\|_\infty}{2c} \right)^{\mathcal{P}_{i,j}} \\ &= \mathcal{L}_{i,j} (2c)^{\mathcal{P}_{i,j} - p} \|w_1 - w_2\|_\infty^{\mathcal{P}}. \end{aligned}$$

Hence,

$$\begin{aligned} |f_i(w_1) - f_i(w_2)| &\leq \sum_{j=1}^{n_y} \mathcal{L}_{i,j} \|w_1 - w_2\|_\infty^{\mathcal{P}_{i,j}} \\ &\leq \sum_{j=1}^{n_y} \mathcal{L}_{i,j} (2c)^{\mathcal{P}_{i,j} - p} \|w_1 - w_2\|_\infty^{\mathcal{P}_{i,j}} \\ &\leq L \|w_1 - w_2\|_\infty^{\mathcal{P}}. \end{aligned}$$

And since there $\exists i^* : |f_{i^*}(w_1) - f_{i^*}(w_2)| = \|f(w_1) - f(w_2)\|_\infty$, hence

$$\|f(w_1) - f(w_2)\|_\infty \leq L \|w_1 - w_2\|^p. \qquad \blacksquare$$

The equivalence between these two definitions guarantees that the inference method proposed in the next section inherits the well-known estimation properties of kinky inference methods.

III. COMPONENTWISE HÖLDER KINKY INFERENCE

The class of learning rules known as *kinky inference* has been growing recently. It makes use of the input and output spaces $\mathcal{W} \in \mathbb{R}^{n_w}$, $\mathcal{Y} \in \mathbb{R}^{n_y}$, the Hölder continuous ground truth function $f : \mathcal{W} \to \mathcal{Y}$, the data set of $N_{\mathcal{D}}$ data points $\mathcal{D} = \{(w_i, y_i)\}$, $i \in \mathbb{I}_1^{N_{\mathcal{D}}}$, and the scalar parameters L and p. The data set containing only inputs is denoted as $\mathcal{W}_{\mathcal{D}} = \text{Proj}_{\mathcal{W}}(\mathcal{D})$. The standard KI method to learn f and to predict a new query point $q \notin \mathcal{W}_{\mathcal{D}}$ is presented in [8]

In [14], L and p are extended to be n_y-dimensional, considering different parameters separately for each of the outputs. With respect to the calculation of these parameters, [9] introduces the *lazily adapted constant kinky inference* (LACKI), where L is the smallest scalar that validates the data. In [15], obtaining L and p is done through the *parameter optimized kinky inference* (POKI) method, by minimising a validation error.

In this paper, making use of the componentwise Hölder continuity defined in (3), we propose a new *componentwise Hölder kinky inference* (CHOKI) predictor, such that given

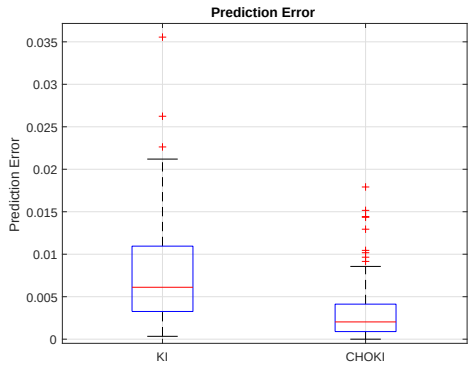

Fig. 1. KI and CHOKI prediction error for $y(w) = w_1^2 + w_1 w_2 - w_2$.

$\Theta = (\mathcal{L}, \mathcal{P})$,

$$\begin{aligned} \tilde{\mathfrak{f}}(q;\Theta,\mathcal{D}) &= \frac{1}{2}\min_{i=1,\dots,N_\mathcal{D}} (\tilde{f}_i + \mathfrak{d}_\mathcal{L}^\mathcal{P}(q-w_i)) \\ &+ \frac{1}{2}\max_{i=1,\dots,N_\mathcal{D}} (\tilde{f}_i - \mathfrak{d}_\mathcal{L}^\mathcal{P}(q-w_i)). \end{aligned} \quad (5)$$

where $\tilde{f}_i$ is the observed map for the i-th data point in $\mathcal{D}$, which may be noisy, and w_i is its corresponding input.

As it was stated in Theorem 1, the approach considered in this paper is a generalization of the Hölder continuity. Hence, the CHOKI predictor will perform better than the standard KI, since the latter is a particular case in which every element of $\mathcal{L}$ and $\mathcal{P}$ equals to the standard KI parameters L and p.

As an example, consider the function

$$f : \{w \in \mathbb{R}^2 \to y \in \mathbb{R}^1 \,|\, y(w) = w_1^2 + w_1 w_2 - w_2\}.$$

If we had $N_\mathcal{D} = 2000$ data points and we were to predict 100 random query points, the resulting estimation error is represented in Figure 1, both for the KI and the CHOKI predictor. With CHOKI, this prediction error decreases up to a 70%.

IV. DATA-BASED PREDICTION MODEL

This section proposes the setting used for modelling, based on data, nonlinear dynamical systems.

Consider the discrete plant governed by the control inputs $u \in \mathbb{R}^{n_u}$ and outputs $y \in \mathbb{R}^{n_y}$ at the sampling time $k \in \mathbb{N}$. Then, the measured output can be modelled as a NARX regression of previous inputs and outputs [13]:

$$y(k+1) = f(x(k), u(k)) + e(k), \quad (6)$$

where

$$x(k) = (y(k), \dots, y(k-n_a), u(k-1), \dots, u(k-n_b)), \quad (7)$$

is the regression state $x \in \mathbb{R}^{n_x}$, $n_x = (n_a+1)n_y + n_b n_u$ for some memory horizons n_a and $n_b \in \mathbb{N}$. The measurements may be noise corrupted. The modelling error can be written as a noise $e(k) \in \mathcal{E}$, where $\mathcal{E}$ is a compact set.

In order to use KI methods, the arguments of f are aggregated into $w = (x, u) \in \mathbb{R}^{n_w}$, with $n_w = (n_a+1)n_y + (n_b+1)n_u$, yielding $f : \mathcal{W} \to \mathcal{Y}$. Then, given some historical trajectories of $u(k)$ and $y(k)$, and some memory horizons, it is possible to construct a data set $\mathcal{D} = \{(w_k, y_{k+1})\}$ for $k = 1, \dots, N_\mathcal{D}$ and hence predict a new output $y(k+1)$ given $\Theta = (\mathcal{L}, \mathcal{P})$ as in (5):

$$\hat{y}(k+1) = \tilde{\mathfrak{f}}(w(k); \Theta, \mathcal{D}). \quad (8)$$

Note that we make use of the following assumptions:

Assumption 1: The ground truth function f is Hölder continuous.

Assumption 2: The prediction error is bounded by some $\mu \in \mathbb{R}^{n_y}$. This bound is known for all admissible x and u, i.e.,

$$\mathfrak{d}(\tilde{\mathfrak{f}}(x, u; \Theta, \mathcal{D}) - f(x, u)) \leq \mu. \quad (9)$$

Notice that this bound exists thanks to Assumption 1 and the compactness of the admissible sets.

V. ROBUST DATA-BASED MPC

This section presents the model predictive controller and its stability and robustness analysis, extended for the CHOKI predictor. The controller requires a prediction model which is derived from experimental data by using the method proposed in sections III and IV. Such prediction model can be formulated in state space as in [13][3]:

$$x(k+1) = F(x(k), u(k)) + \xi(k) \quad (10a)$$
$$y(k) = Mx(k), \quad (10b)$$

Besides, the inputs and outputs of the plant must fulfil certain hard constraints at every sampling time, that is, $u \in \mathcal{U}$ and $y \in \mathcal{Y}$. We propose to use a robust model predictive controller which is stable by design, following the procedure in [14].

This controller is based on the standard design assumption that there exists a local controller with a known Lyapunov function that is able to stabilize the system in a region around the origin. However, instead of including a terminal region constraint (which would need to take into account the possible effects of the predictions errors), the terminal cost weight in the cost function is increased [16]. This approach is appropriate for data-based controllers in order to avoid the definition of the terminal region, which would in general yield conservative controllers.

Thus, the proposed predictive control problem, denoted as $P_N(x(k); \Theta, \mathcal{D})$, is

$$\min_{\mathbf{u}} \quad \sum_{i=0}^{N-1} \ell(\hat{x}(i|k), u(i)) + \lambda V_f(\hat{x}(N|k)) \quad (11a)$$
$$\text{s.t.} \quad \hat{x}(0|k) = x(k) \quad (11b)$$
$$\hat{x}(j+1|k) = \hat{F}(\hat{x}(j|k), u(j)),\ j \in \mathbb{I}_0^{N-1} \quad (11c)$$
$$\hat{y}(j|k) = M\hat{x}(j|k) \quad (11d)$$
$$u(j) \in \mathcal{U} \quad (11e)$$
$$\hat{y}(j|k) \in \mathcal{Y}_j. \quad (11f)$$

Here, N is the prediction horizon, $\ell(\cdot,\cdot)$ is the stage cost, λ is a design parameter greater than 1 and $V_f(\cdot)$ is the terminal

[3]The reader is referred to [13] for further details.

cost. The hard constraint in the output is managed by a set of tightened constraints, which is defined as

$$\mathcal{Y}_j = \mathcal{Y} \ominus \mathcal{B}(s_j). \tag{12}$$

The objective of this set of tightened constraints is to guarantee recursive feasibility by explicitly taking into account the reachability sets of the predictions for all possible predictions errors, see for example [17].

It is important to remark that, for nonlinear systems, there are few results that tackle the computation of the tightened constraints (also known as back-off), often considering only the effect in the following time step, and not an accumulated error. Next, we present a procedure based on the CHOKI predictor to calculate these sets, which in general provides less conservative results than the ones given by standard KI procedures [14]. For its calculation, we will make use of the following lemma:

Lemma 1: The mismatch between a prediction at time step j given the measurement at time step k and the prediction at that time step given the measurement at time $k+1$, for the same sequence of control inputs is bounded by

$$\begin{aligned} \mathfrak{d}(\hat{y}(j|k) - \hat{y}(j-1|k+1)) &\leq c_j \in \mathbb{R}^{n_y} && (13)\\ \mathfrak{d}(\hat{w}(j|k) - \hat{w}(j-1|k+1)) &\leq d_j \in \mathbb{R}^{n_w}. && (14) \end{aligned}$$

The parameters c and d can be obtained from the recursion

$$\begin{aligned} c_{j+1} &= \mathfrak{d}_{\mathcal{L}}^{\mathcal{P}}(d_j) && (15)\\ d_j &= (c_j, \ldots, c_{\sigma(j)}, 0, \ldots, 0), && (16) \end{aligned}$$

with $\sigma(j) = \max(1, j - n_a)$, and $c_1 = \mu$.

Proof: Provided that $w(k) = (x(k), u(k))$, where $x(k)$ is the recursion given by eq. (7), and that the sequence of future inputs $u(k+i)$ for $i \in \mathbb{I}_0^{N-1}$ is given,

$$\begin{aligned} &\mathfrak{d}(\hat{w}(j|k) - \hat{w}(j-1|k+1)) = \\ &= \begin{bmatrix} \mathfrak{d}(\hat{y}(j|k) - \hat{y}(j-1|k+1)) \\ \mathfrak{d}(\hat{y}(j-1|k) - \hat{y}(j-2|k+1)) \\ \vdots \\ \mathfrak{d}(\hat{y}(j-n_a|k) - \hat{y}(j-n_a-1|k+1)) \\ \mathfrak{d}(u(k+j-1) - u(k+j-1)) \\ \vdots \\ \mathfrak{d}(u(k+j-n_b) - u(k+j-n_b)) \\ \mathfrak{d}(u(k+j) - u(k+j)) \end{bmatrix}. \end{aligned} \tag{17}$$

These $\hat{y}$ are predicted values if $j > n_a$, and real measurements if not. Equation (17) directly translates into

$$d_j = (c_j, \ldots, c_{\sigma(j)}, 0, \ldots, 0),$$

with $\sigma(j) = \max(1, j - n_a)$, provided that if $\hat{y}(j|k)$ is a real measurement, $\mathfrak{d}(y(j|k) - y(j-1|k+1)) = \emptyset_{n_y \times 1}$. Hence, d_j has $\max(0, n_a - j + 1)n_y + (n_b + 1)n_u$ zeros.

To obtain c_j we make use of the componentwise Hölder continuity of the predictor $\tilde{\mathfrak{f}}$. Given Assumption 2, we have that $c_1 = \mu$. Note that

$$\hat{y}(j+1|k) = \tilde{\mathfrak{f}}(\hat{x}(j|k), u(k+j); \Theta, \mathcal{D}),$$

and that

$$\mathfrak{d}(\tilde{\mathfrak{f}}_i - \tilde{\mathfrak{f}}_j) \leq \mathfrak{d}_{\mathcal{L}}^{\mathcal{P}}(w_i - w_j).$$

Hence,

$$\begin{aligned} &\mathfrak{d}(\hat{y}(j+1|k) - \hat{y}(j|k+1)) = \\ &= \mathfrak{d}(\tilde{\mathfrak{f}}(\hat{x}(j|k), u(k+j)) - \tilde{\mathfrak{f}}(\hat{x}(j-1|k), u(k+j))) \\ &\leq \mathfrak{d}_{\mathcal{L}}^{\mathcal{P}} \left(\begin{bmatrix} \mathfrak{d}(\hat{y}(j|k) - \hat{y}(j-1|k+1)) \\ \mathfrak{d}(\hat{y}(j-1|k) - \hat{y}(j-2|k+1)) \\ \vdots \\ \mathfrak{d}(\hat{y}(j-n_a|k) - \hat{y}(j-n_a-1|k+1)) \\ \mathfrak{d}(u(k+j-1) - u(k+j-1)) \\ \vdots \\ \mathfrak{d}(u(k+j-n_b) - u(k+j-n_b)) \\ \mathfrak{d}(u(k+j) - u(k+j)) \end{bmatrix} \right), \end{aligned}$$

From here, it immediately follows that $c_{j+1} = \mathfrak{d}_{\mathcal{L}}^{\mathcal{P}}(d_j)$. ■

Then, the tightened constraint is obtained computing s_j as

$$s_j = \sum_{i=1}^{j} c_j. \tag{18}$$

Thus, we can state the following lemma, which under certain assumptions guarantees recursive feasibility (that will be needed in the stability analysis):

Lemma 2: For all $y \in \mathcal{Y}_j$ and for all $\Delta y \in \mathcal{B}(c_j)$, the sets $\mathcal{Y}_j$ are such that $y + \Delta y \in \mathcal{Y}_{j-1}$.

Proof: By definition, we have that

$$y + \Delta y \in \mathcal{Y}_j \oplus \mathcal{B}(c_j) = \mathcal{Y} \ominus \mathcal{B}(s_j) \oplus \mathcal{B}(c_j).$$

Since for $j \geq 1$, $s_j = s_{j-1} + c_j$, we can state that

$$\mathcal{B}(s_j) = \mathcal{B}(s_{j-1}) \oplus \mathcal{B}(c_j),$$

and then

$$\mathcal{Y}_j = \mathcal{Y} \ominus \mathcal{B}(s_j) = \mathcal{Y} \ominus \mathcal{B}(s_{j-1}) \ominus \mathcal{B}(c_j).$$

Thus,

$$\begin{aligned} y + \Delta y &\in \mathcal{Y}_j \oplus \mathcal{B}(c_j) \\ &= \mathcal{Y} \ominus \mathcal{B}(s_{j-1}) \ominus \mathcal{B}(c_j) \oplus \mathcal{B}(c_j) \\ &\subseteq \mathcal{Y} \ominus \mathcal{B}(s_{j-1}) = \mathcal{Y}_{j-1}. \end{aligned}$$

■

With respect to the stability analysis, the ingredients of the optimization problem given by (11) are stated in the following assumptions:

Assumption 3: The estimation error bound μ and the prediction horizon N are such that the set $\mathcal{Y}_N$ is non-empty.

Assumption 4: The stage cost function $\ell(x, u)$ is a componentwise Hölder continuous, positive definite function such that $\ell(x, u) \geq \alpha(\|x\|)$, where α is a $\mathcal{K}$-function. Its Hölder constant is $\mathcal{L}_\ell$.

Assumption 5: There exists a control law $u = \kappa_f(x)$, a function V_f and a level set $\Omega_\gamma = \{x : V_f(x) \leq \gamma\} \subseteq \mathbb{R}^{n_x}$

for a certain $\gamma > 0$ such that for all $x \in \Omega_\gamma$ the following conditions hold:

(a) V_f is a componentwise Hölder continuous, positive definite function, with Hölder constant $\mathcal{L}_f$, such that if α_f and β_f are $\mathcal{K}$-functions, then

$$\alpha_f(\|x\|) \leq V_f(x) \leq \beta_f(\|x\|)$$
$$V_f(\hat{F}(x, \kappa_f(x))) - V_f(x) \leq -\ell(x, \kappa_f(x)).$$

(b) $\kappa_f(x) \in \mathcal{U}$, $Mx \in \mathcal{Y}_N$.

Assumption 6: It is assumed that μ is such that the set

$$\Upsilon = \{x : \ell(x, 0) \leq \nu(\mu)\}$$

is contained in Ω_γ, where ν is defined as

$$\nu(\mu) = \sum_{j=1}^{N} \mathfrak{d}^{\mathcal{P}}_{\mathcal{L}_\ell}(d_j) + \lambda \mathfrak{d}^{\mathcal{P}}_{\mathcal{L}_f}(d_{N+1}).$$

It is also assumed that the positive constant ϕ is such that $\ell(x, 0) > \phi$ for all $x \notin \Omega_\gamma$.

Let Γ define the following level set of the optimal cost function:

$$\Gamma = \{x : V_N^*(x) \leq N\phi + \lambda\gamma\}.$$

Notice that this set is compact and non-empty.

Theorem 2 (ISS stability): Suppose that assumptions 2-6 hold for the optimization problem $P_N(\cdot)$. Let $\kappa_N(x)$ be the control law derived from the solution of $P_N(x)$ applied using a receding horizon policy. Then, for any $x(0) \in \Gamma$, the system to be controlled by the control law $u(k) = \kappa_N(x(k))$ is input-to-state stable with respect to the estimation error and the constraints are always satisfied, that is, $y(k) \in \mathcal{Y}$ and $x(k) \in \Gamma$, for all k. [4]

VI. CASE STUDY

Consider a continuously-stirred tank reactor [13], in which the reaction the reaction $A \rightarrow B$ takes place, being the measurable outputs the concentration of the reactant, C_A (mol/l), the temperature in the tank, T (K), and the temperature of the coolant, T_c (K). They vary according to the control input, which is the reference temperature of the coolant, T_r (K), according to the following set of ordinary differential equations (ODEs):

$$\frac{dC_A(t)}{dt} = \frac{q_0}{V} \cdot (C_{Af} - C_A(t)) - k_0 \cdot e^{\left(-\frac{E}{R \cdot T(t)}\right)} \cdot C_A(t), \quad (19a)$$

$$\frac{dT(t)}{dt} = \frac{q_0}{V} \cdot (T_f - T(t)) + \frac{(-\Delta H_r) \cdot k_0}{\rho \cdot C_p} \cdot e^{\left(-\frac{E}{R \cdot T(t)}\right)} \cdot C_A(t) + \frac{U \cdot A}{V \cdot \rho \cdot C_p} \cdot (T_c(t) - T(t)), \quad (19b)$$

$$\frac{dT_c(t)}{dt} = \frac{T_r(t) - T_c(t)}{\tau}. \quad (19c)$$

[4]The proof of this theorem is omitted in this version, due to the limited number of pages.

TABLE I
PARAMETERS OF THE SYSTEM

Param.	Definition	Value	Units
q_0	Input flow of the reactive	10	l/min
V	Liquid volume in the tank	150	l
k_0	Frequency constant	6×10^{10}	1/min
E/R	Arrhenius constant	9750	K
$-\Delta H_r$	Enthalpy of the reaction	10000	J/mol
UA	Heat transfer coefficient	70000	J/(min K)
ρ	Density	1100	g/l
C_p	Specific heat	0.3	J/(g K)
τ	Time constant	1.5	min
C_{Af}	C_A in the input flow	1	mol/l
T_f	Temperature (input flow)	370	K

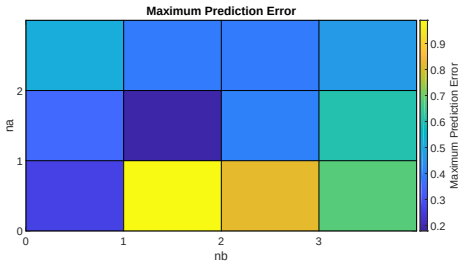

Fig. 2. One-norm of the maximum prediction error (one step ahead) for different memory horizons, after obtaining $\mathcal{L}$ and $\mathcal{P}$ via POKI and performing cross validation tests. Note that these values do not have physical sense, since we are adding one signal in mol/l and two in K.

The parameters of the system are shown in Table I. The sampling time is set to 30 s. The three output sensors add measurement errors to the signals, which are generated randomly using a uniform distribution with a maximum error of 2.5% of the measurement. The constraints are given by $0.368 \leq C_A \leq 0.553$ mol/l, $351.5 \leq T \leq 361.4$ K, $350 \leq T_c \leq 360$ K and $335 \leq T_r \leq 372$ K. The reference equilibrium point is set to $C_A^{\text{ref}} = 0.467$ mol/l, $T^{\text{ref}} = 355.925$ K, $T_c^{\text{ref}} = 354.5$ K and $T_r^{\text{ref}} = 354.5$ K.

The generation of the data sets is obtained via chirp signals for the training data set and pseudorandom binary sequences for the validation data sets. Once the raw data sets are obtained, the regression is constructed for different sets of the memory horizons n_a and n_b.

Setting the prediction horizon to $N = 4$ and applying POKI [15], the parameters $\mathcal{L}$ and $\mathcal{P}$ are obtained minimising the one-norm of s_4, where s is calculated as in equation (18). The minimum value is obtained for $n_a = n_b = 1$. Figure 2 shows the one-norm of the maximum prediction error $\|\mu\|_1$, for different memory horizons.

Both the stage and the terminal cost of the MPC are defined as follows:

$$\ell(x, u) = \|x - x^{\text{ref}}\|_Q^2 + \|u - u^{\text{ref}}\|_R^2,$$

$$V_f(x) = \|x - x^{\text{ref}}\|_P^2.$$

Q is set to $100 I_{n_y}$, $R = I_{n_u}$ (I denotes the identity matrix), and $\lambda = 10$. The reference is given by $x^{\text{ref}} = [C_A^{\text{ref}} \;\; T^{\text{ref}} \;\; T_c^{\text{ref}}]$ and $u^{\text{ref}} = T_r^{\text{ref}}$. Using the model with

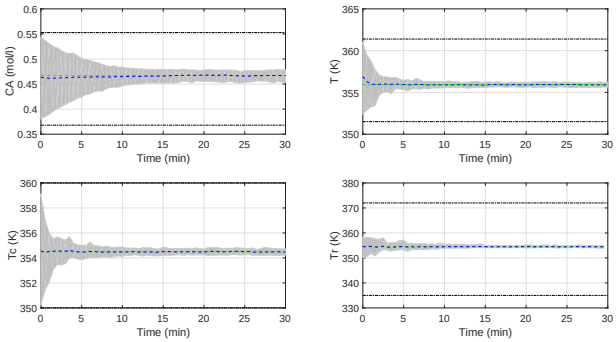

Fig. 3. One hundred simulations of the CHOKI MPC. The grey band groups the trajectories, the blue dotted line its mean, the green one the reference and the black one the constraints.

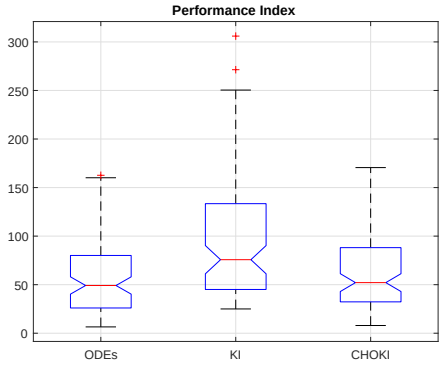

Fig. 4. Boxplot of the performance index of the control applied with different prediction models.

$n_a = n_b = 1$ the terminal cost is obtained solving a LQR for the linearised model around the reference point.

The controller is applied in one hundred closed-loop simulations, with random feasible initial state and random noise. The results are shown in Figure 3. Note that the outputs are stirred to the reference while the constraints are satisfied.

To compare the results, the same setting is used with two other controllers, where the model used to by the MPC is (a) the set of ODEs given by equations (19), in order to resemble the ideal case where the system is known, and (b) a well-tuned KI model with scalar parameters L and p. The comparison of the performance index is shown in Figure 4. Note that the CHOKI model performs better than the KI one. This performance index is obtained as

$$\Phi = \sum_{i=1}^{t_{\text{sim}}} \ell(x(i), u(i)).$$

Note that the interesting contribution of the CHOKI with respect to previous KI approaches relies not only in the improvement of the prediction (which in general leads to better closed-loop performance results), but also in the enlargement of the tightened constraints, enlarging the domain of attraction of the controller. Table II shows s_j (equation (18)) both for the KI and the CHOKI settings. The improvement in the tightened constraints averages a 53.97%, which implies that the multivariate bound is less conservative.

TABLE II

VALUE OF s_j, WHICH CHARACTERIZES THE REDUCTION IN THE CONSTRAINT AS A FUNCTION OF THE PREDICTION STEP, EACH UPPER VALUE FOR THE KI AND LOWER VALUE FOR THE CHOKI SETTING.

		μ	s_2	s_3	s_4
$C_A(\text{mol/l})$	KI	0.0094	0.0189	0.0283	0.0377
	CHOKI	0.0069	0.0096	0.0139	0.0174
$T(\text{K})$	KI	0.2700	0.5401	0.8090	1.0779
	CHOKI	0.0612	0.1295	0.3030	0.4827
$T_c(\text{K})$	KI	0.2159	0.4317	0.6456	0.8595
	CHOKI	0.1111	0.1886	0.3297	0.4982

REFERENCES

[1] E. F. Camacho and C. Bordons, *Model Predictive Control*, 2nd ed. Springer-Verlag, 2004.

[2] L. Ljung, *System identification - theory for the user*. Prentice Hall, 1999.

[3] A. Aswani, H. Gonzalez, S. S. Sastry, and C. Tomlin, "Provably safe and robust learning-based model predictive control," *Automatica*, vol. 49, no. 5, pp. 1216–1226, 2013.

[4] J. R. Salvador, D. Muñoz de la Peña, D. R. Ramírez, and T. Alamo, "Historian data based predictive control of a water distribution network," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 1716–1721.

[5] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, 2018.

[6] M. Maiworm, D. Limon, J. M. Manzano, and R. Findeisen, "Stability of gaussian process learning based output feedback model predictive control," *IFAC-PapersOnLine*, vol. 51, no. 20, pp. 455–461, 2018.

[7] D. Yu and J. Gomm, "Implementation of neural network predictive control to a multivariable chemical reactor," *Control Engineering Practice*, vol. 11, no. 11, pp. 1315–1323, 2003.

[8] J. P. Calliess, "Conservative decision-making and inference in uncertain dynamical systems," Ph.D. dissertation, University of Oxford, 2014.

[9] J. P. Calliess, "Lazily adapted constant kinky inference for nonparametric regression and model-reference adaptive control," *arXiv preprint arXiv:1701.00178*, 2016.

[10] G. Beliakov, "Interpolation of Lipschitz functions," *Journal of computational and applied mathematics*, vol. 196, no. 1, pp. 20–44, 2006.

[11] A. Sukharev, "Optimal method of constructing best uniform approximation for functions of a certain class," *Comput. Math. and Math. Phys.*, 1978.

[12] M. Canale, L. Fagiano, and M. C. Signorile, "Nonlinear model predictive control from data: a set membership approach," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 1, pp. 123–139, 2014.

[13] J. M. Manzano, D. Limon, D. Muñoz de la Peña, and J. P. Calliess, "Output feedback MPC based on smoothed projected kinky inference," *IET Control Theory & Applications*, 2019.

[14] J. M. Manzano, D. Limon, D. Muñoz de la Peña, and J. P. Calliess, "Robust data-based model predictive control for nonlinear constrained systems," *IFAC-PapersOnLine*, vol. 51, no. 20, pp. 505–510, 2018.

[15] J. P. Calliess, "Lipschitz optimisation for Lipschitz interpolation," *American Control Conference (ACC), 2017*, pp. 3141–3146, 2017.

[16] D. Limón, T. Alamo, F. Salas, and E. F. Camacho, "On the stability of constrained mpc without terminal constraint," *IEEE transactions on automatic control*, vol. 51, no. 5, pp. 832–836, 2006.

[17] D. Limon, T. Alamo, and E. Camacho, "Input-to-state stable mpc for constrained discrete-time nonlinear systems with bounded additive uncertainties," in *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, vol. 4. IEEE, 2002, pp. 4619–4624.