Implicit predictors in regularized data-driven predictive control

Manuel Klädtke and Moritz Schulze Darup

arXiv:2307.10750v1 [eess.SY] 20 Jul 2023

Abstract

We introduce the notion of implicit predictors, which characterize the input-(state)-output prediction behavior underlying a predictive control scheme, even if it is not explicitly enforced as an equality constraint (as in traditional model or subspace predictive control). To demonstrate this concept, we derive and analyze implicit predictors for some basic data-driven predictive control (DPC) schemes, which offers a new perspective on this popular approach that may form the basis for modified DPC schemes and further theoretical insights.

I. Introduction

Data-driven predictive control (DPC) is an increasingly popular control approach that utilizes linear combinations of collected trajectory data to make predictions instead of relying on a system model (see, e.g., [1–3]). While exact predictions and equivalence to model predictive control (MPC) are typically only established for linear time-invariant (LTI) systems and exact data (with some nonlinear extensions available; see, e.g., [2, 4]), modified versions of DPC show promising results even in the absence of these requirements (e.g., [1,3,5,6]). Most proposed DPC schemes use regularizations, which have demonstrated benefits and many different interpretations (see, e.g., [3,5]), including relationships to other approaches such as subspace predictive control (SPC) [7]. However, many of these interpretations (e.g. [8, Thms. 3 and 4]) are based on limit behavior, i.e., regularization weights approaching zero or infinity. In this study, we aim to characterize the behavior of DPC for finite weights, as this is how it is typically applied in practice. To achieve a general characterization, we do not focus on a specific system class or noise properties but, instead, just make (reasonable) assumptions about the data. To make DPC predictions more accessible and relatable to traditional schemes such as MPC, where consistency with input-state-output predictors is explicitly enforced via equality constraints, we introduce the notion of implicit predictors (specified in Def. 1 below). Implicit predictors can be interpreted as the predictive behavior that is implicitly attributed to the data-generating system by the predictive scheme, and we view it as a central and elucidating object with much to learn from. To show the benefit of this concept, we derive and analyze implicit predictors for a basic regularized DPC scheme, focusing on two proposed choices of (squared) 2-norm regularization (see [3]), and examine the effect of input and output constraints on these predictors. The paper is organized as follows. First, in Section II, we introduce the notion of implicit predictors, specify the objective of this paper, and summarize important preliminaries. In Section III, we derive implicit predictors for two types of regularized DPC and analyze the impact of constraints on these predictors. Finally, we conclude our work in Section IV and preview future opportunities we envision for implicit predictors.

M. Klädtke and M. Schulze Darup are with the Control and Cyberphysical Systems Group, Faculty of Mechanical Engineering, TU Dortmund University, Germany. E-mails: {manuel.klaedtke, moritz.schulzedarup}@tu-dortmund.de.

This paper is a **reprint** of a contribution to the IEEE Control Systems Letters. The DOI of the original paper is 10.1109/LCSYS.2023.3285104

II. Problem statement and preliminaries

One could argue that (multi-step) predictors, which establish a mapping from an initial state x_0 and an input (sequence) u_f to a predicted output (sequence) y_f, are at the heart of many predictive control schemes. Traditionally, such a predictor $\hat{\mathrm{y}}_f(x_0, \mathrm{u}_f)$ may be explicitly included in an optimal control problem (OCP) as an equality constraint $\mathrm{y}_f = \hat{\mathrm{y}}_f(x_0, \mathrm{u}_f)$. However, even if not included explicitly (such as in DPC), one may still observe the predictions following a similar pattern, which we formalize as follows.

Definition 1. *We call $\hat{\mathrm{y}}(x_0, \mathrm{u}_f)$ an implicit predictor for an OCP if including the constraint $\mathrm{y}_f = \hat{\mathrm{y}}(x_0, \mathrm{u}_f)$ does not alter the (set of) minimizers $(\mathrm{u}_f^*, \mathrm{y}_f^*)$ and the optimal value.*

For DPC, we interpret this definition as the predictive behavior implicitly attributed to the data-generating system by the DPC scheme. Given this interpretation, we take a somewhat opposing viewpoint to the bias-variance hypothesis in [3, Sec. V.C]. In fact, we claim that the choice of regularizer (and other parameters of the OCP) fully specifies a model class given by the structure of its resulting implicit predictor. Therefore, this concept should be seen as a tool to analyze existing schemes and explain their behavior. Although this new viewpoint is mainly theoretical in nature, practitioners can use its results to evaluate whether the predictive behavior of a given scheme matches their prior knowledge of the true system properties, and thus select an appropriate scheme similar to traditional model selection. In this spirit, we analyze an existing DPC scheme without making assumptions on the data-generating system class or properties of measurement noise, allowing for a general characterization. To prepare this analysis, we recall preliminaries on DPC and its relation to MPC and SPC, while highlighting the role of predictors in these schemes.

2475-1456 ©2023 IEEE

A. DPC and its relation to MPC

Instead of utilizing a discrete-time state-space model

$$x(k+1) = f(x(k), u(k), k) \tag{1a}$$
$$y(k) = g(x(k), u(k), k) \tag{1b}$$

with input $u \in \mathbb{R}^m$, state $x \in \mathbb{R}^n$, and output $y \in \mathbb{R}^p$ as in traditional MPC, predictions in DPC are realized based on previously collected trajectory data $(\mathrm{u}^{(1)}, \mathrm{y}^{(1)}), \ldots, (\mathrm{u}^{(\ell)}, \mathrm{y}^{(\ell)})$ via linear combinations

$$\begin{pmatrix} \mathrm{u}_{\text{pred}} \\ \mathrm{y}_{\text{pred}} \end{pmatrix} = \begin{pmatrix} \mathrm{u}^{(1)} \\ \mathrm{y}^{(1)} \end{pmatrix} a_1 + \ldots + \begin{pmatrix} \mathrm{u}^{(\ell)} \\ \mathrm{y}^{(\ell)} \end{pmatrix} a_\ell = \mathcal{D}a.$$

Here, the dimensions of the data matrix $\mathcal{D} \in \mathbb{R}^{L(m+p)\times \ell}$ and generator vector $a \in \mathbb{R}^\ell$ are specified by the length L of recorded (as well as predicted) trajectories and the number ℓ of data trajectories used for predictions. The rationale for this procedure is given by a result for linear time-invariant (LTI) systems, which can be described, e.g., by the specification

$$x(k+1) = Ax(k) + Bu(k) \tag{2a}$$
$$y(k) = Cx(k) + Du(k), \tag{2b}$$

of (1). For these LTI systems and assuming L is greater than the lag of the system, $\mathrm{image}(\mathcal{D})$ is equivalent to the set of all possible system trajectories if and only if [9]

$$\mathrm{rank}(\mathcal{D}) = Lm + n. \tag{3}$$

Note that the condition (3) not only signifies a minimum rank for data-driven predictions, but also the maximum rank that the data matrix $\mathcal{D}$ can achieve for exact data. For the case that the individual trajectories $(\mathrm{u}^{(i)}, \mathrm{y}^{(i)})$ are time-shifted sections of a single long trajectory, a popular sufficient condition for (3) is given by Willems' fundamental lemma [10]. To include the current initial condition of the system as a starting point for predicted trajectories, the predicted I/O-sequence is typically partitioned into a past section $(\mathrm{u}_p, \mathrm{y}_p)$ and a future section $(\mathrm{u}_f, \mathrm{y}_f)$ with N_p respectively N_f time-steps yielding

$$\begin{pmatrix} \mathrm{u}_{\text{pred}} \\ \mathrm{y}_{\text{pred}} \end{pmatrix} = \begin{pmatrix} \mathrm{u}_p \\ \mathrm{u}_f \\ \mathrm{y}_p \\ \mathrm{y}_f \end{pmatrix} = \begin{pmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{pmatrix} a = \mathcal{D}a.$$

The past section of a predicted trajectory is then forced to match the I/O-data ξ recorded in the most recent N_p time-steps during closed-loop operation, i.e., the equality constraints

$$\xi = \begin{pmatrix} \mathrm{u}_p \\ \mathrm{y}_p \end{pmatrix} = \begin{pmatrix} U_p \\ Y_p \end{pmatrix} a = W_p a$$

force any predicted trajectory in (5) to start with the most recently witnessed behavior of the system. In this context, the past trajectory ξ can also be interpreted as the state of a (usually non-minimal) state-space realization of the system and properly specifies its initial condition if N_p is chosen larger or equal to its lag [11].

Remark 1. *Although we have introduced the data-driven predictions in an I/O setting, they can be straightforwardly modified to a state-space setting [12], which we will use for visualization of a low dimensional example in Section III.*

Now, consider a classical OCP

$$\min_{u(k),x(k),y(k)} \sum_{k=0}^{N_f-1} \|y(k)\|_Q^2 + \|u(k)\|_R^2 \tag{4}$$
$$\begin{aligned} \text{s.t.} \quad & x(0) = x_0, \\ & x(k+1) = A\,x(k) + Bu(k), && \forall k \in \{0, ..., N_f-2\}, \\ & y(k) = C\,x(k) + Du(k), && \forall k \in \{0, ..., N_f-1\}, \\ & (u(k), y(k)) \in \mathcal{U}_k \times \mathcal{Y}_k, && \forall k \in \{0, ..., N_f-1\} \end{aligned}$$

for MPC with prediction horizon N_f, positive definite weighing matrices $R \in \mathbb{R}^{m\times m}, Q \in \mathbb{R}^{p\times p}$, convex constraint sets $\mathcal{U}_k \subseteq \mathbb{R}^m, \mathcal{Y}_k \subseteq \mathbb{R}^p$, as well as equality constraints specifying the initial state condition x_0 and predicted behavior based on (2). It has been shown in [1] that DPC given by

$$\min_{\mathrm{u}_f, \mathrm{y}_f, a} \|\mathrm{y}_f\|_{\mathcal{Q}}^2 + \|\mathrm{u}_f\|_{\mathcal{R}}^2 \tag{5a}$$
$$\text{s.t.} \quad \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix} = \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix} a, \tag{5b}$$
$$(\mathrm{u}_f, \mathrm{y}_f) \in \mathcal{U} \times \mathcal{Y} \tag{5c}$$

with $\mathcal{Q} := \mathrm{blkdiag}(Q, ..., Q)$, $\mathcal{R} := \mathrm{blkdiag}(R, ..., R)$, and

$$\mathcal{Y} := \left\{ \mathrm{y}_f \in \mathbb{R}^{pN_f} \,\middle|\, y(k) \in \mathcal{Y}_k,\ \forall k \in \{0, ..., N_f-1\} \right\}$$
$$\mathcal{U} := \left\{ \mathrm{u}_f \in \mathbb{R}^{mN_f} \,\middle|\, u(k) \in \mathcal{U}_k,\ \forall k \in \{0, ..., N_f-1\} \right\}$$

based on exact data generated by an LTI system is equivalent to the MPC in (4). Remarkably, and although the original theory behind exact predictions via linear combination of data does not apply to arbitrary nonlinear systems or with noise and disturbances in the data, DPC has shown good closed-loop performance when applied to these cases. A common feature of these successful applications, is the addition of a regularization term $h(a)$ to the cost function, where different choices of $h(a)$ can have different interpretations for its intended effect on the predictions ([1, 3, 5, 6]).

B. The role of predictors and SPC

In the case of linear MPC, its well-known multi-step predictor can be expressed as

$$\hat{\mathrm{y}}_{\text{MPC}}(x_0, \mathrm{u}_f) = \mathcal{O}x_0 + \mathcal{T}\mathrm{u}_f, \tag{6}$$

where $\mathcal{O}$ and $\mathcal{T}$ are often referred to as the extended observability matrix and the impulse response matrix, respectively (see, e.g. [13]). Using (6), we can state the OCP (4) also as

$$\min_{\mathrm{u}_f, \mathrm{y}_f} \|\mathrm{y}_f\|_{\mathcal{Q}}^2 + \|\mathrm{u}_f\|_{\mathcal{R}}^2 \tag{7}$$
$$\text{s.t.} \quad \mathrm{y}_f = \mathcal{O}x_0 + \mathcal{T}\mathrm{u}_f, \quad (\mathrm{u}_f, \mathrm{y}_f) \in \mathcal{U} \times \mathcal{Y},$$

2475-1456 ©2023 IEEE

which may be useful to eliminate optimization variables and to analyze the structure of its solution. Essentially, instead of using the state-space model (2) as a one-step predictor for each time-step in (4), consecutive applications of the one-step predictor yield the multi-step predictor (6) used in (7). Note that (6) trivially acts as an implicit predictor for the original MPC problem (4). By including the constraint $\mathrm{y}_f = \hat{\mathrm{y}}_{\text{MPC}}(x_0, \mathrm{u}_f)$, other constraints and variables can be eliminated, resulting in (7). However, this removal should be seen separately to Definition 1 and is not our focus.

Instead of estimating a state-space model (2) for MPC, an alternative approach based on subspace identification [13] is given by SPC [7]. Here, a linear multi-step predictor

$$\hat{\mathrm{y}}_{\text{SPC}}(\xi, \mathrm{u}_f) = K_{\text{SPC}} \begin{pmatrix} \xi \\ \mathrm{u}_f \end{pmatrix} \tag{8}$$

can be estimated directly from data as the solution

$$K_{\text{SPC}} := \arg\min_K \left\| Y_f - K \begin{pmatrix} W_p \\ U_f \end{pmatrix} \right\|_F^2 = Y_f \begin{pmatrix} W_p \\ U_f \end{pmatrix}^+ \tag{9}$$

to a least squares problem with the Frobenius norm $\|\cdot\|_F$. Crucially, this form of SPC is related to DPC as follows.

Lemma 1. *The SPC predictor* (8) *is an implicit predictor for the unregularized DPC problem* (5) *with exact trajectory data generated by an LTI system.*

Proof. Follows from [14, Thm. 1]. ■

While the result in [14] is actually stated as (5) being equivalent to SPC, we rather rephrase it in the context of implicit predictors. This is because (5b) cannot be removed (without changing the set of feasible ξ) as it also implies $\xi \in \operatorname{image}(W_p)$, which is not captured by the SPC predictor.

Introducing implicit predictors as explicit equality constraints can be useful to change the solution strategy for some OCPs. This, however, is not our focus in this work. Instead, we want to highlight this notion as a way to think about what kind of behavior is (implicitly) attributed to the system by DPC schemes, even though the predictor is not given as an equality constraint in any of the following cases.

III. Implicit predictors in regularized DPC

While in the case of deterministic LTI systems, adding additional trajectory data, i.e., columns to $\mathcal{D}$, cannot increase its rank past $\operatorname{rank}(\mathcal{D}) \leq mL + n$, this is typically not the case for systems with noise or nonlinearities. Instead, adding trajectory data generated by an LTI system with some output noise and persistently exciting enough input data (see, e.g., [10]) will almost surely increase its rank until $\mathcal{D}$ has full row rank for wide enough (at least square) data matrices [8, Lem. 3] and similar behavior can be observed for nonlinear systems with (or without) noise. Throughout this section, we will thus make the following assumption.

Assumption 1. *The data matrix $\mathcal{D}$ has full row rank.*

We want to emphasize that Assumption 1 takes into account the presence of measurement noise, and we do not make any further assumptions about the class of system generating the data. Once the OCP tuning parameters (including regularization) are determined, the solution becomes deterministic with respect to the provided data $\mathcal{D}$ (including potential output noise, which is included in W_p, Y_f and cannot be distinguished from the true output data), regardless of the type of system that generated it. Since the aim of this work is to characterize the predictive behavior and not to compare it to the (unknown) true system dynamic (as e.g., in [8]), this allows for a very general analysis. The assumption thus helps in shifting the focus away from what initially caused the lack of rank deficiency and towards what predictions are being made by DPC based on the given data (including noise). Furthermore, having more data than necessary in (3) is typically preferred, where some authors report good results with $\mathcal{D}$ being square [15, Sect. 5.2.4]. As the assumption aligns with this case, we believe it is a reasonable starting point but future work will consider data matrices with ranks between the practical minimum given by (3) and the maximum given by Assumption 1. Crucially, due to Assumption 1 there is a vector a satisfying (5b) for *any* triple $(\xi, \mathrm{u}_f, \mathrm{y}_f)$, resulting in arbitrary and meaningless predictions if the scheme is not suitably modified. However, adding a regularization $h(a)$ leading to the regularized DPC scheme

$$\min_{\mathrm{u}_f, \mathrm{y}_f, a} \|\mathrm{y}_f\|_{\mathcal{Q}}^2 + \|\mathrm{u}_f\|_{\mathcal{R}}^2 + h(a) \quad \text{s.t.} \quad \text{(5b)–(5c)} \tag{10}$$

has shown good results with respect to its predictive capabilities. Note that, while (5b) does not restrict the choice of any $(\xi, \mathrm{u}_f, \mathrm{y}_f)$ in terms of feasibility due to Assumption 1, it still defines a relation to a, which can be used to express the effect of $h(a)$ in the following way.

Lemma 2. *Under Assumption 1, the regularized DPC problem* (10) *is equivalent to*

$$\min_{\mathrm{u}_f, \mathrm{y}_f} \|\mathrm{y}_f\|_{\mathcal{Q}}^2 + \|\mathrm{u}_f\|_{\mathcal{R}}^2 + h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) \quad \textit{s.t.} \quad \text{(5c)} \tag{11}$$

with unique

$$h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) := \min_a \; h(a) \quad \textit{s.t.} \quad \text{(5b)}. \tag{12}$$

Proof. The OCP (10) can be trivially decoupled with the outer problem (11) and inner problem (12) by optimizing over one of the optimization variables while treating the others as parameters. The constraint (5c) is irrelevant to the inner problem because $(\mathrm{u}_f, \mathrm{y}_f)$ act as parameters there. On the other hand, the constraint (5b) is irrelevant to the outer problem because a is eliminated by solving the inner problem and (5b) does not restrict the choice of $(\xi, \mathrm{u}_f, \mathrm{y}_f)$ due to Assumption 1. While the minimizer $a^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$ to (12) may be non-unique depending on the choice of $h(a)$ and the number ℓ of data trajectories (i.e., the width of $\mathcal{D}$), the resulting optimal value of the inner problem $h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) = h(a^*(\xi, \mathrm{u}_f, \mathrm{y}_f))$ is trivially unique by optimality. ■

While any trajectory $(\xi, \mathrm{u}_f, \mathrm{y}_f)$ can be predicted via linear combination of collected data trajectories, the regularizer $h(a)$

assigns to every such trajectory a (possibly non-unique) optimal generator vector $a^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$ and associated unique cost $h^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$. Therefore, the regularizer $h(a)$ is, ideally, chosen in a way that incentivizes predictions that seem likely based on the given data in $\mathcal{D}$ by assigning them low (or even zero) cost while avoiding unlikely predictions via high costs.

We will now analyze this behavior for different choices of $h(a)$ and characterize the associated implicit predictors. Note that we treat the influence of equality constraints (5c) on the implied predictor separately in Section III.C.

A. (Squared) 2-norm regularization

Commonly seen in regularized DPC is (squared) 2-norm regularization, where we choose $h(a) = \lambda_a \|a\|_2^2$ with weighing parameter λ_a. Here, the inner problem (12) is a quadratic minimization with linear equality constraints and positive definite Hessian $2\lambda_a I$, which yields the unique minimizer

$$a^*(\xi, \mathrm{u}_f, \mathrm{y}_f) = \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix}^+ \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix} \tag{13}$$

and associated cost

$$h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) = \lambda_a \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix}^\top \left(\begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix} \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix}^\top \right)^{-1} \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix}.$$

Utilizing a block LDU decomposition of the weighing matrix, one can further show that

$$\begin{aligned} h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) &= \lambda_a (\mathrm{y}_f - \hat{\mathrm{y}}_{\text{SPC}})^\top \mathcal{Q}_{\text{reg}} (\mathrm{y}_f - \hat{\mathrm{y}}_{\text{SPC}}) \\ &\quad + \lambda_a \begin{pmatrix} \xi \\ \mathrm{u}_f \end{pmatrix}^\top \left(\begin{pmatrix} W_p \\ U_f \end{pmatrix} \begin{pmatrix} W_p \\ U_f \end{pmatrix}^\top \right)^{-1} \begin{pmatrix} \xi \\ \mathrm{u}_f \end{pmatrix}, \end{aligned} \tag{14}$$

where we introduced

$$\begin{aligned} \mathcal{Q}_{\text{reg}} &:= \left(Y_f \left(I - \Pi \right) Y_f^\top \right)^{-1}, \\ \Pi &:= \begin{pmatrix} W_p \\ U_f \end{pmatrix}^\top \left(\begin{pmatrix} W_p \\ U_f \end{pmatrix} \begin{pmatrix} W_p \\ U_f \end{pmatrix}^\top \right)^{-1} \begin{pmatrix} W_p \\ U_f \end{pmatrix}, \end{aligned} \tag{15}$$

and occasionally omit the arguments of $\hat{\mathrm{y}}_{\text{SPC}}(\xi, \mathrm{u}_f)$ for brevity. The expression (14) highlights the connection between DPC and SPC for this regularization. However, this regularizer not only incentivizes predictions that align with the SPC predictor but also introduces a bias associated with the additional quadratic cost term in (ξ, u_f). This (typically unwelcome) bias justifies the introduction of a projection in the regularizer [3], which we will analyze in Section III.B.

Even though predictions that align with the SPC predictor are incentivized with this regularization, the SPC predictor does not act as an implicit predictor in this case. In contrast to the unregularized deterministic case, where the rank deficiency of $\mathcal{D}$ implies a subspace for predictions, the implicit predictor is instead implied by optimality.

Theorem 3. *Consider the regularized DPC problem* (10) *with regularizer* $h(a) = \lambda_a \|a\|_2^2$ *and without additional input and output constraints. Under Assumption 1,*

$$\hat{\mathrm{y}}_{DPC}(\xi, \mathrm{u}_f) = \left(\lambda_a \mathcal{Q}_{reg} + \mathcal{Q} \right)^{-1} \lambda_a \mathcal{Q}_{reg} \hat{\mathrm{y}}_{SPC}(\xi, \mathrm{u}_f) \tag{16}$$

is an implicit predictor for this problem.

Proof. The implicit predictor (16) is the minimizer

$$\hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f) = \arg\min_{\mathrm{y}_f} \|\mathrm{y}_f\|_{\mathcal{Q}}^2 + \|\mathrm{u}_f\|_{\mathcal{R}}^2 + h^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$$

to an inner optimization problem, where u_f and ξ act as parameters. Since dropping any terms independent of y_f does not change the minimizer, we can simplify the problem as

$$\begin{aligned} &\arg\min_{\mathrm{y}_f} \quad \lambda_a (\mathrm{y}_f - \hat{\mathrm{y}}_{\text{SPC}})^\top \mathcal{Q}_{\text{reg}} (\mathrm{y}_f - \hat{\mathrm{y}}_{\text{SPC}}) + \mathrm{y}_f^\top \mathcal{Q} \mathrm{y}_f \\ &= \arg\min_{\mathrm{y}_f} \quad \mathrm{y}_f^\top \left(\lambda_a \mathcal{Q}_{\text{reg}} + \mathcal{Q} \right) \mathrm{y}_f - 2 \hat{\mathrm{y}}_{\text{SPC}}^\top \lambda_a \mathcal{Q}_{\text{reg}} \mathrm{y}_f, \end{aligned}$$

which yields an unconstrained quadratic minimization problem with the minimizer given by (16). Now, since $\hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f)$ is the parametric minimizer for any (ξ, u_f), the minimizers $(\mathrm{u}_f^*, \mathrm{y}_f^*)$ to the regularized DPC problem must naturally satisfy the relation $\mathrm{y}_f^* = \hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f^*)$ for any ξ. Hence, including the equality constraint $\mathrm{y}_f = \hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f)$ with the regularized DPC problem does not change its optimal value or minimizers, making $\hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f)$ an implicit predictor of this OCP. ■

While this implicit predictor does not match the SPC predictor, it still represents a subspace just like the latter and is visualized in Fig. 1 for different regularization weightings λ_a. For very high λ_a, the predictor starts aligning with the SPC predictor (and even matches it for $\lambda_a \to \infty$) as one would expect from the costs given by (14). For lower λ_a, the implicit predictor tilts towards $\lim_{\lambda_a \to 0} \hat{\mathrm{y}}_{\text{DPC}}(\xi, \mathrm{u}_f) = 0$ as seen in Fig. 1.a, which is also reasonable, since it represents the optimal solution to $\arg\min_{\mathrm{y}_f} \|\mathrm{y}_f\|_{\mathcal{Q}}^2$. These behaviors perfectly match the viewpoint that "control and identification regularize each other" [3] in DPC. However, we emphasize that this means the (implicitly) predicted system behavior of regularized DPC is linear and either consistent with the SPC predictor (for $\lambda_a \to \infty$) or more optimistic than SPC in the sense that it biases these most likely predictions (in the least squares sense, see (9)) towards more favorable predictions given by the control objective. While this optimism in the predictions may indeed be one reason why DPC performs better than SPC in some case studies (e.g. [3, Fig. 2]), it should be recognized that this also leads to potentially undesirable predictive behavior, which we will highlight when discussing additional constraints in Section III.C.

B. Projection-based (squared) 2-norm regularization

While we only introduced Π in (14) for notation purposes, we next exploit its role as a projection matrix. However, instead of Π itself, we will focus on $I - \Pi$, which is an orthogonal projector on $\ker \left(\begin{pmatrix} W_p^\top & U_f^\top \end{pmatrix}^\top \right)$. Intuitively, these projection matrices can be thought of as splitting any $a = \Pi a + (I - \Pi) a$ into a

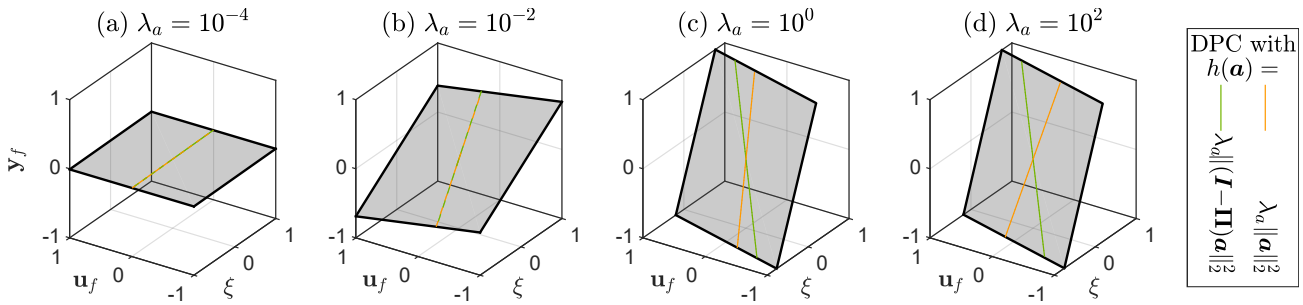

Figure 1: The implicit predictor (grey) is equal for both analyzed unconstrained DPC schemes and its structure is given by a subspace that "tilts" between $\hat{\mathrm{y}}_{\mathrm{DPC}}(\xi, \mathrm{u}_f) = 0$ and the SPC predictor (8) depending on λ_a. The parametric DPC solutions $(\xi, \mathrm{u}_f^*(\xi), \mathrm{y}_f^*(\xi))$ for the different regularizations (green/orange) evolve on this subspace, confirming its validity.

part Πa that aligns with the SPC solution and a part $(I - \Pi)a$ that does not. In [3], the regularizer $h(a) = \lambda_a\|(I - \Pi)a\|_2^2$ was introduced because, although it penalizes deviations from the SPC predictor just as $h(a) = \lambda_a\|a\|_2^2$, it does not introduce any further bias like the latter in (14). In the following, we will phrase this important result in our framework and subsequently analyze the associated implicit predictor. We first note that the inner problem given by (12) is, again, a quadratic minimization problem with linear constraints and the same (unique) minimizer as before in (13). This can easily be verified by plugging $a^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$ into the KKT system of (12). Uniqueness of $a^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$ may not be immediately obvious, since the Hessian $2\lambda_a(I-\Pi)^\top(I-\Pi) = 2\lambda_a(I-\Pi)$ is only positive semi-definite due to $(I - \Pi)$ being a projector on a subspace of $\mathbb{R}^\ell$. However, since

$$\ker \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix} \subset \ker \begin{pmatrix} W_p \\ U_f \end{pmatrix}$$

and $(I-\Pi)$ projects onto the latter kernel, we have $(I-\Pi)a \neq 0$ for any nonzero a in the former kernel, i.e., the Hessian and the constraint matrix only share the trivial nullspace by construction, making the KKT matrix nonsingular and thus $a^*(\xi, \mathrm{u}_f, \mathrm{y}_f)$ unique [16, Sect. 10.1.1]. The additional cost introduced by the regularizer is therefore given as

$$h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) = \lambda_a \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix}^\top \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix}^{+,\top} (I - \Pi) \begin{pmatrix} W_p \\ U_f \\ Y_f \end{pmatrix}^{+} \begin{pmatrix} \xi \\ \mathrm{u}_f \\ \mathrm{y}_f \end{pmatrix},$$

which can be shown to be equivalent to

$$h^*(\xi, \mathrm{u}_f, \mathrm{y}_f) = \lambda_a(\mathrm{y}_f - \hat{\mathrm{y}}_{\mathrm{SPC}})^\top \mathcal{Q}_{\mathrm{reg}}(\mathrm{y}_f - \hat{\mathrm{y}}_{\mathrm{SPC}}) \tag{17}$$

by, again, utilizing block LDU decompositions of the inverses involved in the weighing matrix. The resulting cost is similar to (14), which should not be too surprising, since the projection-based regularization was introduced to remove the bias related to additional costs in (ξ, u_f) and only penalize deviations from $\hat{\mathrm{y}}_{\mathrm{SPC}}(\xi, \mathrm{u}_f)$, which is exactly captured by (17). Furthermore, this expression allows us to state the following result regarding the implicit predictor, which is, again, implied by optimality.

Theorem 4. *Consider the regularized DPC problem* (10) *with regularizer* $h(a) = \lambda_a\|(I-\Pi)a\|_2^2$ *and without additional input and output constraints. Under Assumption 1,* (16) *is an implicit predictor for this problem.*

Proof. The proof can be carried out analogously to Theorem 3 by noting that the only difference between the regularizer costs (17) and (14) is irrelevant to the implied predictor since it does not depend on y_f. ■

This result may be surprising to readers familiar with both DPC regularizations, since the minimizers $(\mathrm{u}_f^*, \mathrm{y}_f^*)$ can differ drastically depending on the chosen regularizer. However, it shows that the predictive aspect of both DPC schemes is the same in the sense that, while $(\mathrm{u}_f^*, \mathrm{y}_f^*)$ may differ (especially for increasing λ_a, as seen in Fig. 1.d), the relation $\mathrm{y}_f^* = \hat{\mathrm{y}}_{\mathrm{DPC}}(\xi, \mathrm{u}_f^*)$ holds for both as visualized in Fig. 1.

Remark 2. *We can easily extend Theorems 3 and 4 towards reference tracking cost* $\|\mathrm{y}_f - \mathrm{y}_{ref}\|_{\mathcal{Q}}^2$*, which adds a linear term in the cost function and thus yields an affine implicit predictor, which is biased towards the unconstrained minimum* y_{ref}*.*

C. Predictions affected by (output) constraints

As previously noted, the implied predictors associated with regularized DPC tend to predict more optimistically (with respect to the control objective) than the SPC predictor. This phenomenon will be further highlighted in this section, where we investigate the effect of additional constraints. In the following, we will restrict our attention to a polyhedral set $\mathcal{Y}$, which is a common assumption in (linear) predictive control. Similar results can be expected for differently shaped sets, although they are not as nice to characterize accurately. Moreover, we introduce $\mathcal{X} \subseteq \mathrm{dom}(\xi)$, which should not be seen as a constraint on an optimization variable (because ξ is a parameter of the OCP), but instead can be viewed as the set of realistically occurring initial conditions in practical operation.

Theorem 5. *Consider the regularized DPC problem* (10) *with regularizer* $h(a) = \lambda_a\|a\|_2^2$ *or* $h(a) = \lambda_a\|(I-\Pi)a\|_2^2$ *and with constraint sets* $\mathcal{U}, \mathcal{Y}$*, where the latter is a polyhedron. Under*

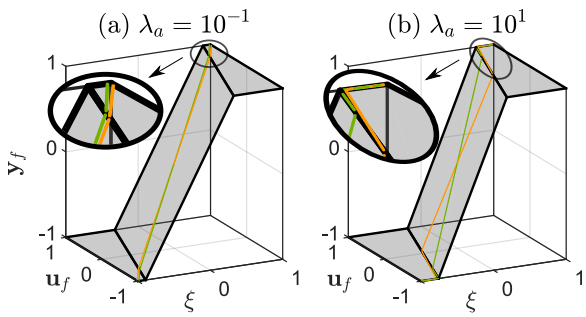

Figure 2: The implicit predictor (grey) for the analyzed DPC schemes with (output) constraints is a PWA function. For increasing λ_a, the parametric DPC solutions (orange/green, see Fig. 1 for a legend) tend to stay on the segment that matches the unconstrained predictor (if possible).

Assumption 1, an implicit predictor $\hat{\mathrm{y}}_{DPC}(\xi, \mathrm{u}_f)$ can be characterized as the minimizer to the multiparametric quadratic program (mpQP)

$$\begin{aligned} &\arg\min_{\mathrm{y}_f} \mathrm{y}_f^\top (\lambda_a \mathcal{Q}_{reg} + \mathcal{Q})\, \mathrm{y}_f - 2\hat{\mathrm{y}}_{SPC}^\top(\xi, \mathrm{u}_f)\lambda_a \mathcal{Q}_{reg} \mathrm{y}_f \\ &\quad \textit{s.t. } \mathrm{y}_f \in \mathcal{Y} \end{aligned} \tag{18}$$

with parameters (ξ, u_f), which is a continuous piecewise affine (PWA) function on partitions of the domain $\mathcal{X} \times \mathcal{U}$.

Proof. Derivation of the mpQP can be carried out analogously to Theorems 3 and 4, by noting that only the output constraints are explicitly relevant to the implied predictor and other constraints only specify the domain of its parametric solution. The structure of the minimizer follows from standard mpQP results (see, e.g., [17, Cor. 5.2]). ■

While PWA functions are not uncommon in (linear) MPC (mainly in relation to its explicit solution [17]), we want to stress that the PWA function characterized in Theorem 5 is in its interpretation very different to an optimal control law $\mathrm{u}_f^*(\xi)$, since it characterizes the predictive behavior we associate with regularized DPC. While this implicit predictor coincides with (16) on the set $\mathcal{Z} := \left\{(\xi, \mathrm{u}_f) \in \mathcal{X} \times \mathcal{U} \,\middle|\, (16) \text{ satisfies } \hat{\mathrm{y}}_{\mathrm{DPC}}(\xi, \mathrm{u}_f) \in \mathcal{Y}\right\}$, i.e., where the unconstrained solution naturally satisfies the output constraints, it changes drastically wherever this coincidence does not occur, as seen in Fig. 2. However, at this point, a discussion on the uniqueness of implicit predictors is due. While the implicit predictors derived so far are kept as general as possible (i.e., they are independent of input cost weights $\mathcal{R}$ and constraints $\mathcal{U}$), by Definition 1 agreement with the predictor needs to occur only at optimal points $(\mathrm{u}_f^*, \mathrm{y}_f^*)$ for it to be valid. In other words, if the actual solution to the constrained regularized DPC problem (10) naturally stays on $\mathcal{Z}$, i.e., we have $(\xi, \mathrm{u}_f^*) \in \mathcal{Z}$ for all $\xi \in \mathcal{X}$, then its predictive behavior is equally well explained by both the linear predictor (16) and the PWA solution to (18). Moreover, the DPC solution favors this phenomenon for increasing λ_a, as highlighted in Fig. 2, since deviations of the unconstrained predictor, in this case, are increasingly penalized by the regularizer. However, it should be obvious that (even with very large λ_a) a match of both predictors cannot occur for such ξ, where (16) cannot be satisfied for any $(\mathrm{u}_f, \mathrm{y}_f) \in \mathcal{U} \times \mathcal{Y}$. In that case, the DPC solution will definitely change its predictive behavior in favor of satisfying the output constraints in accordance with (18). In other words, regularized DPC as in (10) will simply provide a "very unlikely" solution if all "more likely" solutions (in the sense of proximity to its unconstrained predictions or even the SPC predictor) are infeasible. Consequentially, from the viewpoint of an explicit solution to the OCP [17], where $\mathcal{X}$ acts as a constraint of the parameter ξ, its set of feasible states is equal to $\mathcal{X}$ regardless of other constraints $\mathcal{U}, \mathcal{Y}$.

Finally, we want to highlight that the implicit predictor in Theorem 5 is independent of the input constraints, since $\mathcal{U}$ only affects its domain but, in contrast to $\mathcal{Y}$, not its structure. Therefore, the linear implicit predictor (16) remains valid in the absence of output constraints (i.e., $\mathcal{Y} = \mathbb{R}^{pN_f}$), even if additional inputs constraints $\mathcal{U} \subset \mathbb{R}^{mN_f}$ are considered.

D. A remark on the numerical example

To visualize the implicit predictors, we deliberately chose a very low dimensional example with $n = p = m = N_f = \mathcal{Q} = \mathcal{R} = 1, \ell = 3$ and treated it in a state-space setting (see Rem. 1). Furthermore, the constraint sets involved in Fig. 2 are given by $\mathcal{U} = \mathcal{Y} = [-1, 1]$. The data generating system is LTI with parameters $(A, B, C, D) = (2, -0.5, 1, 0)$ and zero-mean Gaussian measurement noise with variance $\sigma^2 = 0.01$. However, we would like to emphasize that the structure of the implicit predictor does not depend on the data generating system class but only the data itself and the OCP parameters.

IV. Conclusions and Outlook

By introducing the notion of implicit predictors, we related the input-(state)-output prediction behavior of regularized DPC schemes to more traditional predictive control schemes such as MPC and SPC, where the predictor is explicitly enforced as an equality constraint. The structure of these implicit predictors seems very relevant to us since they can be interpreted as the behavior that the data-driven predictions attribute to the system based on given data and independently of its actual (unknown) behavior. To demonstrate this concept, we derived implicit predictors for a basic regularized DPC scheme and analyzed their structure for two popular choices of (squared) 2-norm regularization and its dependence on constraints.

In future work, we will continue this structural analysis with general ranks beyond Assumption 1 and further common DPC modifications such as slack variables [1, 6], (terminal) equality constraints [6], and different regularization choices such as the 1-norm [1, 3] or general p-norms. Although currently limited to providing a novel perspective on current DPC practices, we believe that these analyses have the potential to establish a foundation for new schemes that promote advantageous properties of

2475-1456 ©2023 IEEE

the implicit predictor, and for new theoretical insights, including DPC stability proofs such as in [6], but based on the implicit predictor.

REFERENCES

[1] J. Coulson, J. Lygeros, and F. Dörfler. Data-enabled predictive control: In the shallows of the DeePC. *18th European Control Conference*, pp. 307–312, 2019.

[2] J. Berberich and F. Allgöwer. A trajectory-based framework for data-driven system analysis and control. *2020 European Control Conference*, pp. 1365–1370, 2020.

[3] F. Dörfler, J. Coulson, and I. Markovsky. Bridging direct and indirect data-driven control formulations via regularizations and relaxations. *IEEE Transactions on Automatic Control*, 68(2):883–897, 2023.

[4] M. Alsalti, J. Berberich, V. G. Lopez, F. Allgöwer, and M. A. Müller. Data-based system analysis and control of flat nonlinear systems. *2021 60th IEEE Conference on Decision and Control*, pp. 1484–1489, 2021.

[5] J. Coulson, J. Lygeros, and F. Dörfler. Regularized and distributionally robust data-enabled predictive control. *2019 IEEE 58th Conference on Decision and Control*, pp. 2696–2701, 2019.

[6] J. Berberich, J. Köhler, M. Müller, and F. Allgöwer. Data-driven model predictive control with stability and robustness guarantees. *IEEE Transactions on Automatic Control*, 66(4):1702–1717, 2021.

[7] W. Favoreel, B. De Moor, and M. Gevers. Spc: Subspace predictive control. *IFAC Proceedings Volumes*, 32(2):4004–4009, 1999. 14th IFAC World Congress 1999, Beijing, China.

[8] V. Breschi, A. Chiuso, and S. Formentin. Data-driven predictive control in a stochastic setting: a unified framework. *Automatica*, 152, 2023.

[9] I. Markovsky and F. Dörfler. Identifiability in the behavioral setting. *IEEE Transactions on Automatic Control*, 68(3):1667–1677, 2023.

[10] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor. A note on persistency of excitation. *Systems & Control Letters*, 54(4):325–329, 2005.

[11] I. Markovsky and P. Rapisarda. Data-driven simulation and control. *International Journal of Control*, 81:1946 – 1959, 2008.

[12] C. De Persis and P. Tesi. Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control*, 65:909–924, 2020.

[13] P. van Overschee and B. de Moor. *Subspace identification for linear systems: Theory, implementation, applications*. Kluwer Academic Publishers, 1996.

[14] F. Fiedler and S. Lucia. On the relationship between data-enabled predictive control and subspace predictive control. *2021 European Control Conference*, pp. 222–229, 2021.

[15] I. Markovsky and F. Dörfler. Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control*, 52:42–64, 2021.

[16] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[17] F. Borrelli, A. Bemporad, and M. Morari. *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2017.

2475-1456 ©2023 IEEE