Robust Data-Enabled Predictive Control: Tractable Formulations and Performance Guarantees

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3241282

Robust Data-Enabled Predictive Control: Tractable Formulations and Performance Guarantees

Linbin Huang†, Jianzhe Zhen†, John Lygeros, and Florian Dörfler

***Abstract*—We introduce a general framework for robust data-enabled predictive control (DeePC) for linear time-invariant (LTI) systems, which enables us to obtain robust and optimal control in a receding-horizon fashion based on inexact input and output data. Robust DeePC solves a min-max optimization problem to compute the optimal control sequence that is resilient to all possible realizations of the uncertainties in data within a prescribed uncertainty set. We present computationally tractable reformulations of the min-max problem with various uncertainty sets. Moreover, we show that even though an accurate prediction of the future behavior is unattainable due to inaccessibility of exact data, the obtained control sequence provides performance guarantees for the actually *realized* input and output cost in open loop. Finally, we demonstrate the performance of robust DeePC using high-fidelity simulations of a power converter system.**

***Index Terms*—Data-driven control, predictive control, regularization, robust control, robust optimization.**

I. Introduction

Data-driven control seeking an optimal control strategy from data is attracting increasing interest from both academia and industry. It can be applied in scenarios where data is readily available, but the system and uncertainty models are too complex to obtain or maintain. There are mainly two paradigms of data-driven control: 1) *indirect data-driven control* that first identifies a model and then conducts control design based on the identified model, and 2) *direct data-driven control* that circumvents the step of system identification and obtains control policies directly from data. Indirect data-driven control has a long history in control, where many methods have been developed for system identification [1]; the subsequent control design can use, for instance, model predictive control (MPC) [2]–[4]. Direct data-driven control gained increasing attention and became popular thanks to iterative feedback tuning [5], virtual reference feedback tuning [6], reinforcement learning [7], learning-based control [8]–[10], etc. The connection between indirect data-driven control and direct data-driven control was investigated in [11] from the perspective of regularizations and convex relaxations in optimization. A central promise is that direct data-driven control may have higher flexibility and better performance than indirect data-driven control thanks to the data-centric representation that avoids using a specific model from identification. Moreover, it is generally difficult to map uncertainty specifications from system identification over to robust control in indirect data-driven control, while, as we will show in this paper, this may become easier in direct data-driven control.

In recent years, a result formulated by Willems et al. has received renewed attention. This result, known as the *Fundamental Lemma* [12], shows that the subspace of trajectories of a linear time-invariant system can be obtained from the column span of a data Hankel matrix. Multiple direct data-driven methods have been proposed based on the Fundamental Lemma, e.g., [13]–[20]. Here we concentrate on the **d**ata-**e**nabl**e**d **p**redictive **c**ontrol (DeePC) proposed in [13], which uses input and output data to perform safe and optimal control (with quadratic cost) in a receding-horizon manner. DeePC has been successfully applied in many scenarios, e.g., power systems [21], [22], motor drives [23], and quadcopters [24].

When exact (noiseless and uncorrupted) data is accessible, DeePC can accurately predict the system's future behavior. However, in practice, exact data is in general not accessible due to measurement noise and input noise, which leads to inaccurate estimations and predictions and may degrade the control performance. In fact, a key question for data-driven control is: how does the system perform when applying control policies computed from inexact (noisy) data? For DeePC, it has been frequently observed that regularization is important to ensure good performance under noisy measurements [11], [15], [17]. In [25], we showed that including a quadratic regularization in DeePC enables the reformulation as a min-max problem, which provides robustness against uncertainties in the output data. Nonetheless, it still remains unclear whether the performance can be guaranteed under inexact data, and whether different structural assumptions on the uncertainties (e.g., with Hankel structure) can be taken into account to reduce the conservativeness induced from coarse robustification.

To this end, we present a robust DeePC framework that involves solving a min-max problem to robustify the control sequence against uncertainties in data. We justify our min-max formulation by showing that it leads to a performance guarantee. We consider different uncertainty sets as tight estimates when different types of data matrices (e.g., Hankel or Page matrices) are used. For instance, to make the considered set tight, one may incorporate Hankel structure on uncertainties when Hankel matrices are used, and may consider column-wise uncertainties when Page/trajectory matrices are used. For such uncertainty sets we explicitly show how the min-max problems can be reduced to tractable minimization problems.

The rest of this paper is organized as follows: Section II gives a brief review on DeePC. Section III presents robust DeePC. Section IV derives the tractable formulations under

†: The first two authors contributed equally to this work.

The authors are with the Department of Information Technology and Electrical Engineering at ETH Zürich, Switzerland. (Emails: linhuang@ethz.ch, trevorzhen@gmail.com, jlygeros@ethz.ch, dorfler@ethz.ch)

This research was supported by the SNSF under NCCR Automation, the ERC under the project OCAL (grant number 787845), and ETH Zurich Funds.

different uncertainty sets. Section V tests the robust DeePC on a power converter system. Section VI concludes the paper.

Notation: Let $\mathbb{Z}$ denote the set of integers, and $[n]$ the index set with cardinality $n \in \mathbb{Z}_{>0}$, i.e., $[n] = \{1, ..., n\}$. Let $\|x\|$ ($\|A\|$) denote the 2-norm of the vector x (the matrix A) and $\|x\|_1$ the 1-norm of x. For a vector x, we use $\|x\|_A^2$ to denote $x^\top Ax$ and x_i (or $(x)_i$) to denote the i-th element of x. For a matrix A, we use $\|A\|_Q$ to denote $\|Q^{\frac{1}{2}}A\|$, $\|A\|_F$ to denote the Frobenius norm of A, A_i (or $(A)_i$) to denote the i-th column of A, and A_{ij} (or $(A)_{ij}$) to denote the element in the i-th row and j-th column of A. We use $\mathbf{1}_n$ to denote a vector of ones of length n, $\mathbf{1}_{m\times n}$ to denote a m-by-n matrix of ones, and I_n to denote an n-by-n identity matrix (abbreviated as I when the dimensions can be inferred from the context). We use A^+ to denote the pseudoinverse of A, and $A^\perp = I - A^+A$ to denote the orthogonal projector onto the kernel of A. We use $[Z_0; Z_1; ...; Z_\ell]$ to denote the matrix $[Z_0^\top \ Z_1^\top \ \cdots \ Z_\ell^\top]^\top$. We use $\otimes$ to denote the Kronecker product.

II. Data-Enabled Predictive Control

A. Preliminaries on the Fundamental Lemma

Consider a discrete-time linear time-invariant (LTI) system

$$\begin{cases} x_{t+1} = Ax_t + Bu_t \\ y_t = Cx_t + Du_t \end{cases} \tag{1}$$

in a minimal representation, where $x_t \in \mathbb{R}^n$ is the state of the system, $u_t \in \mathbb{R}^m$ is the input vector, and $y_t \in \mathbb{R}^p$ is the output vector, all at time $t \in \mathbb{Z}_{\geq 0}$. Recall the respective extended observability matrix and convolution (impulse-response) matrices

$$\mathscr{O}_\ell(A, C) := [C; CA; ...; CA^{\ell-1}], \quad \text{and}$$
$$\mathcal{T}_N := \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & CA^{N-4}B & \cdots & D \end{bmatrix}. \tag{2}$$

The *lag* of the system (1) is defined by the smallest integer $\ell \in \mathbb{Z}_{\geq 0}$ such that the observability matrix $\mathscr{O}_\ell(A, C)$ has rank n, i.e., the state can be reconstructed from ℓ measurements.

In a data-driven setting, ℓ and n are generally unknown, but upper bounds can usually be inferred from knowledge of the system. Consider a length-T ($T \in \mathbb{Z}_{\geq 0}$) trajectory of (1): $\bar{u}^{\mathrm{d}} = [\bar{u}_0^{\mathrm{d}}; \bar{u}_1^{\mathrm{d}}; \ldots; \bar{u}_{T-1}^{\mathrm{d}}] \in \mathbb{R}^{mT}$, $\bar{y}^{\mathrm{d}} = [\bar{y}_0^{\mathrm{d}}; \bar{y}_1^{\mathrm{d}}; \ldots; \bar{y}_{T-1}^{\mathrm{d}}] \in \mathbb{R}^{pT}$, i.e., the length-$T$ restricted behavior. For the inputs $\bar{u}^{\mathrm{d}}$, define the Hankel matrix of depth L (with $T \geq L > \ell$) as

$$\mathscr{H}_L(\bar{u}^{\mathrm{d}}) := \begin{bmatrix} \bar{u}_0^{\mathrm{d}} & \bar{u}_1^{\mathrm{d}} & \cdots & \bar{u}_{T-L}^{\mathrm{d}} \\ \bar{u}_1^{\mathrm{d}} & \bar{u}_2^{\mathrm{d}} & \cdots & \bar{u}_{T-L+1}^{\mathrm{d}} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{u}_{L-1}^{\mathrm{d}} & \bar{u}_L^{\mathrm{d}} & \cdots & \bar{u}_{T-1}^{\mathrm{d}} \end{bmatrix}. \tag{3}$$

Similarly, for the outputs define the Hankel matrix $\mathscr{H}_L(\bar{y}^{\mathrm{d}})$. The length-$L$ restricted behavior of (1) equals the image of $\mathscr{H}_L(\bar{u}^{\mathrm{d}}, \bar{y}^{\mathrm{d}}) := [\mathscr{H}_L(\bar{u}^{\mathrm{d}}); \mathscr{H}_L(\bar{y}^{\mathrm{d}})]$ if and only if $\operatorname{rank}\big(\mathscr{H}_L(\bar{u}^{\mathrm{d}}, \bar{y}^{\mathrm{d}})\big) = mL+n$ [26, Corollary 19]. Note that this result extends and includes the original Fundamental Lemma [12, Theorem 1] which requires controllability and persistency of excitation (i.e., $\mathscr{H}_{L+n}(\bar{u}^{\mathrm{d}})$ must have full row rank) as sufficient conditions. These behavioral results can be leveraged for data-driven prediction as follows. Consider $T_{\mathrm{ini}}, N \in \mathbb{Z}_{\geq 0}$ so that $\operatorname{rank}\big(\mathscr{H}_{T_{\mathrm{ini}}+N}(\bar{u}^{\mathrm{d}}, \bar{y}^{\mathrm{d}})\big) = m(T_{\mathrm{ini}} + N) + n$. Then, we partition Hankel matrix $\mathscr{H}_{T_{\mathrm{ini}}+N}(\bar{u}^{\mathrm{d}}, \bar{y}^{\mathrm{d}})$ into

$$[\bar{U}_{\mathrm{P}}; \bar{U}_{\mathrm{F}}] := \mathscr{H}_{T_{\mathrm{ini}}+N}(\bar{u}^{\mathrm{d}}) \quad \text{and} \quad [\bar{Y}_{\mathrm{P}}; \bar{Y}_{\mathrm{F}}] := \mathscr{H}_{T_{\mathrm{ini}}+N}(\bar{y}^{\mathrm{d}}),$$

where $\bar{U}_{\mathrm{P}} \in \mathbb{R}^{mT_{\mathrm{ini}}\times H_c}$, $\bar{U}_{\mathrm{F}} \in \mathbb{R}^{mN\times H_c}$, $\bar{Y}_{\mathrm{P}} \in \mathbb{R}^{pT_{\mathrm{ini}}\times H_c}$, $\bar{Y}_{\mathrm{F}} \in \mathbb{R}^{pN\times H_c}$, and $H_c = T - T_{\mathrm{ini}} - N + 1$. In the sequel, the data in the partition with subscript P (for "past") will be used to implicitly estimate the initial condition of the system, whereas the data with subscript F will be used to predict the "future" behavior. Recall that the image of $\mathscr{H}_{T_{\mathrm{ini}}+N}(\bar{u}^{\mathrm{d}}, \bar{y}^{\mathrm{d}})$ spans all length-$(T_{\mathrm{ini}} + N)$ trajectories, that is, $[\bar{u}_{\mathrm{ini}}; u; \bar{y}_{\mathrm{ini}}; y]$ is a trajectory of (1) if and only if there exists $g \in \mathbb{R}^{H_c}$ so that

$$[\bar{U}_{\mathrm{P}}; \bar{Y}_{\mathrm{P}}; \bar{U}_{\mathrm{F}}; \bar{Y}_{\mathrm{F}}]g = [\bar{u}_{\mathrm{ini}}; \bar{y}_{\mathrm{ini}}; u; y]. \tag{4}$$

The initial trajectory $[\bar{u}_{\mathrm{ini}}; \bar{y}_{\mathrm{ini}}]$ can be thought of as setting the initial condition for the future trajectory $[u; y]$. In particular, if $T_{\mathrm{ini}} \geq \ell$, for every given u, the future outputs y are uniquely determined through (4) [27].

One can also use (Chinese) Page matrices or trajectory matrices as data-driven predictors [26], where the Page matrix (of depth L) for a signal $u \in \mathbb{R}^{mLH_c}$ is

$$\mathscr{P}_L(u) := \begin{bmatrix} u_0 & u_L & \cdots & u_{L(H_c-1)} \\ u_1 & u_{L+1} & \cdots & u_{L(H_c-1)+1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{L-1} & u_{2L-1} & \cdots & u_{LH_c-1} \end{bmatrix}, \tag{5}$$

and a trajectory matrix (of depth L) can be constructed from H_c independent length-L trajectories $u^i \in \mathbb{R}^{mL}$ ($i \in [H_c]$) as

$$\mathscr{T}_L(u) := \begin{bmatrix} u^1 & u^2 & \cdots & u^{H_c} \end{bmatrix}. \tag{6}$$

Notice that the columns in a Hankel matrix are correlated, while the columns in a Page/trajectory matrix are independent.

B. Review of DeePC

Based on (4), the DeePC method in [13] directly uses input and output data to perform safe and optimal control:

$$\begin{aligned} \min_{g, [u;y]\in\mathcal{C}} \ & \|u\|_R^2 + \|y - r\|_Q^2 \\ \text{s.t.} \ & [\bar{U}_{\mathrm{P}}; \bar{Y}_{\mathrm{P}}; \bar{U}_{\mathrm{F}}; \bar{Y}_{\mathrm{F}}]g = [\bar{u}_{\mathrm{ini}}; \bar{y}_{\mathrm{ini}}; u; y], \end{aligned} \tag{7}$$

where the input/output constraints are defined as $\mathcal{C} = \{[u; y] \in \mathbb{R}^{(m+p)N} \mid W[u; y] \leq w\}$ for $W \in \mathbb{R}^{n_w\times(m+p)N}$ and $w \in \mathbb{R}^{n_w}$ (we assume that $\mathcal{C}$ is large enough such that (7) is feasible). The positive definite matrix R and positive semi-definite matrix Q are the cost matrices. The vector $r \in \mathbb{R}^{pN}$ is a reference trajectory for the future outputs. DeePC solves (7) in a receding horizon fashion, that is, after calculating the optimal control sequence $u^\star$, we apply $(u_t, ..., u_{t+k-1}) = (u_0^\star, ..., u_{k-1}^\star)$ to the system for $k \leq N$ time steps. Then, we reinitialize (7) by updating $[\bar{u}_{\mathrm{ini}}; \bar{y}_{\mathrm{ini}}]$ to the most recent measurements, and setting t to $t + k$, to calculate the new optimal control for the next k time steps. We refer to, for example [11], [15], [28], for selecting the hyper-parameters.

Exact data and inexact data. We use $\bar{U}_{\mathrm{P}}$, $\bar{U}_{\mathrm{F}}$, $\bar{u}_{\mathrm{ini}}$, $\bar{Y}_{\mathrm{P}}$, $\bar{Y}_{\mathrm{F}}$, and $\bar{y}_{\mathrm{ini}}$ to denote the exact data generated from the system (1), which accurately captures the system dynamics according

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3241282

to (4); we use $\hat{U}_\mathrm{P}$, $\hat{U}_\mathrm{F}$, and $\hat{u}_\mathrm{ini}$ to denote the corresponding recorded (inexact) input data, and use $\hat{Y}_\mathrm{P}$, $\hat{Y}_\mathrm{F}$, and $\hat{y}_\mathrm{ini}$ to denote the corresponding measured (inexact) output data.

In this paper, we assume that there is no process noise in the state variables. The formulation in (7) assumes exact data, while in practice, exact data may not be accessible to the controller due to, for example, measurement noise and input noise (i.e., noise that enters through the input channels). With inexact data, the equality constraints in (7) may be infeasible, either with a small T or a large T [13]. Moreover, the equality constraints may restrict the feasible set of g to an undesired region that is determined by random noise. Therefore, we propose to penalize the violation of the equality constraints in the cost function (i.e., consider them as soft constraints [29])

$$\begin{aligned}\min_{g:[\hat{U}_\mathrm{F}g;\hat{Y}_\mathrm{F}g]\in\mathcal{C}} \quad & \|\hat{U}_\mathrm{F}g\|_R^2 + \|\hat{Y}_\mathrm{F}g - r\|_Q^2 \\ & +\lambda_u\|\hat{U}_\mathrm{P}g - \hat{u}_\mathrm{ini}\|^2 + \lambda_y\|\hat{Y}_\mathrm{P}g - \hat{y}_\mathrm{ini}\|^2,\end{aligned} \tag{8}$$

where $\lambda_u, \lambda_y > 0$ are the weighting coefficients. Note that (8) is always feasible as long as the set $\mathcal{C}$ is nonempty. In what follows, we adopt robust optimization approach to address uncertainties in the data matrices and trajectories of (8).

III. Robust DeePC

In this section, we propose a framework for robust DeePC to handle different types of uncertainties in input/output data.

A. Robust DeePC

To model the uncertainties in the input/output data, we introduce an *uncertainty vector* $\xi \in \mathcal{D}$, where $\mathcal{D} \subset \mathbb{R}^{n_\xi}$ is a *prescribed uncertainty set*. We consider parameterized data matrices $U_\mathrm{P}(\xi), Y_\mathrm{P}(\xi), U_\mathrm{F}(\xi), Y_\mathrm{F}(\xi)$, and initial trajectories $u_\mathrm{ini}(\xi)$, $y_\mathrm{ini}(\xi)$, which are affine functions of ξ and satisfy $U_\mathrm{P}(0) = \hat{U}_\mathrm{P}, Y_\mathrm{P}(0) = \hat{Y}_\mathrm{P}, U_\mathrm{F}(0) = \hat{U}_\mathrm{F}, Y_\mathrm{F}(0) = \hat{Y}_\mathrm{F}$, $u_\mathrm{ini}(0) = \hat{u}_\mathrm{ini}$ and $y_\mathrm{ini}(0) = \hat{y}_\mathrm{ini}$. For instance,

$$Y_\mathrm{P}(\xi) = \hat{Y}_\mathrm{P} + \sum_{j=1}^{n_\xi} Y_\mathrm{P}^{(j)}\xi_j \text{ and } y_\mathrm{ini}(\xi) = \hat{y}_\mathrm{ini} + \sum_{j=1}^{n_\xi} y_\mathrm{ini}^{(j)}\xi_j,$$

where $Y_\mathrm{P}^{(j)} \in \mathbb{R}^{pT_\mathrm{ini}\times H_c}$, $y_\mathrm{ini}^{(j)} \in \mathbb{R}^{pT_\mathrm{ini}}$ for every $j \in [n_\xi]$. The parameterized matrices and trajectories are built upon the inexact data. We assume that the uncertainties in the input/output data are captured by the set $\mathcal{D}$.

Assumption 1. *There exists a $\bar{\xi} \in \mathcal{D}$ such that $U_\mathrm{P}(\bar{\xi}) = \bar{U}_\mathrm{P}$, $Y_\mathrm{P}(\bar{\xi}) = \bar{Y}_\mathrm{P}$, $U_\mathrm{F}(\bar{\xi}) = \bar{U}_\mathrm{F}$, $Y_\mathrm{F}(\bar{\xi}) = \bar{Y}_\mathrm{F}$, $u_\mathrm{ini}(\bar{\xi}) = \bar{u}_\mathrm{ini}$, and $y_\mathrm{ini}(\bar{\xi}) = \bar{y}_\mathrm{ini}$.*

Then, we introduce the following *robust DeePC* formulation which robustifies (8) against the uncertainties in data

$$\begin{aligned}\min_{g\in\mathcal{G}} \max_{\xi\in\mathcal{D}} \quad & \|U_\mathrm{F}(\xi)g\|_R^2 + \|Y_\mathrm{F}(\xi)g - r\|_Q^2 \\ & +\lambda_u\|U_\mathrm{P}(\xi)g - u_\mathrm{ini}(\xi)\|^2 + \lambda_y\|Y_\mathrm{P}(\xi)g - y_\mathrm{ini}(\xi)\|^2,\end{aligned} \tag{9}$$

where $\mathcal{G} = \{g \mid \forall \xi \in \mathcal{D} : [U_\mathrm{F}(\xi)g; Y_\mathrm{F}(\xi)g] \in \mathcal{C}\}$. Note that the input/output constraints in $\mathcal{G}$ are also robustified. We assume that $\mathcal{G}$ is nonempty. Since $\mathcal{C}$ is a polytope and the uncertainty set $\mathcal{D}$ is a compact and convex set with nonempty relative interior, each semi-infinite constraint in $\mathcal{G}$ admits a tractable robust counterpart, that is, it can be reformulated into a finite set of convex constraints by using standard techniques from robust optimization [30]; see Table I. If some of the equality constraints in (7) are certain, that is, not affected by uncertainties, then one can include such equalities in $\mathcal{G}$ instead of softening them, in a similar way as in [25]. The minimizer $g^\star$ of (9) is then used to compute the control sequence $u^\star = \hat{U}_\mathrm{F}g^\star$, which is optimal with regards to the worst-case scenario in $\mathcal{D}$ (i.e., robust and optimal). We illustrate the relationship between the unknown system and robust DeePC in Fig. 1, where the realized control actions u_sys of the system may be different from $u^\star$ due to input noise. The realized output trajectory y_sys (in response to u_sys) is in general different from the predicted trajectory $y^\star = \hat{Y}_\mathrm{F}g^\star$. With inexact data, the (worst-case) prediction in (9) may be non-causal. We relegate the investigation on the causality issue to future research.

TABLE I

Tractable robust counterparts for $\forall\xi \in \mathcal{D} : (a + M^\top\xi)^\top x \le b$. LO: linear optimization problems. CQO: convex quadratic optimization problems. Note that ν is an additional optimization variable in the resulting tractable robust counterparts. This table is adapted from [31].

tractability	$\mathcal{D}$	robust counterpart
box (LO)	$\|\xi\|_\infty \le \rho$	$a^\top x + \rho\|Mx\|_1 \le b$
ellipsoidal (CQO)	$\|\xi\| \le \rho$	$a^\top x + \rho\|Mx\| \le b$
budget (LO)	$D\xi \le d$	$a^\top x + d^\top\nu \le b,$ $D^\top\nu = Mx,\ \nu \ge 0$
polyhedral (LO)	$\|\xi\|_\infty \le \rho,$ $\|\xi\|_1 \le \tau$	$a^\top x + \rho\|\nu\|_1 + \tau\|Mx - \nu\|_\infty \le b$

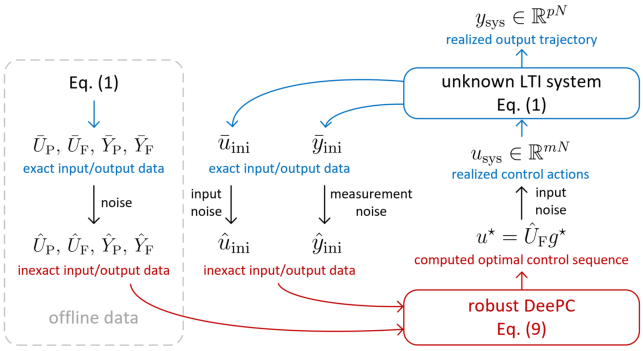

Fig. 1. Interaction between the system and robust DeePC. The data matrices $\hat{U}_\mathrm{P}$, $\hat{Y}_\mathrm{P}$, $\hat{U}_\mathrm{F}$, and $\hat{Y}_\mathrm{F}$ are obtained offline.

B. Performance Guarantees of Robust DeePC

Although (9) considers uncertainties in data, it is still not obvious that the realized cost can be guaranteed by applying $u^\star = \hat{U}_\mathrm{F}g^\star$ to the system. This is because the violations of the initial conditions are penalized in (9), and thus we generally have $\bar{U}_\mathrm{P}g \neq \bar{u}_\mathrm{ini}$ and $\bar{Y}_\mathrm{P}g \neq \bar{y}_\mathrm{ini}$, hence, the decision variable g cannot be used to accurately predict the realized future trajectory. As a first result, we link the *realized cost* of the system, defined by $c_\mathrm{realized} = \|u_\mathrm{sys}\|_R^2 + \|y_\mathrm{sys} - r\|_Q^2$, to the optimization cost of (9), and show how the realized cost can be certified by applying robust DeePC. Here "realized cost" refers to the cost accrued by the real system if u_sys is applied in open loop for the whole horizon N (see Fig. 1).

Theorem III.1. *If Assumption 1 holds and* $\|u_{\rm sys}-\bar{U}_{\rm F}g^\star\| \le \eta_{\rm i}$ *where* $g^\star$ *is the minimizer of* (9)*, then there exists sufficiently large* (λ_u, λ_y) *such that*

$$2\sqrt{c_{\rm opt}} + \eta_{\rm i}(\sqrt{2}\,\|I_{mN}\|_R + \|\mathcal{T}_N\|_Q) \ge \sqrt{c_{\rm realized}}\,,$$

where $c_{\rm opt}$ *is the minimum of* (9)*, and* $\mathcal{T}_N$ *is defined in* (2)*.*

Proof. If Assumption 1 holds, it follows the min-max problem

$$\begin{aligned} c_{\rm opt} \ge & \|\bar{U}_{\rm F}g^\star\|_R^2 + \|\bar{Y}_{\rm F}g^\star - r\|_Q^2 + \lambda_u\|\bar{U}_{\rm P}g^\star - \bar{u}_{\rm ini}\|^2 \\ & + \lambda_y\|\bar{Y}_{\rm P}g^\star - \bar{y}_{\rm ini}\|^2\,. \end{aligned} \tag{10}$$

Since the realized $[u_{\rm sys}; y_{\rm sys}]$ is an input/output trajectory departing from $[\bar{u}_{\rm ini}; \bar{y}_{\rm ini}]$, according to the Fundamental Lemma [12], there exists a $\bar{g}$ that satisfies

$$[\bar{U}_{\rm P}; \bar{Y}_{\rm P}; \bar{U}_{\rm F}; \bar{Y}_{\rm F}]\bar{g} = [\bar{u}_{\rm ini}; \bar{y}_{\rm ini}; u_{\rm sys}; y_{\rm sys}]. \tag{11}$$

By defining $\Delta_g = \bar{g} - g^\star$, we have

$$\begin{bmatrix} \bar{U}_{\rm P} \\ \bar{Y}_{\rm P} \\ \bar{U}_{\rm F} \\ \bar{Y}_{\rm F} \end{bmatrix} \Delta_g = \begin{bmatrix} \bar{u}_{\rm ini} - \bar{U}_{\rm P}g^\star \\ \bar{y}_{\rm ini} - \bar{Y}_{\rm P}g^\star \\ u_{\rm sys} - \bar{U}_{\rm F}g^\star \\ y_{\rm sys} - \bar{Y}_{\rm F}g^\star \end{bmatrix} =: \begin{bmatrix} \epsilon_{u_{\rm ini}} \\ \epsilon_{y_{\rm ini}} \\ \epsilon_{u_{\rm sys}} \\ y_{\rm sys} - \bar{Y}_{\rm F}g^\star \end{bmatrix}, \tag{12}$$

which implies that

$$\Delta_g = \begin{bmatrix} \bar{U}_{\rm P} \\ \bar{Y}_{\rm P} \\ \bar{U}_{\rm F} \end{bmatrix}^+ \begin{bmatrix} \epsilon_{u_{\rm ini}} \\ \epsilon_{y_{\rm ini}} \\ \epsilon_{u_{\rm sys}} \end{bmatrix} + \begin{bmatrix} \bar{U}_{\rm P} \\ \bar{Y}_{\rm P} \\ \bar{U}_{\rm F} \end{bmatrix}^\perp x\,,$$

$$\text{and} \quad \bar{Y}_{\rm F}\Delta_g = \bar{Y}_{\rm F}\begin{bmatrix} \bar{U}_{\rm P} \\ \bar{Y}_{\rm P} \\ \bar{U}_{\rm F} \end{bmatrix}^+ \begin{bmatrix} \epsilon_{u_{\rm ini}} \\ \epsilon_{y_{\rm ini}} \\ \epsilon_{u_{\rm sys}} \end{bmatrix} =: K\epsilon_{\rm ini} + \mathcal{T}_N\epsilon_{u_{\rm sys}}\,,$$

where $x \in \mathbb{R}^{H_c}$ and $\epsilon_{\rm ini} = [\epsilon_{u_{\rm ini}}; \epsilon_{y_{\rm ini}}]$, while the second equality follows from $\bar{Y}_{\rm F}[\bar{U}_{\rm P}; \bar{Y}_{\rm P}; \bar{U}_{\rm F}]^\perp = 0$ because the output trajectory is uniquely determined given an initial trajectory and a future input trajectory [27]. Note that $\mathcal{T}_N$ maps the N-step future inputs to the N-step future outputs; see (2).

By defining $\Lambda := \operatorname{diag}(\lambda_u I_{mT_{\rm ini}}, \lambda_y I_{pT_{\rm ini}})$, it follows from (10) and (12) that

$$\begin{aligned} & c_{\rm opt} - \|u_{\rm sys} - \epsilon_{u_{\rm sys}}\|_R^2 \\ \ge & \|y_{\rm sys} - K\epsilon_{\rm ini} - \mathcal{T}_N\epsilon_{u_{\rm sys}} - r\|_Q^2 + \|\epsilon_{\rm ini}\|_\Lambda^2 \\ \ge & \tfrac{1}{2}(\|y_{\rm sys} - r - K\epsilon_{\rm ini} - \mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q + \|\epsilon_{\rm ini}\|_\Lambda)^2 \\ \ge & \tfrac{1}{2}(\|y_{\rm sys} - r - \mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q - \|K\epsilon_{\rm ini}\|_Q + \|\epsilon_{\rm ini}\|_\Lambda)^2 \\ \ge & \tfrac{1}{2}\|y_{\rm sys} - r - \mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q^2\,. \end{aligned} \tag{13}$$

Here, the second inequality follows from $a^2 + b^2 \ge \frac{1}{2}(a+b)^2$; the third inequality holds thanks to the reverse triangle inequality; and the last inequality is satisfied if we take λ_u and λ_y large enough (by assumption) to ensure $\Lambda \succeq K^\top QK$ such that $\|\epsilon_{\rm ini}\|_\Lambda \ge \|K\epsilon_{\rm ini}\|_Q$. From (13), we further have

$$c_{\rm opt} \ge \tfrac{1}{4}(\sqrt{2}\|u_{\rm sys} - \epsilon_{u_{\rm sys}}\|_R + \|y_{\rm sys} - r - \mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q)^2\,,$$

considering that $a^2 + b^2 \ge \frac{1}{2}(a+b)^2$, and

$$\begin{aligned} & 2\sqrt{c_{\rm opt}} \ge \sqrt{2}\|u_{\rm sys} - \epsilon_{u_{\rm sys}}\|_R + \|y_{\rm sys} - r - \mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q \\ & \ge \sqrt{2}\|u_{\rm sys}\|_R - \sqrt{2}\,\|\epsilon_{u_{\rm sys}}\|_R + \|y_{\rm sys} - r\|_Q - \|\mathcal{T}_N\epsilon_{u_{\rm sys}}\|_Q \\ & \ge \sqrt{2}\|u_{\rm sys}\|_R + \|y_{\rm sys} - r\|_Q - \eta_{\rm i}(\sqrt{2}\|I_{mN}\|_R + \|\mathcal{T}_N\|_Q)\,, \end{aligned}$$

where the second inequality follows from the reverse triangle inequality, and the third due to the Cauchy–Schwarz inequality. Notice that $\sqrt{2}\,\|u_{\rm sys}\|_R + \|y_{\rm sys} - r\|_Q \ge \sqrt{c_{\rm realized}}$ because $a + b \ge \sqrt{a^2 + b^2}$ $(a, b \ge 0)$. This completes the proof. □

Theorem III.1 shows that by solving the min-max problem in (9), one can obtain a robust control sequence that guarantees the realized cost. The choice of (λ_u, λ_y) depends only on Q and K, where K is implicitly determined by the state-space matrices A, B, C, and D. Since Theorem III.1 only requires (λ_u, λ_y) to be sufficiently large, one can use an estimate on the upper bound of the largest eigenvalue of $K^\top QK$ and choose (λ_u, λ_y) accordingly, where we can estimate K using inexact data. The value of $\eta_{\rm i}$ increases with a higher level of input noise. In the absence of input noise (i.e., $u_{\rm sys} = \bar{U}_{\rm F}g^\star$) we construct a new upper bound on the realized cost.

Corollary III.2. *If Assumption 1 holds and the input data is exact, then there exist sufficiently large* (λ_u, λ_y) *such that*

$$2c_{\rm opt} \ge c_{\rm realized}\,.$$

Proof. Since $\epsilon_{u_{\rm sys}} = u_{\rm sys} - \bar{U}_{\rm F}g^\star = 0$ by assumption, the claim follows from (13) and the definition of $c_{\rm realized}$. □

With $\eta_{\rm i} = 0$, the inequality in Corollary III.2 is at least as tight as the one in Theorem III.1. The minimum of (9) will in general be increasing when a larger uncertainty set is needed to cover possible realized uncertainties, which indicates that the realized cost may also increase. Hence, one should also make the considered uncertainty set tight to reduce the conservativeness. We observe in our case studies (in Section IV) that $c_{\rm realized}$ is close to $2c_{\rm opt}$ when the considered uncertainty set is a tight estimate of the realized uncertainties.

IV. Tractable Reformulations of Robust DeePC

We next derive tractable reformulations for robust DeePC when various uncertainty sets are considered, and we will compare the conservativeness and the resulting robust counterparts. We first compactly and equivalently rewrite (9) as

$$\min_{g\in\mathcal{G}} \max_{\xi\in\mathcal{D}} \quad \|A(\xi)g - b(\xi)\|^2\,, \tag{14}$$

where $A(\xi)$ and $b(\xi)$ are affine functions of ξ, namely,

$$A(\xi) = A^{(0)} + \sum\nolimits_{j=1}^{n_\xi} A^{(j)}\xi_j, \text{ and } b(\xi) = b^{(0)} + \sum\nolimits_{j=1}^{n_\xi} b^{(j)}\xi_j,$$

where

$$\begin{aligned} A^{(j)} &= [\lambda_u^{\frac{1}{2}}U_{\rm P}^{(j)}; \lambda_y^{\frac{1}{2}}Y_{\rm P}^{(j)}; R^{\frac{1}{2}}U_{\rm F}^{(j)}; Q^{\frac{1}{2}}Y_{\rm F}^{(j)}] \in \mathbb{R}^{H_r\times H_c}, \\ b^{(j)} &= [\lambda_u^{\frac{1}{2}}u_{\rm ini}^{(j)}; \lambda_y^{\frac{1}{2}}y_{\rm ini}^{(j)}; 0; 0] \in \mathbb{R}^{H_r},\ \forall j \in [n_\xi]\,, \\ A^{(0)} &= [\lambda_u^{\frac{1}{2}}\hat{U}_{\rm P}; \lambda_y^{\frac{1}{2}}\hat{Y}_{\rm P}; R^{\frac{1}{2}}\hat{U}_{\rm F}; Q^{\frac{1}{2}}\hat{Y}_{\rm F}] \in \mathbb{R}^{H_r\times H_c}, \\ b^{(0)} &= [\lambda_u^{\frac{1}{2}}\hat{u}_{\rm ini}; \lambda_y^{\frac{1}{2}}\hat{y}_{\rm ini}, ; 0; Q^{\frac{1}{2}}r] \in \mathbb{R}^{H_r}, \end{aligned}$$

and $H_r = (m+p)(T_{\rm ini} + N)$. Eq. (14) can be considered as a robust least-squares problem, and we refer to, e.g., [30], [32], for existing techniques to solve this problem. Below we discuss various uncertainty quantifications and tractable reformulations for (14) sorted from coarse to fine, as listed in Table II. Note that the squared 2-norms in the soft constraints of (8) and (9) are essential for obtaining the robust least-square problem in (14) and deriving the tractable reformulations.

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3241282

TABLE II
COMPARISONS OF DIFFERENT UNCERTAINTY SETS FOR ROBUST DEEPC.

Section	uncertainty type	tractability	set different bounds for different columns	remove the effects of scaling matrices	consider Hankel structure
IV-A	unstructured	conic quadratic	✗	✗	✗
IV-B	column-wise	conic quadratic	✓	✗	✗
IV-C	interval	convex quadratic	✓	✓	✗
IV-D	structured	semi-definite program	✗	✓	✓

A. Unstructured Uncertainties

Consider the following variant of (14) with unstructured uncertainties in the following form

$$\min_{g\in\mathcal{G}} \max_{(\Delta_A,\Delta_b)\in\mathcal{D}_{\text{uns}}} \|(A^{(0)}+\Delta_A)g-(b^{(0)}+\Delta_b)\|, \quad (15)$$

where the unstructured uncertainty set is defined through

$$\mathcal{D}_{\text{uns}} = \{(\Delta_A,\Delta_b) \mid \|[\Delta_A\ \Delta_b]\|_F \le \rho_u\},$$

with $\rho_u \in \mathbb{R}_{\ge 0}$. Here we take the square-root of the quadratic objective function of (14), which will not affect the minimizer.

Note that (15) is a special (albeit unstructured) case of (14) with $\xi = [(\Delta_A)_1; (\Delta_A)_2; \dots; (\Delta_A)_{H_c}; \Delta_b]$, and it can be easily verified that the elements of $A(\xi) = A^{(0)} + \Delta_A$ and $b(\xi) = b^{(0)} + \Delta_b$ are affine functions of ξ. Moreover, Assumption 1 can be easily satisfied by choosing a sufficiently large ρ_u, which leads to a performance certificate according to Theorem III.1. The following result shows that (15) can be reformulated into a tractable conic quadratic problem.

Proposition IV.1. *A vector $g^\star$ is a minimizer of* (15) *if and only if $g^\star$ also minimizes the following problem*

$$\min_{g\in\mathcal{G}} \ \|A^{(0)}g-b^{(0)}\| + \rho_u\sqrt{\|g\|^2+1}. \quad (16)$$

Moreover, the minima of (15) *and* (16) *coincide.*

Proof. The claimed result can be proved by referring to [32, Theorem 3.1]. □

Eq. (16) can be seen as adding a regularizer $\sqrt{\|g\|^2+1}$ to the formulation in (8) [25]. If $\Delta_b = 0$, the regularizer becomes $\|g\|$. This regularizer robustifies the obtained control policy against inexact data, which avoids the incorrect solution ($u^\star = \hat{U}_{\text{F}} g^\star = 0$ and $y^\star = r$) of (7) when the data matrix becomes full row rank due to noise. We refer to [28] for a detailed discussion on the role of regularization in data-driven control. The considered uncertainties in (15) may be overly conservative for several reasons: 1) the worst-case realization of the uncertainties satisfies $\text{rank}([\Delta_A\ \Delta_b]) = 1$ [32], which cannot be achieved in general when Hankel matrices are used; 2) Hankel structure cannot be imposed on Δ_A; 3) the uncertainties are indirectly added to the raw data through the scaling matrices, e.g., $Q^{\frac{1}{2}}$ and $\lambda_y^{\frac{1}{2}} I$; 4) the columns in $[\Delta_A\ \Delta_b]$ are uncorrelated when Page matrices or trajectory matrices are used, but $\mathcal{D}_{\text{uns}}$ considers the Frobenius norm of $[\Delta_A\ \Delta_b]$ to be bounded which may impose correlations among the columns. In the following sections we alleviate these sources of conservatism. First we consider column-wise uncertainties.

B. Generalized Column-wise Uncertainties

Consider the following variant of (14) with generalized column-wise uncertainties, that is,

$$\min_{g\in\mathcal{G}} \max_{(\Delta_A,\Delta_b)\in\mathcal{D}_{\text{gcol}}} \|(A^{(0)}+\Delta_A)g-(b^{(0)}+\Delta_b)\|. \quad (17)$$

Here the generalized column-wise uncertainty set [33, Section 3] is defined column-by-column as

$$\mathcal{D}_{\text{gcol}} = \{(\Delta_A,\Delta_b) \mid \exists \rho_A \in \mathbb{R}^{H_c}: \|(\Delta_A)_i\| \le (\rho_A)_i, \forall i\in[H_c], \ f_j(\rho_A)\le 0,\ \forall j\in[J], \|\Delta_b\|\le\rho_b\},$$

where the function $f_j : \mathbb{R}^{H_c} \to [-\infty, +\infty]$ correlating the $(\rho_A)_i$ is a proper, closed and convex function for each $j \in [J]$.

Proposition IV.2. *If $\mathcal{D}_{\text{gcol}}$ admits a Slater point, then a vector $g^\star$ is a minimizer of* (17) *if and only if there exists a $(\lambda^\star, \{y_j^\star\}_{j=1}^J)$ such that $g^\star$ also minimizes the following convex optimization problem*

$$\begin{aligned} \min_{\substack{g\in\mathcal{G},\lambda \\ \{y_j\}_{j=1}^J}} \quad & \|A^{(0)}g-b^{(0)}\| + \sum\nolimits_{j\in[J]} \psi_j(y_j,\lambda_j) + \rho_b \\ \text{s.t.} \quad & \lambda_j \ge 0\ \forall j\in[J] \ \text{ and } \ \sum\nolimits_{j\in[J]} y_j \le -|g|, \end{aligned} \quad (18)$$

where $\psi_j : \mathbb{R}^{H_c}\times\mathbb{R}_{\ge 0} \to [-\infty,+\infty]$ denotes the perspective function[1] *of the conjugate function $f_j^* : \mathbb{R}^{H_c} \to [-\infty,+\infty]$ of f_j. Moreover, the minima of* (17) *and* (18) *coincide.*

Proof. According to [33, Theorem 1], we have

$$\begin{aligned} &\max\nolimits_{(\Delta_A,\Delta_b)\in\mathcal{D}_{\text{gcol}}} \|(A^{(0)}+\Delta_A)g-(b^{(0)}+\Delta_b)\| \\ &= \|A^{(0)}g-b^{(0)}\| + \max_{\rho_A: f_j(\rho_A)\le 0\ \forall j\in[J]} \rho_A^\top|g| + \rho_b. \end{aligned} \quad (19)$$

Then, one can reformulate the inner maximization into its dual and obtain (18) thanks to the strong duality of convex programs, which applies because $\mathcal{D}_{\text{gcol}}$ admits a Slater point. □

Eq. (17) is useful if (column-independent) Page/trajectory matrices are used. Different (correlated) bounds for different columns can be considered, which can come in handy when the data comes from a time-varying system or the columns are from independent experiments (i.e., trajectory matrices are used). The following result shows that for a special case of (17), where ρ_A is an uncorrelated vector, then (17) becomes a least-square problem with a weighted 1-norm regularization.

Corollary IV.3. *Consider $\mathcal{D}_{\text{gcol}} = \mathcal{D}_{\text{col}}$ in* (17)*, where*

$$\mathcal{D}_{\text{col}} := \{(\Delta_A,\Delta_b) \mid \|(\Delta_A)_i\| \le (\rho_A)_i, \forall i\in[H_c], \|\Delta_b\|\le\rho_b\},$$

then $g^\star$ is a minimizer of (17) *if and only if $g^\star$ minimizes*

$$\min_{g\in\mathcal{G}} \ \|A^{(0)}g-b^{(0)}\| + \rho_A^\top|g| + \rho_b\,. \quad (20)$$

Moreover, the minima of (17) *and* (20) *coincide.*

Proof. The result follows from (19) with a constant ρ_A. □

[1] We adopt the definition of the perspective function from [34], where $\psi(x,t) = tf^*(x/t)$ if $t > 0$, and $\psi(x,t) = \delta^*_{\text{dom}(f)}(x)$ if $t = 0$, and $\delta^*_{\mathcal{S}}$ denotes the support function over the set $\mathcal{S}$.

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3241282

In (17), the uncertainties are still indirectly added to the raw data through the scaling matrices. Although one can consider more sophisticated column-wise sets to cancel the scaling matrices, a tractable formulation may not be obtained. We next show that with interval uncertainties, one can directly consider uncertainties added to data and obtain a tractable formulation.

C. Interval Uncertainties

Consider the following special case of (14)

$$\min_{g\in\mathcal{G}} \max_{(\Delta_A,\Delta_b)\in\mathcal{D}_{\rm int}} \|(A^{(0)}+\Delta_A)g-(b^{(0)}+\Delta_b)\| \quad (21)$$

with the nonempty interval uncertainty set

$$\mathcal{D}_{\rm int}=\{(\Delta_A,\Delta_b)\mid |(\Delta_A)_{ij}|\le \bar{A}_{ij},\forall(i,j),|(\Delta_b)_i|\le\bar{b}_i,\forall i\},$$

where $\bar{A}\in\mathbb{R}^{H_r\times H_c}_{\ge 0}$ and $\bar{b}\in\mathbb{R}^{H_r}_{\ge 0}$.

Proposition IV.4. *A vector $g^\star$ is a minimizer of* (21) *if and only if there exists a $(\gamma^\star,\nu^\star)$ such that $(g^\star,\gamma^\star,\nu^\star)$ also minimizes the following convex quadratic problem*

$$\begin{array}{ll}\min\limits_{g\in\mathcal{G},\gamma,\nu} & \|\gamma+\bar{b}+\bar{A}\nu\|^2\\ \text{s.t.} & -\gamma\le A^{(0)}g-b^{(0)}\le\gamma\\ & -\nu\le g\le\nu.\end{array} \quad (22)$$

Moreover, the minimum of (21) *coincides with $\|\gamma^\star+\bar{b}+\bar{A}\nu^\star\|$.*

Proof. This proof is adapted from [30, Section 6.2], which shows that (21) can be equivalently reformulated as

$$\begin{array}{l}\min\limits_{g\in\mathcal{G},\tau}\ \tau^\top\tau\\ \text{s.t. }\ |(A^{(0)}g-b^{(0)})_i|+\bar{b}_i+\sum\limits_{j=1}^{H_c}\bar{A}_{ij}|g_j|\le\tau_i,\ \forall i\in[H_r],\end{array}$$

where the minimum of (21) coincides with $\|\tau^\star\|$. One can now eliminate τ by introducing γ_i and ν_j to replace $|(A^{(0)}g-b^{(0)})_i|$ and $|g_j|$, respectively, for every $i\in[H_r]$ and $j\in[H_c]$. □

With the interval uncertainty set, one can directly consider uncertainties on the raw data and cancel the effects of the scaling matrices. For instance, when the measurement error of the i-th output is bounded by $\tilde{y}_i\in\mathbb{R}_{\ge 0}$ and the input noise of the i-th input is bounded by $\tilde{u}_i\in\mathbb{R}_{\ge 0}$, we can choose

$$\bar{A}=\begin{bmatrix}\lambda_u^{\frac12}\mathbf{1}_{T_{\rm ini}\times H_c}\otimes\tilde{u}\\ \lambda_y^{\frac12}\mathbf{1}_{T_{\rm ini}\times H_c}\otimes\tilde{y}\\ R^{\frac12}\mathbf{1}_{N\times H_c}\otimes\tilde{u}\\ Q^{\frac12}\mathbf{1}_{N\times H_c}\otimes\tilde{y}\end{bmatrix}\quad\text{and}\quad\bar{b}=\begin{bmatrix}\lambda_u^{\frac12}\mathbf{1}_{T_{\rm ini}}\otimes\tilde{u}\\ \lambda_y^{\frac12}\mathbf{1}_{T_{\rm ini}}\otimes\tilde{y}\\ 0\\ 0\end{bmatrix}. \quad (23)$$

In the above settings, uncertainties with Hankel structure cannot be considered. We address this issue in what follows.

D. Structured Uncertainties

Consider (14) with the uncertainties residing within the following uncertainty set

$$\mathcal{D}=\mathcal{D}_{\rm struct}:=\{\xi\in\mathbb{R}^{n_\xi}\mid\|\xi\|\le\rho_s\}.$$

Recall that ξ enters $A(\xi)$ and $b(\xi)$ with affine structures. We show below how these affine structures allow us to construct a Hankel structure for the uncertainties. We start by noting that (14) can be reformulated as

$$\min_{g\in\mathcal{G}}\max_{\xi\in\mathcal{D}}\ \|D(g)\xi-c(g)\|^2, \quad (24)$$

where $D(g)$ and $c(g)$ that are affine functions of g, namely,

$$D(g)=D^{(0)}+\sum\nolimits_{\ell=1}^{H_c}D^{(\ell)}g_\ell,\ \text{and}\ c(g)=c^{(0)}+\sum\nolimits_{\ell=1}^{H_c}c^{(\ell)}g_\ell.$$

The equivalence between (14) and (24) directly follows from

$$\begin{aligned}(A^{(0)})_\ell&=c^{(\ell)}\ \forall\ell\in[H_c],\ (D^{(0)})_j=b^{(j)}\ \forall j\in[n_\xi],\\ (A^{(j)})_\ell&=(D^{(\ell)})_j\ \forall\ell\in[H_c]\ \forall j\in[n_\xi],\ b^{(0)}=c^{(0)}.\end{aligned}$$

Example 1 (Additive uncertainties with Hankel structure)**.** *Consider uncertainties directly on the inexact data (i.e., $\hat{u}^{\rm d}$, $\hat{u}_{\rm ini}$, $\hat{y}^{\rm d}$, and $\hat{y}_{\rm ini}$). Let $\xi=[\xi_1\in\mathbb{R}^{mT};\xi_2\in\mathbb{R}^{pT};\xi_3\in\mathbb{R}^{mT_{\rm ini}};\xi_4\in\mathbb{R}^{pT_{\rm ini}}]\in\mathcal{D}_{\rm struct}$ and $L=T_{\rm ini}+N$ such that*

$$\begin{aligned}[U_{\rm P}(\xi);U_{\rm F}(\xi)]&=\mathscr{H}_L(\hat{u}^{\rm d}+\alpha_1\xi_1)=[\hat{U}_{\rm P};\hat{U}_{\rm F}]+\alpha_1\mathscr{H}_L(\xi_1),\\ [Y_{\rm P}(\xi);Y_{\rm F}(\xi)]&=\mathscr{H}_L(\hat{y}^{\rm d}+\alpha_2\xi_2)=[\hat{Y}_{\rm P};\hat{Y}_{\rm F}]+\alpha_2\mathscr{H}_L(\xi_2),\\ u_{\rm ini}(\xi)&=\hat{u}_{\rm ini}+\alpha_3\xi_3,\ y_{\rm ini}(\xi)=\hat{y}_{\rm ini}+\alpha_4\xi_4,\end{aligned}$$

where α_1, α_2, α_3, and α_4 are the scaling coefficients. Note that $\alpha_1=\alpha_3=0$ in the absence of input noise.

We now show that (24) *with additive uncertainties in the input/output data matrices (that obey the prescribed Hankel structure) admit a compact representation. Notice that*

$$\begin{aligned}[U_{\rm P}(\xi);U_{\rm F}(\xi)]g&=[\hat{U}_{\rm P};\hat{U}_{\rm F}]g+\alpha_1\mathscr{H}_L(\xi_1)g\\ &=[\hat{U}_{\rm P};\hat{U}_{\rm F}]g+\alpha_1\mathscr{M}_L(g)\otimes I_m\xi_1,\\ [Y_{\rm P}(\xi);Y_{\rm F}(\xi)]g&=[\hat{Y}_{\rm P};\hat{Y}_{\rm F}]g+\alpha_2\mathscr{H}_L(\xi_2)g\\ &=[\hat{Y}_{\rm P};\hat{Y}_{\rm F}]g+\alpha_2\mathscr{M}_L(g)\otimes I_p\xi_2,\end{aligned}$$

where $\mathscr{M}_L(x)\in\mathbb{R}^{L\times(n_x+L-1)}$ ($x\in\mathbb{R}^{n_x}$) is defined as

$$\mathscr{M}_L(x)=\begin{bmatrix}x_1&x_2&\cdots&x_L&\cdots&x_{n_x}&&\\ &x_1&x_2&\cdots&x_L&\cdots&x_{n_x}&\\ &&\ddots&\ddots&&\ddots&&\ddots\\ &&&x_1&x_2&\cdots&x_L&\cdots&x_{n_x}\end{bmatrix}.$$

One can further partition $\mathscr{M}_L(g)$ into

$$[\mathscr{M}_L^{\rm P}(g)\in\mathbb{R}^{T_{\rm ini}\times T};\mathscr{M}_L^{\rm F}(g)\in\mathbb{R}^{N\times T}]:=\mathscr{M}_L(g).$$

According to the definition of $A(\xi)$ and $b(\xi)$, we then have

$$\begin{aligned}D(g)&=[D_1(g)\ \ D_2(g)],\\ c(g)&=[\lambda_u^{\frac12}\hat{U}_{\rm P};\lambda_y^{\frac12}\hat{Y}_{\rm P};R^{\frac12}\hat{U}_{\rm F};Q^{\frac12}\hat{Y}_{\rm F}]g\\ &\quad-[\lambda_u^{\frac12}\hat{u}_{\rm ini};\lambda_y^{\frac12}\hat{y}_{\rm ini};0;Q^{\frac12}r]\end{aligned} \quad (25)$$

in (24)*, where*

$$D_1(g)=\begin{bmatrix}\lambda_u^{\frac12}\alpha_1\mathscr{M}_L^{\rm P}(g)\otimes I_m&0\\ 0&\lambda_y^{\frac12}\alpha_2\mathscr{M}_L^{\rm P}(g)\otimes I_p\\ R^{\frac12}\alpha_1\mathscr{M}_L^{\rm F}(g)\otimes I_m&0\\ 0&Q^{\frac12}\alpha_2\mathscr{M}_L^{\rm F}(g)\otimes I_p\end{bmatrix},$$

$$\text{and}\quad D_2(g)=\begin{bmatrix}\lambda_u^{\frac12}\alpha_3 I_{mT_{\rm ini}}&0\\ 0&\lambda_y^{\frac12}\alpha_4 I_{pT_{\rm ini}}\\ 0&0\\ 0&0\end{bmatrix}.$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3241282

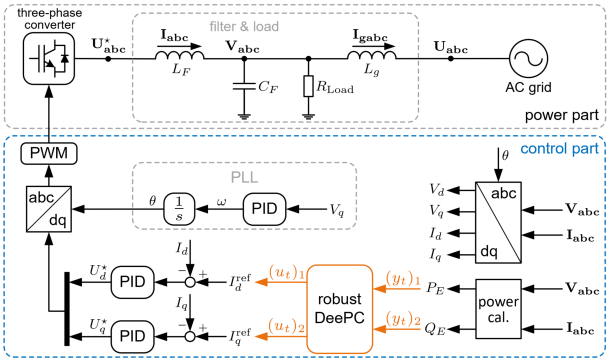

Fig. 2. Data-driven control for a grid-connected converter.

Example 1 shows that the formulation in (24) admits a compact representation of uncertainties with Hankel structure. The following result explicitly shows how (24) with $\mathcal{D} = \mathcal{D}_{\text{struct}}$ can be solved by considering a semi-definite program.

Proposition IV.5. *A vector $g^\star$ is a minimizer of* (24) *with $\mathcal{D} = \mathcal{D}_{\text{struct}}$ if and only if there exists a $(\tau^\star, \lambda^\star)$ such that $g^\star$ also minimizes the following semi-definite programming problem*

$$\begin{aligned} \min_{g\in\mathcal{G},\tau,\lambda} \quad & \tau \\ \text{s.t.} \quad & \begin{bmatrix} \tau - \lambda\rho_s^2 & 0 & c(g)^\top \\ 0 & \lambda I & D(g)^\top \\ c(g) & D(g) & I \end{bmatrix} \succeq 0. \end{aligned} \tag{26}$$

Moreover, the minima of (24) *and* (26) *coincide.*

Proof. The result can be proved by referring to [32], [35]. □

The structured uncertainty set in Example 1 is tighter than the other uncertainty sets when Hankel matrices are used as predictors, and thus it leads to less conservative results. However, it involves solving a semi-definite program, which usually reqires more computational effort. We will show that though the unstructured set is conservative, it can generally lead to satisfactory performance.

V. Simulation Results

In this section, we provide simulation results to illustrate the effectiveness of robust DeePC. We refer to the preprint of this paper [35] for more detailed simulations. We consider a grid-connected power converter in Fig. 2, and apply robust DeePC to regulate active/reactive power. Conventionally, power regulation is achieved by PID controllers. However, the power grid is highly complex, featuring variable load and generation, and in general unknown from the perspective of a converter, which significantly affects the performance of (fixed) PID controllers and may result in instabilities [36]. As a remedy, we use robust DeePC to perform data-driven, robust, and optimal control.

We choose active and reactive power to be the outputs, and robust DeePC provides control inputs for current references I_d^{ref}, I_q^{ref}; see Fig. 2. The inputs/outputs are all in per-unit values (p.u.). In our simulations, we use the nonlinear converter model; the same trends in the results are, however, also observed if one uses a linearization. The sampling time of robust DeePC is 1ms, and the main parameters are: $T_{\text{ini}} = 5$, $N = 25$, $T = 120$, $R = I$, $Q = 10^5 I$, $\lambda_u = \lambda_y = 10^5$,

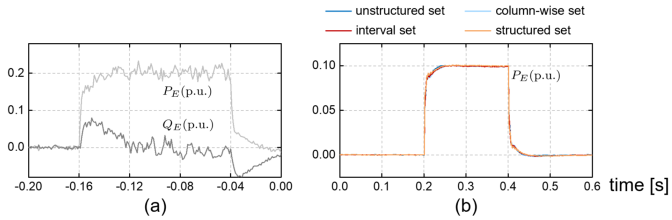

Fig. 3. Time-domain responses of the converter (a) during data collection and (b) with robust DeePC activated at $t = 0$s (control horizon $k = 5$).

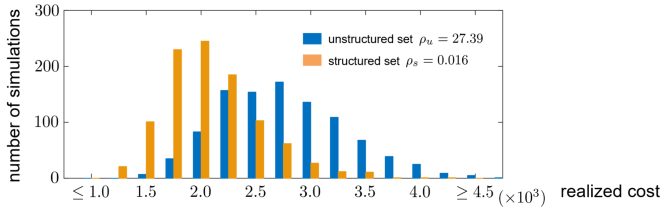

Fig. 4. Comparison of robust DeePC with structured set and unstructured set when the uncertainty samples are drawn from a structured set.

and $r = \mathbf{1}_N \otimes [P_0; Q_0]$, where P_0 and Q_0 are respectively the active and reactive power references. In robust DeePC, we assume that the output data is inexact and the input data is exact, while we show the exact data when demonstrating the results. Before activating robust DeePC, persistently exciting white noise signals are injected into the system through I_d^{ref} and I_d^{ref} for 0.12s to collect data, as shown in Fig. 3 (a). Fig. 3 (b) shows the active power responses of robust DeePC with different uncertainty sets: a) the upper/lower bounds of the considered interval set are $\pm\rho_i$ ($\rho_i = 0.001$); b) we choose $\rho_u = 23.49$, $\rho_A = 2.45 \times \mathbf{1}_{H_c}$, and $\rho_s = 0.016$, which are the smallest values such that the corresponding uncertainty sets contains the interval set; c) the outputs are subject to measurement noise (variance: 4×10^{-7}); and d) P_0 steps from 0 to 0.1(p.u.) at $t = 0.2$s and back to 0 at $t = 0.4$s. We observe that the converter has excellent performance in all cases [35].

We next compare the average *realized cost* of applying the open-loop optimal control sequence (of length N) obtained at $t = 0.2$s. We start with a study where Hankel matrices are used as predictors. Theoretical intuition from Section IV suggests that when Hankel matrices are used, robust DeePC with structured set should be less conservative than that with unstructured set. To test this intuition, we assume that the output data $\hat{y}^{\text{d}}$ and $\hat{y}_{\text{ini}}$ are affected by uncertainties residing in a structured set with $\rho_s = 0.016$, and draw 1000 samples that are uniformly-distributed over this set. Fig. 4 plots the results of applying robust DeePC with structured set and unstructured set (compactly containing the structured set). It shows that robust DeePC with structured set performs better than that with unstructured set, which confirms our intuition.

Section IV suggests that column-wise set and interval set are more appropriate choices when Page matrices are used. To validate this, we next consider Page matrices as predictors, by collecting a longer trajectory such that the constructed Page matrices have the same dimensions as the Hankel matrices. We assume that the output data is affected by uncertainties residing in an interval set with $\rho_i = 0.001$. We consider 1000 samples uniformly-distributed over this interval set. We still

choose for the column-wise set $\rho_A = 2.45 \times \mathbf{1}_{H_c}$ and $\rho_u = 23.49$ for the unstructured set, and we observe that robust DeePC with column-wise and unstructured sets achieve similar performance, while the average realized cost is improved by 20% when incorporating the interval set (a tighter estimate).

One advantage of column-wise uncertainties is that one can compactly consider different bounds for different columns in the data matrices. To demonstrate this, we use trajectory matrices as predictors and test the performance of robust DeePC when the columns in the trajectory matrices are subject to different levels of uncertainties, which could be the case when the columns are from different experiments. We consider 1000 uncertainty samples uniformly-distributed over the column-wise set where for all $i \in [H_c]$: $(\rho_A)_i = 0.63i$ and $\rho_b = 0.63H_c$. We choose $\rho_u = 324.73$ for the unstructured set so that it contains the considered column-wise set. Under this setting, we observe that the performance of applying the column-wise set in robust DeePC is 8.5% better than the unstructured set, thanks to the tighter uncertainty representation.

VI. Conclusions

This paper proposed a robust DeePC framework to perform data-driven, robust, and optimal control in a receding-horizon manner. Robust DeePC solves a min-max optimization problem to robustify the control sequence against uncertainties in data that is used for predictions. We showed that by applying robust DeePC, the realized cost of the system is bounded if the considered uncertainty set captures the exact data. We explicitly derived tractable formulations for robust DeePC when different geometries of uncertainty sets are incorporated. In particular, when Hankel matrices are used as predictors, we illustrated how uncertainties with Hankel structure can be taken into account in a structured uncertainty set to reduce the conservativeness. We confirm that 2-norm (1-norm) regularization corresponds to an inherent unstructured (column-wise) uncertainty set and plays a vital role in robustification. Future work should include closed-loop performance, feasibility, and stability certificates of robust DeePC.

References

[1] L. Ljung, "System identification," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–19, 1999.

[2] 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.

[3] W. Favoreel, B. De Moor, and M. Gevers, "SPC: Subspace predictive control," *IFAC Proceedings Volumes*, vol. 32, no. 2, 1999.

[4] M. Morari and J. H. Lee, "Model predictive control: past, present and future," *Computers & Chemical Engineering*, vol. 23, no. 4-5, pp. 667–682, 1999.

[5] H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, "Iterative feedback tuning: theory and applications," *IEEE control systems magazine*, vol. 18, no. 4, pp. 26–41, 1998.

[6] M. C. Campi, A. Lecchini, and S. M. Savaresi, "Virtual reference feedback tuning: a direct method for the design of feedback controllers," *Automatica*, vol. 38, no. 8, pp. 1337–1346, 2002.

[7] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, pp. 253–279, 2019.

[8] K. P. Wabersich, L. Hewing, A. Carron, and M. N. Zeilinger, "Probabilistic model predictive safety certification for learning-based control," *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 176–188, 2021.

[9] Y. Li, Y. Tang, R. Zhang, and N. Li, "Distributed reinforcement learning for decentralized linear quadratic control: A derivative-free policy optimization approach," *IEEE Trans. Autom. Control*, 2021.

[10] Y. Chen, Y. Shi, and B. Zhang, "Optimal control via neural networks: A convex approach," *ICLR*, 2019.

[11] F. Dorfler, J. Coulson, and I. Markovsky, "Bridging direct & indirect data-driven control formulations via regularizations and relaxations," *IEEE Trans. Autom. Control*, 2022.

[12] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

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

[14] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 909–924, 2019.

[15] J. Berberich, J. Köhler, M. A. Muller, and F. Allgower, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Trans. Autom. Control*, vol. 66, no. 4, pp. 1702–1717, 2021.

[16] H. J. Van Waarde, J. Eising, H. Trentelman, and M. Camlibel, "Data informativity: a new perspective on data-driven analysis and control," *IEEE Trans. Autom. Control*, vol. 65, no. 11, pp. 4753–4768, 2020.

[17] J. Coulson, J. Lygeros, and F. Dorfler, "Distributionally robust chance constrained data-enabled predictive control," *IEEE Trans. Autom. Control*, vol. 67, no. 7, pp. 3289–3304, 2022.

[18] Y. Lian and C. N. Jones, "Nonlinear data-enabled prediction and control," in *Learning for Dynamics and Control*, 2021, pp. 523–534.

[19] A. B. Alexandru, A. Tsiamis, and G. J. Pappas, "Data-driven control on encrypted data," *arXiv preprint arXiv:2008.12671*, 2020.

[20] J. G. Rueda-Escobedo, E. Fridman, and J. Schiffer, "Data-driven control for linear discrete-time delay systems," *IEEE Trans. Autom. Control*, vol. 67, no. 7, pp. 3321–3336, 2022.

[21] L. Huang, J. Coulson, J. Lygeros, and F. Dörfler, "Decentralized data-enabled predictive control for power system oscillation damping," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 3, pp. 1065–1077, 2021.

[22] ——, "Data-enabled predictive control for grid-connected power converters," in *2019 IEEE 58th Conference on Decision and Control*, 2019.

[23] P. G. Carlet, A. Favato, S. Bolognani, and F. Dörfler, "Data-driven predictive current control for synchronous motor drives," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2020.

[24] E. Elokda, J. Coulson, P. N. Beuchat, J. Lygeros, and F. Dörfler, "Data-enabled predictive control for quadcopters," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 18, pp. 8916–8936, 2021.

[25] L. Huang, J. Zhen, J. Lygeros, and F. Dörfler, "Quadratic regularization of data-enabled predictive control: Theory and application to power converter experiments," *19th IFAC Symposium on System Identification*, vol. 54, no. 7, pp. 192–197, 2021.

[26] I. Markovsky and F. Dörfler, "Identifiability in the behavioral setting," *IEEE Trans. Autom. Control*, 2022.

[27] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *International Journal of Control*, vol. 81, no. 12, pp. 1946–1959, 2008.

[28] I. Markovsky, L. Huang, and F. Dörfler, "Data-driven control based on the behavioral approach: From theory to applications in power systems," *IEEE Control Systems*, 2023, to appear.

[29] E. C. Kerrigan and J. M. Maciejowski, "Soft constraints and exact penalty functions in model predictive control," in *Control 2000 Conference*, 2000.

[30] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009.

[31] A. Ben-Tal, D. den Hertog, and J.-P. Vial, "Deriving robust counterparts of nonlinear uncertain inequalities," *Mathematical Programming*, vol. 149, no. 1, pp. 265–299, 2015.

[32] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on matrix analysis and applications*, vol. 18, no. 4, pp. 1035–1064, 1997.

[33] H. Xu, C. Caramanis, and S. Mannor, "Robust regression and lasso," *IEEE Trans. Information Theory*, vol. 56, no. 7, pp. 3561–3574, 2010.

[34] R. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[35] L. Huang, J. Zhen, J. Lygeros, and F. Dörfler, "Robust data-enabled predictive control: Tractable formulations and performance guarantees," *arXiv preprint arXiv:2105.07199*, 2021.

[36] L. Huang *et al.*, "Grid-synchronization stability analysis and loop shaping for PLL-based power converters with different reactive power control," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 501–516, 2019.