IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Distributionally robust uncertainty quantification via data-driven stochastic optimal control

Guanru Pan *Graduate Student Member, IEEE* and Timm Faulwasser* *Senior Member, IEEE*

Abstract— **This paper studies optimal control problems of unknown linear systems subject to stochastic disturbances of uncertain distribution. Uncertainty about the stochastic disturbances is usually described via ambiguity sets of probability measures or distributions. Typically, stochastic optimal control requires knowledge of underlying dynamics and is as such challenging. Relying on a stochastic fundamental lemma from data-driven control and on the framework of polynomial chaos expansions, we propose an approach to reformulate distributionally robust optimal control problems with ambiguity sets as uncertain conic programs in a finite-dimensional vector space. We show how to construct these programs from previously recorded data and how to relax the uncertain conic program to numerically tractable convex programs via appropriate sampling of the underlying distributions. The efficacy of our method is illustrated via a numerical example.**

Keywords: Distributional ambiguity, optimal control, Willems' fundamental lemma, uncertainty propagation, polynomial chaos expansion

I. INTRODUCTION

In many real-world applications, stochastic disturbances pose significant challenges, such as distributed energy systems facing uncertain wind speed and renewable energy generation, or building control systems dealing with uncertain weather conditions and occupancy. To hedge against the uncertainty surrounding the disturbance statistics, distributionally robust formulations optimize over an *ambiguity set* of possible disturbance distributions ensuring robust satisfaction of equality and inequality constraints [1]. Additionally, the complexity and time-consuming nature of first principles modeling and system identification further motivates the need for data-driven approaches.

There are two prominent data-driven avenues to distributionally robust optimal control: data-based synthesis of ambiguity sets to capture the uncertainty surrounding the distribution of disturbances while requiring explicit knowledge of a system model [2]–[4] and robustness analysis of data-driven system descriptions with respect to uncertainty surrounding the distribution of the measurement noise [5]. However, uncertainty propagation through dynamics without explicit knowledge of the system model and considering distributional uncertainty of the disturbance is still an open problem. In this work, we address this gap by generalizing the data-driven description of stochastic linear systems based on Polynomial Chaos Expansion (PCE) from [6], [7] towards uncertainty surrounding the disturbance distribution.

Specifically, the present paper appears to be the first to combine data-driven descriptions of stochastic systems via PCE and Hankel matrices, exact convex reformulation of Gelbrich ambiguity sets, and exact reformulation of chance constraints towards distributionally robust stochastic optimal control without explicit model knowledge. The main contributions are threefold: (i) we present a novel formulation of ambiguity sets for distributionally robust optimization using PCE including an exact convex reformulation for Gelbrich ambiguity sets. Moreover, while [4], [5], [8] use the conditional value at risk to reformulate chance constraints, we consider an exact reformulation applicable under distributional uncertainty. (ii) in contrast to [9], which considers ambiguity sets specified by fixed values of the first two moments, we allow for ranges of the first two moments via Gelbrich sets. (iii) we present mild conditions under which a distributionally robust Optimal Control Problem (OCP) with Gelbrich ambiguity and stated in random variables can be equivalently reformulated as an uncertain conic program without explicit knowledge of the system matrices. We also propose an approach to approximate this uncertain conic program with sampled uncertainty distributions. Finally, we draw upon a simulation example to demonstrate the efficacy of the proposed scheme.

Notation: Given a vector $x \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{n\times m}$, we specify $\|x\|$ as the 2-norm and $\|M\| = \sqrt{\mathrm{tr}(MM^\top)}$ as the Frobenius norm. We denote the set of all positive semi-definite (positive definite) matrices in $\mathbb{R}^{n\times n}$ as $\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$). The principal square root of $Q \in \mathbb{S}^n_+$ is written as $Q^{\frac{1}{2}}$. The vectorization of $\{x_k\}_{k=0}^{N-1}$ is denoted as $x_{[0,N-1]}$.

II. PROBLEM STATEMENT AND PRELIMINARIES

We first revisit the essential notions of probability theory. For rigorous definitions, we refer to the textbook [10]. A measurable space is a pair $(\Omega, \mathcal{F})$ where Ω is the sample space and $\mathcal{F}$ is a σ-algebra on Ω. A *probability measure* on the measurable space $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0,1]$ with $\mu(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, \mu)$ is a probability space. A *random variable* V is a measurable function $V : \Omega \to \mathbb{R}^{n_v}$ from the probability space $(\Omega, \mathcal{F}, \mu)$ to the measurable space $(\mathbb{R}^{n_v}, \mathcal{B})$ where $\mathcal{B}$ represents the Borel σ-algebra. Moreover, an $\mathcal{L}^2$ random variable $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_v})$ is finite in the $\mathcal{L}^2$ norm, i.e., $\|V\|^2 \doteq \int_\Omega V(\omega)^\top V(\omega)\mathrm{d}\mu(\omega) < +\infty$. The random variable V induces the probability measure μ_V on $(\mathbb{R}^{n_v}, \mathcal{B})$, i.e., for all $\mathcal{E} \in \mathcal{B}$, $\mu_V(\mathcal{E}) = \mu(\{\omega \in \Omega \,|\, V(\omega) \in$

*: Corresponding author.
The authors gratefully acknowledge funding by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) under project number 499435839. GP and TF are with the Institute for Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund University, Dortmund, Germany guanru.pan@tu-dortmund.de and timm.faulwasser@ieee.org

$\mathcal{E}\}$), denoted as the *distribution* of the random variable V. For compactness, we write $V \sim \mu_V$. Consider two random variables $V, \tilde{V} \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_v})$. The expectation of V is written as $\mathbb{E}[V] \in \mathbb{R}^{n_v}$, its variance is $\mathbb{V}[V] \in \mathbb{R}^{n_v}$, and the covariance of V and $\tilde{V}$ is denoted by $\Sigma[V, \tilde{V}]$.

Definition 1 (Gelbrich distance [11]): Consider two tuples of mean vectors and covariance matrices (m, Γ) and $(\bar{m}, \bar{\Gamma})$, their Gelbrich distance $\mathbb{G}((m, \Gamma), (\bar{m}, \bar{\Gamma})) \doteq d \geq 0$ is

$$d = \sqrt{\|m - \bar{m}\|^2 + \operatorname{tr}(\Gamma + \bar{\Gamma} - 2(\bar{\Gamma}^{\frac{1}{2}} \Gamma \bar{\Gamma}^{\frac{1}{2}})^{\frac{1}{2}})}. \qquad \square$$

A. Stochastic linear time-invariant systems

We consider stochastic discrete-time Linear Time-Invariant (LTI) systems

$$X_{k+1} = AX_k + BU_k + EW_k, \quad X_0 = x_{\text{ini}} \tag{1a}$$
$$Y_k = CX_k + DU_k + FW_k, \tag{1b}$$

with state $X_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_x})$, input $U_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_u})$, output $Y_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_y})$, and stochastic disturbance $W_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_w})$ for $k \in \mathbb{N}$. Note that the stochastic processes X, Y, and U are adapted to the filtration containing all historical information, cf. [10]. In this paper, we consider a deterministic initial condition $x_{\text{ini}} \in \mathbb{R}^{n_x}$ for (1) and identically independently distributed (i.i.d.) (not necessarily Gaussian) disturbances $W_k \sim \mu_W$.

Instead of exact knowledge of μ_W, we model it as an element of a given ambiguity set. The most commonly used ambiguity sets employ the Wasserstein metric. However, tractable reformulations of Wasserstein ambiguity sets are limited to certain empirical distributions [1] or to ambiguity sets comprising Gaussians [12]. As an alternative, *Gelbrich ambiguity sets* include all distributions with moments that closely match a given empirical pair $(\bar{m}, \bar{\Gamma})$ based on the Gelbrich distance in Definition 1. Specifically, we consider the Gelbrich ambiguity set with a given radius $\rho \in \mathbb{R}^+$

$$\mathcal{A} \doteq \left\{ \mu_W \in (\mathbb{R}^{n_w}, \mathcal{B}) \;\middle|\; \begin{array}{c} \mu_W \in \mathcal{D}(m, \Gamma),\ \Gamma \succeq 0 \\ \mathbb{G}\left((m, \Gamma), (\bar{m}, \bar{\Gamma})\right) \leq \rho \end{array} \right\}. \tag{2}$$

Here $\mathcal{D}(m, \Gamma)$ is the set of distributions with mean $m \in \mathbb{R}^{n_w}$ and covariance $\Gamma \in \mathbb{S}_+^{n_w}$. It is worth to be noted that the Gelbrich ambiguity set is an outer approximation for the corresponding Wasserstein set [11]. Additionally, we remark that [9] considers the special case of more restrictive ambiguity sets with fixed first two moments, i.e. $\mathcal{D}(\bar{m}, \bar{\Gamma})$. This corresponds to Gelbrich sets with $\rho = 0$.

Moving from distributions (or probability measures) to random variables, we note that the ambiguity set induces an uncertainty set for the sequence of i.i.d. random variables $W_{[0,N-1]}$ with respect to $N \in \mathbb{N}$

$$\mathcal{W} \doteq \left\{ W_{[0,N-1]} \;\middle|\; \begin{array}{c} \forall i \neq k, \quad i, k \in [0, N-1], \\ W_k \sim \mu_W \in \mathcal{A},\ \Sigma[W_k, W_i] = 0 \end{array} \right\}. \tag{3}$$

Note that $\mathcal{A} \subset (\mathbb{R}^{n_w}, \mathcal{B})$ while $\mathcal{W} \subset (\mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_w}))^N$.

B. Model-based distributionally robust optimal control

Our analysis begins with a distributionally robust OCP with the explicit knowledge of the system model, while its data-driven counterpart is presented in Section IV-B. Consider the uncertainty set (3), we have

$$\min_{\bar{u}, K, \alpha, U, Y, X} \alpha \tag{4a}$$
$$\text{s.t. } \forall W_{[0,N-1]} \in \mathcal{W}, \quad \forall k \in \mathbb{I}_{[0,N-1]},$$
$$\textstyle\sum_{k=0}^{N-1} \mathbb{E}\left[\|Y_k\|_Q^2 + \|U_k\|_R^2\right] \leq \alpha, \tag{4b}$$
$$X_{k+1} = AX_k + BU_k + EW_k, \quad X_0 = x_{\text{ini}}, \tag{4c}$$
$$Y_k = CX_k + DU_k + FW_k, \tag{4d}$$
$$U_k = \bar{u}_k + \textstyle\sum_{i=0}^{k-1} K_{k,i} W_i, \quad U_0 = \bar{u}_0, \tag{4e}$$
$$\mathbb{P}[a_{u,i}^\top U_k \leq 1] \geq 1 - \varepsilon_u, \quad \forall i \in \mathbb{I}_{[1,N_u]}, \tag{4f}$$
$$\mathbb{P}[a_{y,i}^\top Y_k \leq 1] \geq 1 - \varepsilon_y, \quad \forall i \in \mathbb{I}_{[1,N_y]}. \tag{4g}$$

Given the uncertainy set $\mathcal{W} \subset (\mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_w}))^N$, we minimize the worst-case value $\alpha \in \mathbb{R}$ of the objective function over the horizon $N \in \mathbb{N}$ in (4a)–(4b). The objective function is the expected value of a quadratic form with $Q \in \mathbb{S}_+^{n_y}$ and $R \in \mathbb{S}_{++}^{n_u}$. We consider i.i.d. disturbances directly entering the dynamics in (4c)-(4d). Similar to [4], [13] we aim at affine and causal disturbance feedback. This is encoded in (4e) and it can be written as

$$U_{[0,N-1]} = \bar{u}_{[0,N-1]} + K_w W_{[0,N-1]},$$
$$K_w = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ K_{1,0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \mathbf{0} & \vdots \\ K_{N-1,0} & \cdots & K_{N-1,N-2} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{Nn_u \times Nn_w}. \tag{5}$$

Chance constraints are specified as individual half-space constraints by $a_{u,i} \in \mathbb{R}^{n_u}$, $i \in \mathbb{I}_{[1,N_u]}$, and $a_{y,i} \in \mathbb{R}^{n_y}$, $i \in \mathbb{I}_{[1,N_y]}$ with probabilities of $1-\varepsilon_u$ and $1-\varepsilon_y$, respectively, in (4f)-(4g).

We remark that the conceptual formulation (4) poses several challenges. First, the optimization involves infinite-dimensional $\mathcal{L}^2$ random variables. Second, distributional robustness requires (4b)–(4g) to be satisfied for all possible random variable sequences in $\mathcal{W}$, resulting in infinitely many infinite-dimensional constraints. To address these challenges, we use the PCE framework to reformulate the random variables, the ambiguity sets, and the chance constraints.

III. The PCE Perspective on Gelbrich Ambiguity

A. Primer on polynomial chaos expansion

The core idea of PCE is that $\mathcal{L}^2$ random variables can be expressed as a series expansion in a suitable basis [10]. To this end, consider an orthogonal polynomial basis $\{\phi^j(\xi)\}_{j=0}^{\infty}$ which spans $\mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R})$, i.e. $\langle \phi^i, \phi^j \rangle = \int_{-\infty}^{\infty} \phi^i\left(\xi(\omega)\right) \phi^j\left(\xi(\omega)\right) \mathrm{d}\mu(\omega) = \delta^{ij} \|\phi^j\|^2$ where δ^{ij} is the Kronecker delta. We remark that it is customary in PCE to consider $\phi^0 = 1$.

Definition 2 (Polynomial chaos expansion): The PCE of a random variable $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R})$ with respect to the basis $\{\phi^j\}_{j=0}^{\infty}$ is $V = \sum_{j=0}^{\infty} \mathsf{v}^j \phi^j$ with $\mathsf{v}^j = \langle V, \phi^j \rangle / \|\phi^j\|^2$, where v^j is called the jth PCE coefficient. $\square$

We remark that by applying PCE component-wise the jth PCE coefficient vector of a vector-valued random variable $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_v})$ reads $\mathsf{v}^j = \begin{bmatrix} \mathsf{v}^{1,j} & \mathsf{v}^{2,j} & \cdots & \mathsf{v}^{n_v,j} \end{bmatrix}^\top$, where $\mathsf{v}^{i,j}$ is the jth PCE coefficient of component V^i. Moreover, we introduce a shorthand of the matrix generated by horizontally stacking the PCE coefficients as $\mathsf{V}^{[0,L-1]} \doteq \begin{bmatrix} \mathsf{v}^0, \mathsf{v}^1, \dots, \mathsf{v}^{L-1} \end{bmatrix} \in \mathbb{R}^{n_v \times L}$.

Definition 3 (Exact PCE representation [14]): The PCE of a random variable $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_v})$ is said to be exact with dimension L if $V - \sum_{j=0}^{L-1} \mathsf{v}^j \phi^j = 0$. □

Furthermore, with Definition 3, the expectation $\mathbb{E}[V]$, the variance $\mathbb{V}[V]$, and the covariance $\Sigma[V, \tilde{V}]$ can be obtained from the PCE coefficients as $\mathbb{E}[V] = \mathsf{v}^0$, $\mathbb{V}[V] = \sum_{j=1}^{L-1} \mathsf{v}^j \circ \mathsf{v}^j \|\phi^j\|^2$, $\Sigma[V, \tilde{V}] = \sum_{j=1}^{L-1} \mathsf{v}^j \tilde{\mathsf{v}}^{j\top} \|\phi^j\|^2$, where $\mathsf{v}^j \circ \mathsf{v}^j$ refers to the Hadamard product [15].

B. PCE representation of disturbances

For i.i.d. (not necessarily Gaussian) disturbances W_k, $k \in \mathbb{N}$, we first construct an exact PCE of finite dimension. For starters, we denote the map $\Psi : \mathbb{S}_+^{n_w} \to \mathbb{R}^{n_w \times n_w}$, $\Gamma \mapsto \Psi(\Gamma)$ as a *generalized matrix square root* if it is bijective and satisfies $\Gamma = \Psi(\Gamma)\Psi(\Gamma)^\top$.

Consider ξ_k with $\mathbb{E}[\xi_k] = 0$ and $\Sigma[\xi_k, \xi_k] = I_{n_w}$ such that $W_k = m + \Psi(\Gamma)\xi_k$ holds. Notice that the elements of ξ_k—i.e. ξ_k^i, $i \in \mathbb{I}_{[1,n_w]}$—are independently distributed and satisfy $\mathbb{E}[\xi_k^i] = 0$ as well as $\mathbb{V}[\xi_k^i] = 1$. Using the basis $\{\phi_w^j(\xi_k)\}_{j=0}^{n_w} = \{1, \{\xi_k^i\}_{i=1}^{n_w}\}$ with polynomials of degree of at most 1, the exact and finite PCE of W_k is obtained as

$$W_k = m + \Psi(\Gamma)\xi_k = \textstyle\sum_{j=0}^{n_w} \mathsf{w}^j \phi_w^j(\xi_k), \tag{6}$$

with $\mathsf{w}^0 = m$ and $\mathsf{W}^{[1,n_w]} \doteq \begin{bmatrix} \mathsf{w}^1, \dots, \mathsf{w}^{n_w} \end{bmatrix} = \Psi(\Gamma)$.

For any finite horizon $N \in \mathbb{N}$ in OCP (4) and let the inputs U_k satisfy (4e) the following orthonormal basis

$$\{\phi^j(\boldsymbol{\xi})\}_{j=0}^{L-1} = \{1, \{\{\xi_k^i\}_{i=1}^{n_w}\}_{k=0}^{N-1}\}, \tag{7}$$

where $\boldsymbol{\xi} = [\xi_0^\top, ..., \xi_{N-1}^\top]^\top \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{Nn_w})$ and $L = Nn_w + 1$, allows exact PCEs for $(U, Y, W, X)_k$, $k \in \mathbb{I}_{[0,N-1]}$, cf. [6].

Applying Galerkin projection onto the basis in (7) yields the dynamics of the PCE coefficients

$$\mathsf{x}_{k+1}^j = A\mathsf{x}_k^j + B\mathsf{u}_k^j + E\mathsf{w}_k^j, \quad \mathsf{x}_0^j = \delta^{0j} x_{\text{ini}}, \tag{8a}$$

$$\mathsf{y}_k^j = C\mathsf{x}_k^j + D\mathsf{u}_k^j + F\mathsf{w}_k^j, \quad \forall j \in \mathbb{I}_{[0,L-1]} \tag{8b}$$

where δ^{0j} is the Kronecker delta [16]. Due to the i.i.d. property of W_k, the PCE coefficients for $W_{[0,N-1]}$ satisfy

$$\left[\mathbf{1}_N \otimes \mathsf{w}^0, I_N \otimes \mathsf{W}^{[1,n_w]}\right] = \mathsf{W}_{[0,N-1]}^{[0,L-1]} \tag{9}$$

where $\mathsf{W}_{[0,N-1]}^{[0,L-1]} \in \mathbb{R}^{Nn_w \times L}$ is the vertically stacked block matrix comprising $\{\mathsf{W}_k^{[0,L-1]}\}_{k=0}^{N-1}$.

At first glance, the PCE representation of W_k in (6) seemingly resembles a usual moment-based representation. However, using the generalized square root of the covariance, we obtain a linear parametrization of W_K, which in turn simplifies the data-driven uncertainty propagation. Furthermore, for all W_k collecting the normalized random variables ξ_k, $k \in \mathbb{I}_{[0,N-1]}$ in the basis (7), we obtain the coefficient dynamics (8). These dynamics are structurally similar to the original dynamics in random variables (1). Put differently, for all $j \in \mathbb{I}_{[0,L-1]}$ the coefficient dynamics (8) capture the influence of the corresponding disturbance component. We remark that considering the explicit state covariance propagation $\Sigma_{k+1} = (A+BK)\Sigma_k(A+BK)^\top + E\Sigma[W_k, W_k]E^\top$ would render it more difficult to work with data-driven system descriptions. We refer to [7] for a more detailed comparison of moment propagation and PCE.

C. Representation of Gelbrich ambiguity sets

The PCE reformulation of W_k in (6) suggest to translate the Gelbrich ambiguity set $\mathcal{A}$ to an uncertainty set of the PCE coefficients, i.e., translation to a set of matrices with real numbers. Specifically, the distributions in $\mathcal{A}$ are bijectively paired to the PCE coefficients matrices by the map

$$\Pi_\Psi: \ \mu_W \in \mathcal{D}(m, \Gamma) \mapsto \left[m \,|\, \Psi(\Gamma)\right]. \tag{10a}$$

Notice the design degree of freedom to use any generalized matrix square root Ψ. As the principal square root $(\cdot)^{\frac{1}{2}}$ in the Gelbrich metric (Def. 1) is a non-convex function, we choose

$$\Psi(\Gamma) = \bar{\Gamma}^{-\frac{1}{2}}(\bar{\Gamma}^{\frac{1}{2}}\Gamma\bar{\Gamma}^{\frac{1}{2}})^{\frac{1}{2}}. \tag{10b}$$

The map $\Gamma \mapsto \Psi(\Gamma)$ is bijective and it satisfies $\Psi(\Gamma)\Psi(\Gamma)^\top = \Gamma$. For $(\bar{\Gamma})^{-\frac{1}{2}}$ to exist, we assume $\bar{\Gamma} \in \mathbb{S}_{++}^{n_w}$. Moreover, consider the PCE coefficient ambiguity set

$$\mathbb{A} = \left\{ \begin{array}{c} \mathsf{W}^{[0,n_w]} \in \\ \mathbb{R}^{n_w \times (n_w+1)} \end{array} \middle| \begin{array}{c} \left\| \mathsf{W}^{[0,n_w]} - \left[\bar{m} \,\middle|\, \bar{\Gamma}^{\frac{1}{2}}\right] \right\| \leq \rho \\ \bar{\Gamma}^{\frac{1}{2}} \mathsf{W}^{[1,n_w]} \succeq 0 \end{array} \right\}. \tag{11}$$

Lemma 1 ($\Pi_\Psi(\mathcal{A}) = \mathbb{A}$): Given the empirical moments $(\bar{m}, \bar{\Gamma})$ with mean $\bar{m} \in \mathbb{R}^{n_w}$ and covariance $\bar{\Gamma} \in \mathbb{S}_{++}^{n_w}$. Consider Π_Ψ from (10), the Gelbrich ambiguity set $\mathcal{A}$ from (2), and the PCE coefficient ambiguity set $\mathbb{A}$ from (11). Then, the element-wise image of $\mathcal{A}$ under Π_Ψ is given by $\mathbb{A}$. □

Proof: First, we show that under the map Π_Ψ, the Gelbrich distance $\mathbb{G}$ in the definition of $\mathcal{A}$ (2) corresponds to the norm expression in $\mathbb{A}$ (11). With Π_Ψ and Ψ as specified in (10), we have $\mathsf{w}^0 = m$, $\mathsf{W}^{[1,n_w]} = \Psi(\Gamma)$ and $d = \mathbb{G}\left((m, \Gamma), (\bar{m}, \bar{\Gamma})\right) = \sqrt{\|\mathsf{w}^0 - \bar{m}\|^2 + \operatorname{tr}(\bar{\Gamma} + \Gamma - 2\bar{\Gamma}^{\frac{1}{2}}\mathsf{W}^{[1,n_w]})}$. Moreover, with $M = \bar{\Gamma}^{\frac{1}{2}} - \mathsf{W}^{[1,n_w]}$ and since $\Gamma = \mathsf{W}^{[1,n_w]}\mathsf{W}^{[1,n_w]\top}$, we have $d = \sqrt{\|\mathsf{w}^0 - \bar{m}\|^2 + \operatorname{tr}(MM^\top)} = \|\mathsf{W}^{[0,n_w]} - [\bar{m} \,|\, \bar{\Gamma}^{\frac{1}{2}}]\|$, where we used the properties of the Frobenius norm.

Next we prove that $\bar{\Gamma}^{\frac{1}{2}}\mathsf{W}^{[1,n_w]} \succeq 0$ in (11) is equivalent to $\Gamma \succeq 0$ in (2) provided $\mathsf{W}^{[1,n_w]} = \Psi(\Gamma)$ as in (10). That is, we aim to show

$$\mathsf{W}^{[1,n_w]} = \Psi(\Gamma), \Gamma \succeq 0 \Leftrightarrow \bar{\Gamma}^{\frac{1}{2}}\mathsf{W}^{[1,n_w]} \succeq 0. \tag{12}$$

The $\Rightarrow$ implication holds, since $\bar{\Gamma}^{\frac{1}{2}}\mathsf{W}^{[1,n_w]} = (\bar{\Gamma}^{\frac{1}{2}}\Gamma\bar{\Gamma}^{\frac{1}{2}})^{\frac{1}{2}} \succeq 0$. $\Leftarrow$: since Ψ is bijective, its inverse map $\Psi^{-1} : \mathbb{R}^{n_w \times n_w} \to \mathbb{S}_+^{n_w}$, $\mathsf{W}^{[1,n_w]} \mapsto \Psi^{-1}(\mathsf{W}^{[1,n_w]})$ exists. Thus, if the right hand side of (12) holds, we find $\Gamma = \Psi^{-1}(\mathsf{W}^{[1,n_w]}) \in \mathbb{S}_+^{n_w}$ and then the left hand side holds. ■

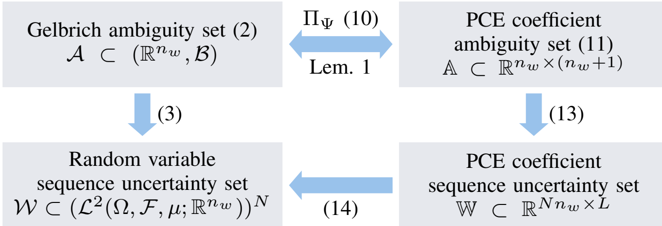

Fig. 1. Relations and maps between the sets $\mathcal{A}$, $\mathbb{A}$, $\mathcal{W}$, and $\mathbb{W}$.

Recall that the Gelbrich distance in Definition 1 is a non-convex function of (m, Γ). However, it is convex in the PCE coefficients $\mathsf{W}^{[0,n_w]}$. Hence the PCE ambiguity set $\mathbb{A}$ from (11) is a compact and convex subset of $\mathbb{R}^{n_w \times (n_w+1)}$. Finally, we arrive at the uncertainty description for the PCE coefficient sequences $W_{[0,N-1]}$

$$\mathbb{W} \doteq \left\{ \begin{array}{c} \mathsf{W}^{[0,L-1]}_{[0,N-1]} \in \\ \mathbb{R}^{Nn_w \times L} \end{array} \middle| \begin{array}{c} \mathsf{W}^{[0,L-1]}_{[0,N-1]} \left(\mathsf{W}^{[0,n_w]}\right) \text{ s.t. (9)} \\ \mathsf{W}^{[0,n_w]} \in \mathbb{A} \end{array} \right\}, \quad (13)$$

and at the PCE reformulation of $\mathcal{W}$ from (3)

$$\mathcal{W} = \left\{ W \middle| \begin{array}{l} W = \sum_{j=0}^{L-1} \mathsf{W}^j \phi^j(\boldsymbol{\xi}),\ \phi \text{ cf. (7)} \\ \mathsf{W}^{[0,L-1]} \in \mathbb{W},\ \boldsymbol{\xi} \in \mathcal{D}(0, I_{Nn_w}) \end{array} \right\}. \quad (14)$$

Figure 1 summarizes the relations and maps between the ambiguity sets $\mathcal{A}, \mathbb{A}$ and the sequence uncertainty descriptions $\mathcal{W}, \mathbb{W}$.

IV. Data-driven distributionally robust optimal control in PCE cofficients

The above reformulation of the ambiguity set $\mathcal{A}$ to $\mathbb{W}$ enables us to the cast the distributionally robust OCP (4) as an *uncertain conic problem*, whereby we will use a data-driven representation in lieu of explicit knowledge of the system matrices.

A. Data-driven representation of stochastic LTI systems

For a specific uncertainty outcome $\omega \in \Omega$ the realization of W_k is written as $w_k \doteq W_k(\omega)$, $k \in \mathbb{N}$. Likewise, the realizations of inputs, outputs, and states are $u_k \doteq U_k(\omega)$, $y_k \doteq Y_k(\omega)$, and $x_k \doteq X_k(\omega)$, respectively. Given $\{w_k\}_{k\in\mathbb{N}}$, the stochastic system (1) induces the realization dynamics

$$x_{k+1} = Ax_k + Bu_k + Ew_k, \quad x_0 = x_{\text{ini}} \quad (15a)$$

$$y_k = Cx_k + Du_k + Fw_k. \quad (15b)$$

Assumption 1 (System properties and data): Consider stochastic LTI system (1) and its realization dynamics (15), we assume that $(A, [B, E])$ is a controllable pair, and respectively, (A, C) is a observable pair. In addition, we suppose the matrices (A, B, C, D, E, F) are unknown, while measurements of past input-output-disturbance realizations u_k, y_k and w_k are available. □

Definition 4 (Persistency of excitation [17]): Let $T, t \in \mathbb{N}^+$. A sequence of inputs $u_{[0,T-1]}$ is said to be persistently exciting of order t if the Hankel matrix $\mathcal{H}_t(u_{[0,T-1]}) \doteq \begin{bmatrix} u_0 & \cdots & u_{T-t} \\ \vdots & \ddots & \vdots \\ u_{t-1} & \cdots & u_{T-1} \end{bmatrix}$ is of full row rank. □

Next we recall crucial insights from [6, Lem. 4, Cor. 2] which allow to represent the PCE coefficients dynamics (8) by previous recorded data of the realization dynamics (15).

Lemma 2 ([6]): Let Assumption 1 hold. Consider system (1) and a T-length realization trajectory tuple $(u, w, y)_{[0,T-1]}$ of its corresponding realization dynamics (15). We suppose that $(u, w)_{[0,T-1]}$ is persistently exciting of order $n_x + t$. Then $(U, W, Y)_{[0,t-1]}$ is a trajectory of (1) if and only if there exists $G \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{T-t+1})$ such that $\mathcal{H}_t(v_{[0,T-1]})G = V_{[0,t-1]}$ holds for all $(v, V) \in \{(u, U), (w, W), (y, Y)\}$.

Moreover, $(\mathsf{u}, \mathsf{w}, \mathsf{y})^j_{[0,t-1]}$, $j \in \mathbb{I}_{[0,L-1]}$ is a trajectory of the dynamics of PCE coefficients (8) if and only if there exists $\mathsf{g}^j \in \mathbb{R}^{T-t+1}$ such that $\mathcal{H}_t(v_{[0,T-1]})\mathsf{g}^j = \mathsf{v}^j_{[0,t-1]}$, $j \in \mathbb{I}_{[0,L-1]}$, holds for all $(v, \mathsf{v}) \in \{(u, \mathsf{u}), (w, \mathsf{w}), (y, \mathsf{y})\}$. □

It is worth be remarked that the structural similarity of the PCE coefficient dynamics (8) with (1) and (15) is at the core of the above lemma. Note that this similarity is jeopardized by co-variance based uncertainty propagation.

B. Distributionally robust data-driven OCP

Combining the above results, we turn to the data-driven reformulation of OCP (4) in terms of PCE coefficients.

Assumption 2 (Data availability): Consider a given T-length realization trajectory tuple $(u, w, y)_{[0,T-1]}$ of the corresponding realization dynamics (15). We suppose that $(u, w)_{[0,T-1]}$ is persistently exciting of order $n_x + N + T_{\text{ini}}$ with T_{ini} not smaller than the system lag of (1), cf. [6]. □

Consider T_{ini} past measurements of $(u, y, w)_{[-T_{\text{ini}},-1]}$ and a T-length realization trajectory tuple $(u, y, w)_{[0,T-1]}$ satisfying Assumption 2. Let p and f denote the ranges $[-T_{\text{ini}}, -1]$ and $[0, N-1]$, respectively. Let $\mathcal{H}_{v,\text{p}}$ and $\mathcal{H}_{v,\text{f}}$ be the first $T_{\text{ini}}n_v$ rows and, respectively, the remaining Nn_v rows of the Hankel matrix $\mathcal{H}_{N+T_{\text{ini}}}(v_{[0,T-1]})$ for $v \in \{u, y, w\}$. Consider the stacked Hankel matrices as $\mathcal{H}_\text{p} \doteq [\mathcal{H}_{u,\text{p}}^\top, \mathcal{H}_{y,\text{p}}^\top, \mathcal{H}_{w,\text{p}}^\top]^\top$ and $\mathcal{H}_\text{f} \doteq [\mathcal{H}_{u,\text{f}}^\top, \mathcal{H}_{y,\text{f}}^\top, \mathcal{H}_{w,\text{f}}^\top]^\top$. The uncertainty set $\mathbb{W}$ for the PCE coefficient sequences (13) gives the finite-dimensional and convex reformulation of OCP (4)

$$\min_{\bar{u}, K, \alpha, \mathsf{u}, \mathsf{y}, \mathsf{g}} \alpha \quad (16a)$$

$$\text{s.t.} \quad \forall \mathsf{W}_\text{f}^{[0,L-1]} \in \mathbb{W},\ \forall k \in \mathbb{I}_{[0,N-1]},$$

$$\textstyle\sum_{k=0}^{N-1} \sum_{j=0}^{L-1} (\|\mathsf{y}_k^j\|_Q^2 + \|\mathsf{u}_k^j\|_R^2) \leq \alpha, \quad (16b)$$

$$\mathcal{H}_\text{p}\mathsf{g}^j = \delta^{0j}[u_\text{p}^\top, y_\text{p}^\top, w_\text{p}^\top]^\top, \forall j \in \mathbb{I}_{[0,L-1]}, \quad (16c)$$

$$\mathcal{H}_\text{f}\mathsf{g}^j = [\mathsf{u}_\text{f}^{j\top}, \mathsf{y}_\text{f}^{j\top}, \mathsf{w}_\text{f}^{j\top}]^\top, \forall j \in \mathbb{I}_{[0,L-1]}, \quad (16d)$$

$$\mathsf{u}_\text{f}^0 = \bar{u} + K_w \mathsf{w}_\text{f}^0,\ \mathsf{u}_\text{f}^j = K_w \mathsf{w}_\text{f}^j, \forall j \in \mathbb{I}_{[1,L-1]}, \quad (16e)$$

$$a_{u,i}^\top \mathsf{u}_k^0 + \sigma(\varepsilon_u)\|a_{u,i}^\top \mathsf{U}_k^{[1,L-1]}\| \leq 1, \forall i \in \mathbb{I}_{[1,N_u]}, \quad (16f)$$

$$a_{y,i}^\top \mathsf{y}_k^0 + \sigma(\varepsilon_y)\|a_{y,i}^\top \mathsf{Y}_k^{[1,L-1]}\| \leq 1, \forall i \in \mathbb{I}_{[1,N_y]}, \quad (16g)$$

where δ^{0j} is the Kronecker delta, K_w collects all feedback gains $K_{k,i}$ similar to (5), and $\sigma(\varepsilon) = \sqrt{(1-\varepsilon)/\varepsilon}$.

Lemma 2 justifies the data-driven representation of the dynamics of PCE coefficients (8) in (16c)-(16d). Note that

δ^{0j} in (16c) specifies the PCE coefficients of the initial condition to be zero for $j > 0$, i.e., we consider a deterministic initial condition. Causality and affiness of polices in (4e) are stated in (16e). The next result gives the exactness of the reformulation of the chance constraints from (4f)-(4g) to (16f)-(16g).

Proposition 1 (PCEs for DRO chance constraints): Consider a random variable $V \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{n_v})$ with its PCE $V(\boldsymbol{\xi}) = \sum_{j=0}^{L-1} \mathsf{v}^j \phi^j(\boldsymbol{\xi})$ regarding the basis (7). For $a \in \mathbb{R}^{n_v}$, the distributionally robust chance constraint

$$\mathbb{P}[a^\top V(\boldsymbol{\xi}) \le 1]] \ge 1 - \varepsilon, \quad \forall \boldsymbol{\xi} \in \mathcal{D}(0, \mathbb{I}_{Nn_w})$$

is equivalent to $a^\top \mathsf{v}^0 + \sqrt{(1-\varepsilon)/\varepsilon}\|a^\top \mathsf{V}^{[1,L-1]}\| \le 1$. □

Proof: Using (7) we have $V(\boldsymbol{\xi}) = \mathsf{v}^0 + \mathsf{V}^{[1,L-1]}\boldsymbol{\xi}$, and thus the DRO chance constraint reads

$$\mathbb{P}\left[a^\top \mathsf{V}^{[1,L-1]}\boldsymbol{\xi} \le 1 - a^\top \mathsf{v}^0\right] \ge 1 - \varepsilon, \forall \boldsymbol{\xi} \in \mathcal{D}(0, \mathbb{I}_{Nn_w}).$$

Since this expression is bilinear in $\boldsymbol{\xi} \in \mathcal{L}^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^{Nn_w})$ and the decision variables $\mathsf{V}^{[0,L-1]} \in \mathbb{R}^{n_v \times L}$, it is equivalent to $a^\top \mathsf{V}^{[1,L-1]}\mathbb{E}[\boldsymbol{\xi}] + \sqrt{(1-\varepsilon)/\varepsilon}(a^\top \mathsf{V}^{[1,L-1]}\Sigma[\boldsymbol{\xi}, \boldsymbol{\xi}]\, \mathsf{V}^{[1,L-1]\top} a)^{1/2} \le 1 - a^\top \mathsf{v}^0$, cf. [18, Th. 3.1]. With $\mathbb{E}[\boldsymbol{\xi}] = 0$ and $\Sigma[\boldsymbol{\xi}, \boldsymbol{\xi}] = I_{Nn_w}$, we conclude the assertion. ■

Theorem 1 (Equivalence of OCP minimizers): Consider OCP (4) with the random-variable uncertainty set $\mathcal{W}$ (3) and OCP (16) with the PCE uncertainty set $\mathbb{W}$ (13). Let Assumptions 1–2 and the conditions of Lemma 1 hold. Then, for any given initial condition $(u, y, w)_{[-T_{\text{ini}}, -1]}$ for OCP (16), there exists $x_{\text{ini}} \in \mathbb{R}^{n_x}$ for OCP (4) such that—provided they are non-empty—the sets of minimizers $(\bar{u}^\star, K^\star, \alpha^\star)$ of OCP (4) and OCP (16) are the same. □

Proof: The proof relies on that the PCE reformulation of all random variables in the basis (7) is exact and the omission of the basis from OCP (4) to OCP (16) is without loss of information. Due to Assumption 1 the system is observable and the measurements of (u, y, w) are exact. Hence $(u, y, w)_{[-T_{\text{ini}}, -1]}$ determines a unique initial state x_{ini} in OCP (4) given T_{ini} is not smaller than the system lag.

Using the basis (7) all random variables in OCP (4) admit exact PCEs with at most $L = Nn_w + 1$ terms cf. [6, Prop. 1]. Replacing all random variables with their PCEs the constraint (4b) is equivalent to (16b) due to the orthonormality of the basis (7). With Assumption 2, (16c)-(16d) exactly captures the PCE coefficient dynamics, cf. Lemma 2. Moreover, (16e) expresses the causal and affine policies (4e) in PCE coefficients. With (14) we split the uncertainty description $\forall W_{\text{f}} \in \mathcal{W}$ into two conditions $\forall \mathsf{W}_{\text{f}}^{[0,L-1]} \in \mathbb{W}$ and $\forall \boldsymbol{\xi} \in \mathcal{D}(0, \mathbb{I}_{Nn_w})$. Using the latter condition and applying Proposition 1 the chance constraints (4f)-(4g) are exactly reformulated to (16f)-(16g). Notice that the reformulated objective constraint and chance constraints are independent of the PCE basis (7). Thus, without loss of information, we drop the basis and finally obtain OCP (16). Since the reformulation from OCP (4) to OCP (16) is exact, the sets of their minimizers restricted to the variables $(\bar{u}^\star, K^\star, \alpha^\star)$ coincide. ■

C. Numerical implementation

Observe that OCP (16) is an uncertain conic problem. Hence tractable reformulations are possible for specific types of uncertainty sets [19]. In our approach, we approximate the uncertainty set $\mathbb{A}$ in (11) by a polytope and than define the approximation of $\mathbb{W}$ accordingly. To this end, we uniformly sample $s \in \mathbb{N}^+$ points $\delta^j \in \mathbb{R}^{n_w \times (n_w+1)}$ from $\mathbb{A}$ for $j \in \mathbb{I}_{[1,s]}$. [1] We approximate $\mathbb{A}$ by the convex hull of the sample points, denoted as $\widetilde{\mathbb{A}} = \text{Conv}(\delta^1, ..., \delta^s) \subset \mathbb{A}$. By linearly lifting each vertex of $\widetilde{\mathbb{A}}$ via (9), we obtain $\widetilde{\mathbb{W}}$ similarly as in (13).

We denote the vertices of $\widetilde{\mathbb{W}}$ by $\Delta \doteq \left\{\widetilde{\delta}^j, j \in \{1, \ldots, \tilde{s}\}\right\}$ which are a subset of the lifted sample points with $\tilde{s} \le s$. Replacing $\mathbb{W}$ with the countable set Δ, we obtain

$$\min_{\bar{u}, K, \alpha, \mathsf{u}, \mathsf{y}, \mathsf{g}} \alpha \quad \text{s.t. } \forall \mathsf{W}_{\text{f}}^{[0,L-1]} \in \Delta, \text{(16b)} - \text{(16g)}. \qquad (17)$$

Observe that with (16f)–(16g), (17) is a second-order cone program whose computational complexity is $O(\sqrt{N(N_u + N_y)\tilde{s}})$, cf. [20]. Due to the tight page limit, a detailed analysis of the sample efficiency of the proposed approximation strategy is postponed to future work. Instead, we demonstrate its efficacy numerically.

V. Numerical Example

We consider the discrete-time stochastic double integrator

$$X_{k+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X_k + \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} U_k + W_k, \; Y_k \; = \begin{bmatrix} 1 & 0 \end{bmatrix} X_k,$$

where the W_k are i.i.d. with Gaussian mixture distributions. Especially, μ_W is the mixture of $\mathcal{N}(\left[\begin{smallmatrix} 0.1 \\ 0.1 \end{smallmatrix}\right], 0.01 I_2)$ and $\mathcal{N}(\left[\begin{smallmatrix} -0.1 \\ -0.1 \end{smallmatrix}\right], 0.01 I_2)$ with mean $m_{\text{true}} = [0, 0]^\top$ and covariance $\Gamma_{\text{true}} = \left[\begin{smallmatrix} 0.03 & 0.02 \\ 0.02 & 0.03 \end{smallmatrix}\right]$. Notice that these true values of mean and covariance are unknown to the OCP. We specify $T_{\text{ini}} = 2$ which corresponds to the system lag. The weighting matrices are $Q = R = 1$ for Y and U, and the prediction horizon is $N = 10$. Chance constraints on the input require $U_k \le 0.5$ and $U_k \ge -0.5$ to be satisfied individually with probability of no less than 80% for $k = 0, ..., 9$.

To construct OCP (17) based on measured data, we first apply 70 random inputs to the system and record the output responses as well as the realized disturbances. Then we use this data to construct Hankel matrices and to estimate the moments of W as $[\bar{m} \,|\, \bar{\Gamma}] = \left[\begin{smallmatrix} 0.0025 & | & 0.0211 & 0.0100 \\ 0.0025 & | & 0.0100 & 0.0157 \end{smallmatrix}\right]$. Using $[\bar{m} \,|\, \bar{\Gamma}]$ as the empirical moment pair and setting the radius $\rho = \bar{\rho} \cdot \|[\bar{m} \,|\, \bar{\Gamma}]\|$ for a user-chosen $\bar{\rho} \in \mathbb{R}^+$, we obtain Gelbrich ambiguity sets $\mathcal{A}$ (2) and the corresponding PCE uncertainty sets $\mathbb{A}$ (11). To construct Δ, we uniformly sample s points from $\mathbb{A}$. Subsequently, we investigate the effect of varying radius ρ and the number of samples s.

We consider three cases of OCP (17) :

(I) The robust case, where OCP (17) is solved with Δ for different values of $\bar{\rho}$ and s.

[1] An intuitive strategy is to sample uniformly over hypercubes which contain $\mathbb{A}$ and to neglect any samples which are not in $\mathbb{A}$.

TABLE I

COMPARISON OF THE AVERAGED COST J AND THE NUMBER OF CONSTRAINTS VIOLATION $\#_V$ FOR 1000 REALIZATION TRAJECTORIES.

case I	$\bar{\rho} = 0.1$		$\bar{\rho} = 0.3$		$\bar{\rho} = 0.5$		$\bar{\rho} = 0.7$	
s	J	$\#_V$	J	$\#_V$	J	$\#_V$	J	$\#_V$
10	24.54	138	24.79	68	25.24	27	25.56	11
50	24.58	124	25.15	53	25.79	24	26.65	6
100	24.58	124	25.15	53	25.79	24	26.65	6

	J	$\#_V$		J	$\#_V$
case II	24.47	184	case III	25.04	26

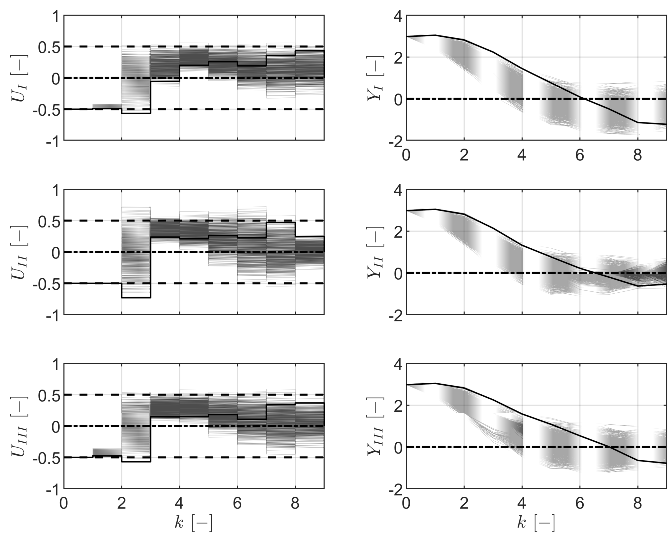

Fig. 2. Input and output response for 1000 disturbance sequences of case I with $\bar{\rho} = 0.5$, $s = 50$ (top), case II (middle), and case III (bottom). The most constraint-violating realization is highlighted.

(II) The optimistic case, where OCP (17) is solved with $\Delta = \{[\bar{m} \,|\, \Psi(\bar{\Gamma})]\}$, using the empirical moments estimated from the 70 recorded disturbance samples.

(III) The ideal case, i.e., OCP (17) with $\Delta = \{[m_{\text{true}} \,|\, \Psi(\Gamma_{\text{true}})]\}$, utilizing the true moments.

Each OCP is solved using the same initial data u_{p}, y_{p}, and w_{p}. Note that with ambiguity sets of fixed moments, cases II and III are instances of the approach in [9].

Using 1000 different sampled disturbance realization sequence of length 10 each, Table I compares the averaged cost J and the number of constraint violations $\#_V$ for case I with different values of $\bar{\rho}$ and s with cases II & III. We see that increasing $\bar{\rho}$ and s leads to fewer constraint violations and decreased performance. Comparing case I with cases II & III, it is evident that the former provides a more robust solution. Figure 2 shows the corresponding input and output responses of case I with $\bar{\rho} = 0.5$ and $s = 50$ as well as cases II & III. Observe that the input responses of case I violate the constraints much less frequently compared to case II (with moments estimated from data) and still achieve similar output responses as case III (with the true moments).

VI. CONCLUSION AND OUTLOOK

This paper discussed distributionally robust uncertainty propagation for LTI systems via data-driven stochastic optimal control. We leveraged polynomial chaos expansions to derive an exact reformulation of model-based distributionally robust OCPs with Gelbrich ambiguity sets to data-driven uncertain conic problems with a finite-dimensional convex uncertainty set in PCE coefficients. A tractable approximation to convex programs has been proposed and illustrated via an example. Future work will consider tailored sampling strategies for the PCE coefficient ambiguity set, exact reformulations for robust second-order cone constraints [19], and the effect of the size of the previously recorded data.

REFERENCES

[1] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.

[2] M. Fochesato and J. Lygeros, "Data-driven distributionally robust bounds for stochastic model predictive control," in *Proc. IEEE 61th Conf. on Decis. and Control*. IEEE, 2022, pp. 3611–3616.

[3] S. Lu, J. H. Lee, and F. You, "Soft-constrained model predictive control based on data-driven distributionally robust optimization," *AIChE Journal*, vol. 66, no. 10, p. e16546, 2020.

[4] P. Coppens and P. Patrinos, "Data-driven distributionally robust MPC for constrained stochastic systems," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 1274–1279, 2021.

[5] J. Coulson, J. Lygeros, and F. Dörfler, "Distributionally robust chance constrained data-enabled predictive control," *IEEE Trans. Automat. Contr.*, vol. 67, no. 7, pp. 3289–3304, 2022.

[6] G. Pan, R. Ou, and T. Faulwasser, "On a stochastic fundamental lemma and its use for data-driven optimal control," *IEEE Trans. Automat. Contr.*, pp. 1–16, 2022.

[7] T. Faulwasser, R. Ou, G. Pan, P. Schmitz, and K. Worthmann, "Behavioral theory for stochastic systems? A data-driven journey from Willems to Wiener and back again," *Annu. Rev. Control*, vol. 55, pp. 92–117, 2023.

[8] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Trans. Automat. Contr.*, vol. 61, no. 2, pp. 430–442, 2015.

[9] B. Li, Y. Tan, A.-G. Wu, and G.-R. Duan, "A distributionally robust optimization based method for stochastic model predictive control," *IEEE Trans. Automat. Contr.*, vol. 67, no. 11, pp. 5762–5776, 2021.

[10] T. J. Sullivan, *Introduction to Uncertainty Quantification*. Springer, 2015, vol. 63.

[11] C. R. Givens and R. M. Shortt, "A class of Wasserstein metrics for probability distributions." *Michigan Math. J.*, vol. 31, no. 2, pp. 231–240, 1984.

[12] V. A. Nguyen, S. Shafieezadeh-Abadeh, D. Kuhn, and P. Mohajerin Esfahani, "Bridging Bayesian and minimax mean square error estimation via Wasserstein distributionally robust optimization," *Math. Oper. Res.*, vol. 48, no. 1, pp. 1–37, 2021.

[13] Y. Lian and C. N. Jones, "From system level synthesis to robust closed-loop data-enabled predictive control," in *Proc. IEEE 60th Conf. on Decis. and Control*. IEEE, 2021, pp. 1478–1483.

[14] T. Mühlpfordt, R. Findeisen, V. Hagenmeyer, and T. Faulwasser, "Comments on quantifying truncation errors for polynomial chaos expansions," *IEEE Contr. Syst. Lett.*, vol. 2, no. 1, pp. 169–174, 2018.

[15] T. Lefebvre, "On moment estimation from polynomial chaos expansion models," *IEEE Contr. Syst. Lett.*, vol. 5, no. 5, pp. 1519–1524, 2020.

[16] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, revised ed. Springer New York, 2003.

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

[18] G. C. Calafiore and L. E. Ghaoui, "On distributionally robust chance-constrained linear programs," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1–22, 2006.

[19] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton Univ. Press, 2009.

[20] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1-3, pp. 193–228, 1998.