2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Distributionally Robust Density Control with Wasserstein Ambiguity Sets

Joshua Pilipovsky and Panagiotis Tsiotras

Abstract— **Precise control under uncertainty requires a good understanding and characterization of the noise affecting the system. This paper studies the problem of steering state distributions of dynamical systems subject to partially known uncertainties. We model the distributional uncertainty of the noise process in terms of Wasserstein ambiguity sets, which, based on recent results, have been shown to be an effective means of capturing and propagating uncertainty through stochastic LTI systems. To this end, we propagate the distributional uncertainty of the state through the dynamical system, and, using an affine feedback control law, we steer the ambiguity set of the state to a prescribed, terminal ambiguity set. We also enforce distributionally robust CVaR constraints for the transient motion of the state so as to reside within a prescribed constraint space. The resulting optimization problem is formulated as a semi-definite program, which can be solved efficiently using standard off-the-shelf solvers. We illustrate the proposed distributionally-robust framework on a path planning problem and compare with non-robust solutions.**

I. INTRODUCTION

When controlling a dynamical system affected by noise, one needs to be able to discern the statistical properties of the exogenous disturbances acting on the system. When such a characterization is unknown, or is only approximately known, care must be taken to ensure robust performance of the system under a range of uncertainties that can potentially affect the system. Indeed, if, for example, a control designer naively assumes a normally distributed noise process, the resulting control law may severely underestimate the probability of violating the constraints or it may fail to reach a given desired terminal state [1].

To this end, we would like to systematically and tractably solve a stochastic optimal control problem that can not only control the dispersion of system states to a prescribed terminal distribution, but also steer the uncertainty of this dispersion for all disturbances sufficiently close to the true disturbance acting on the system. The theory of covariance control, originally introduced in the 80's with works of Hotz and Skelton [2], solved the problem of steering the first two moments of the state distribution in the infinite horizon setting. In recent years, this theory has been extended to the finite-horizon setting [3], [4], as well as extensions involving chance constraints [5], [6], partially observed systems [7], and data-driven scenarios [8] under the term covariance steering (CS) to emphasize the finite-horizon problem formulation. The baseline theory is mathematically tractable and elegant, however, it assumes Gaussian noise entering the system, as well as boundary Gaussian distributions for the initial and terminal states.

Additionally, the framework assumes exact knowledge of the noise model affecting the system, which is unrealistic in practice. We are rarely fully aware of the disturbances acting on the system, and at best we can characterize partial statistical information from collected data, e.g., the first two moments. As such, it is fruitful to consider the problem of steering the distribution of the state under distributional uncertainty in the noise model. A natural framework to accomplish this goal is to the model the noise as residing in an *ambiguity set*, which is characterized by a whole family of distributions that the noise can follow. The goal, then, is to optimize the control law and satisfy constraints under the *worst-case* disturbance that nature imposes within the allowable ambiguity set. The work in [1] has solved this problem by characterizing the distributional noise uncertainty as a Chebyshev ambiguity set, which is a family of distributions that have common first two moments, and by tractably enforcing chance constraints using concentration inequalities. This ambiguity set, however, is still quite limited in its expressivity due to the assumption of common moments.

Recently, there has been a great promise in capturing distributional uncertainty via Wasserstein ambiguity sets, which are defined through the natural Wasserstein metric on probability spaces. Indeed, [9], [10], [11] has shown that distributionally robust optimization (DRO) over Wasserstein ambiguity sets is tractable in cases where the nominal distribution is either empirical or elliptical. In the context of stochastic control, the work in [12] has outlined a general framework for capturing distributional noise uncertainty through empirical data collected, and have provided a procedure to propagate Wasserstein ambiguity sets through stochastic LTI systems, which is analytically exact under some mild assumptions. Subsequent works have applied this framework to design optimal open-loop controllers while satisfying conditional value-at-risk (CVaR) constraints, as well as in the context of model-predictive control [13].

Our contributions are as follows. To the best of our knowledge, this is the first work that solves the open problem [12] of optimizing over both open-loop and *feedback* controllers for distributionally-robust optimal control problems, which is accomplished using an affine state feedback control law using established techniques from the CS literature. Secondly, we show that it is possible to *steer* the distributional

J. Pilipovsky is a PhD student at the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA30332-0150, USA. Email: jpilipovsky3@gatech.edu

P. Tsiotras is the David & Lewis Chair and Professor at the School of Aerospace Engineering and the Institute for Robotics & Intelligent Machines, Georgia Institute of Technology, Atlanta, GA30332-0150, USA. Email: tsiotras@gatech.edu

979-8-3503-1632-2/24/$31.00 ©2024 IEEE

state uncertainty to a desired terminal ambiguity set, thereby controlling the dispersion of system states under all possible noise realizations within the Wasserstein ambiguity set.

II. NOTATION

We assume a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for all random objects. Real valued-vectors are denoted by lowercase letters, $u \in \mathbb{R}^m$, matrices are denoted by uppercase letters, $V \in \mathbb{R}^{n\times M}$, and random vectors are denoted by boldface, $\boldsymbol{x} \in \mathbb{R}^n$. The space of probability distributions over $\mathbb{R}^d$ with finite qth moment is denoted by $\mathcal{P}_q(\mathbb{R}^d)$. Given $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_q(\mathbb{R}^d)$, we denote by $\mathbb{P}\otimes\mathbb{Q}$ their product distribution and by $\mathbb{P}^{\otimes N}$ the N-fold product distribution $\mathbb{P}\otimes\cdots\otimes\mathbb{P}$ with N terms. Given a matrix $A \in \mathbb{R}^{m\times d}$, the pushforward of $\mathbb{P}$ is given by $A_{\#}\mathbb{P}$ and is defined by $(A_{\#}\mathbb{P}(\mathcal{B})) = \mathbb{P}(A^{-1}(\mathcal{B}))$, for all Borel sets $\mathcal{B} \subset \mathbb{R}^m$. We denote by δ_x the Dirac delta distribution concentrating unit mass at the atom $x \in \mathbb{R}^n$. The convolution of $\mathbb{P}$ and δ_x is denoted by $\delta_x * \mathbb{P}$, and is defined by $(\delta_x * \mathbb{P})(\mathcal{B}) = \mathbb{P}(\mathcal{B} - x)$. With slight abuse of notation, the operator $\|\cdot\|$ denotes the Euclidean norm for vectors and the spectral norm for matrices. The Moore-Penrose pseudoinverse of a matrix A is denoted by $A^{\dagger}$. Lastly, for any $t \in \mathbb{Z}_+$, we set $[t] = \{0, \ldots, t\}$.

III. PROBLEM STATEMENT

Consider the discrete-time, stochastic linear dynamics system

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + B\boldsymbol{u}_k + D\boldsymbol{w}_k, \quad \forall k \in [N-1], \tag{1}$$

with states $\boldsymbol{x}_k \in \mathbb{R}^n$, control inputs $\boldsymbol{u}_k \in \mathbb{R}^m$, and process noise sequence $\{\boldsymbol{w}_k\}_{k\in[N-1]} \subset \mathbb{R}^d$ that is neither identically nor necessarily independently distributed. In this work, we assume the system model $\{A, B, D\}$ is known. The noise process $\{\boldsymbol{w}_k\}_{k\in[N-1]}$, on the other hand, is unknown but belongs to an ambiguity set $\mathcal{W}$, which is defined rigorously in the following definition.

Definition 1: The Wasserstein ambiguity set of radius ε with transportation cost c centered at the nominal distribution $\mathbb{P}$ is defined by

$$\mathbb{B}^c_{\varepsilon,p}(\mathbb{P}) = \{\mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^d) : \mathbb{W}^c_p(\mathbb{Q}, \mathbb{P}) \le \varepsilon\}, \tag{2}$$

with respect to the type-p Wasserstein metric

$$\mathbb{W}^c_p(\mathbb{P}, \mathbb{P}') \triangleq \left(\inf_{\pi\in\Pi(\mathbb{P},\mathbb{P}')} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(\xi, \xi')^p \; \pi(\mathrm{d}\xi, \mathrm{d}\xi') \right)^{\frac{1}{p}}, \tag{3}$$

where $\Pi(\mathbb{P}, \mathbb{P}')$ denotes the set of all joint probability distributions of $\xi \in \mathbb{R}^d$ and $\xi' \in \mathbb{R}^d$ with marginals $\mathbb{P}$ and $\mathbb{P}'$, respectively.

In what follows, we will work with the type-2 Wasserstein distance ($p = 2$) and the Euclidean norm transportation cost, i.e., $c = \|\cdot\|$. For simplicity, we denote $\mathbb{B}^{\|\cdot\|}_{\varepsilon} \triangleq \mathbb{B}^{\|\cdot\|}_{\varepsilon,2}$ and $\mathbb{W} \triangleq \mathbb{W}^{\|\cdot\|}_2$. A customary way to construct the noise ambiguity set is by defining the nominal distribution as $\hat{\mathbb{P}}_w = \frac{1}{T}\sum_{i=1}^{T} \delta_{\hat{w}^{(i)}}$, where $\hat{w}^{(i)} \triangleq [(\hat{w}_0^{(i)})^\intercal, \ldots, (\hat{w}_{N-1}^{(i)})^\intercal]^\intercal \in \mathbb{R}^{Nd}$ is a noise realization sampled from the underlying *true* distribution $\mathbb{P}_w$. It can be shown that by choosing a suitable radius $\varepsilon(T, \beta)$, the true distribution lies in the ball $\mathbb{B}^{\|\cdot\|}_{\varepsilon}(\hat{\mathbb{P}}_w)$ with probability $1-\beta$ [14]. In the present work, however, we assume that the noise sequence ambiguity set $\mathcal{W}$ is centered on a zero-mean normal distribution $\hat{\mathbb{P}}_w = \mathcal{N}(0, \Sigma_w)$ with noise covariance matrix $\Sigma_w \in \mathbb{R}^{Nn\times Nn} \succ 0$ and radius $\varepsilon > 0$.

Remark 1: It is possible, and in fact customary, to define the noise ambiguity set for an individual disturbance w_k via $\mathbb{B}^{\|\cdot\|}_{\varepsilon}(\hat{\mathbb{P}}_w)$. Assuming the noise is i.i.d., then it can be shown that the ambiguity set for the entire noise *sequence* is $\mathbb{B}^{\|\cdot\|}_{N\varepsilon}(\hat{\mathbb{P}}_w^{\otimes N})$. However, in this work we choose to define the ambiguity set directly in terms of the disturbance sequence to include cases where the noise terms are not independent of one another, thus prohibiting us from writing the joint distribution of $\boldsymbol{w}$ as an N-fold product distribution.

We assume that the initial state $x_0 = x_i$ is deterministic, which implies that all the distributional uncertainty in the state results from the noise ambiguity set $\mathcal{W} = \mathbb{B}^{\|\cdot\|}_{\varepsilon}(\hat{\mathbb{P}})$. We define the set π of *admissible* control inputs as the set of control sequences $\{\boldsymbol{u}_k\}_{k\in[N]}$ where the input $\boldsymbol{u}_k$ is an affine function of the state. Further, we define the *nominal* state as the deterministic part of the state governed by the nominal dynamics

$$\bar{x}_{k+1} = Ax_k + B\bar{u}_k, \tag{4}$$

where $\bar{u}_k \in \mathbb{R}^m$ is the nominal control, and we define the *error* state $\tilde{\boldsymbol{x}}_k \triangleq \boldsymbol{x}_k - \bar{x}_k$, which obeys the dynamics

$$\tilde{\boldsymbol{x}}_{k+1} = A\tilde{\boldsymbol{x}}_k + B\tilde{\boldsymbol{u}}_k + D\boldsymbol{w}_k, \tag{5}$$

where $\tilde{\boldsymbol{u}}_k$ is the error control.

Remark 2: The nominal state as defined in this work can no longer be associated with the *mean* state $\mathbb{E}_{\mathbb{P}_k}[\boldsymbol{x}_k]$, as is customarily done in the CS literature [6]. In fact, the expectation of the state cannot even be computed because the underlying state distribution is ambiguous by definition.

The goal is to steer to a terminal ambiguity set $\mathbb{S}_f \triangleq \mathbb{B}^{\|\cdot\|}_{\delta}(\mathbb{P}_f)$, where $\delta > 0$ is a given, desired terminal radius, and $\mathbb{P}_f = \mathcal{N}(\mu_f, \Sigma_f)$ is the desired terminal center distribution, while minimizing the distributionally-robust objective function

$$\mathcal{J} = \beta \sum_{k=0}^{N-1} \|\bar{u}_k\| + \max_{\mathbb{P}\in\mathcal{W}} \mathbb{E}_{\mathbb{P}} \left[\sum_{k=0}^{N-1} \tilde{\boldsymbol{x}}_k^\intercal Q_k \tilde{\boldsymbol{x}}_k + \tilde{\boldsymbol{u}}_k^\intercal R_k \tilde{\boldsymbol{u}}_k \right], \tag{6}$$

where $Q_k \succeq 0$ and $R_k \succ 0$ represent the state and input cost weights, respectively, and $\beta > 0$ denotes the weight of the nominal control. Lastly, we would also like to enforce distributionally-robust constraints on the trajectory of the state along the planning horizon.

Letting the state constraint space be the polyehdron $\mathcal{X} \triangleq \{x : \max_{j\in[J]} \alpha_j^\intercal x + \beta_j \le 0\}$, the traditional way of enforcing probabilistic constraints is to enforce chance constraints, which limit the probability of violating the constraints to be smaller than some prescribed risk γ [15]. It is well-known, however, that Value-at-Risk (VaR) constraints are *not* convex, and are only exactly tractable when the underlying state distribution is normal; otherwise, they are approximated using concentration inequalities [16]. In this work, we choose the CVaR risk measure, which is defined as follows.

Definition 2: Given $f : \mathbb{R}^n \to \mathbb{R}$ and a random variable $\boldsymbol{x} \sim \mathbb{P}$ on $\mathbb{R}^n$, the CVaR of $f(\boldsymbol{x})$ at the quantile $1-\gamma$ is

$$\mathrm{CVaR}^{\mathbb{P}}_{1-\gamma}(f(x)) = \inf_{\tau \in \mathbb{R}} \left(\tau + \frac{1}{\gamma} \mathbb{E}_{\mathbb{P}}[\max\{0, f(x) - \tau\}] \right). \tag{7}$$

The CVaR of a random variable is by definition convex [16], implicitly satisfies the VaR constraint, and mitigates the effects of extreme "black swan" events by reducing the tail probability of violating the constraints. To this end, we enforce the distributionally-robust CVaR (DR-CVaR) constraints

$$\sup_{\mathbb{P}_k \in \mathbb{S}_k} \mathrm{CVaR}^{\mathbb{P}_k}_{1-\gamma} \left(\max_{j \in [J]} \alpha_j^\intercal x_k + \beta_j \right) \leq 0, \quad \forall k \in [N], \tag{8}$$

where $\mathbb{S}_k$ denotes the ambiguity set of the state at time step k. In summary, the distributionally-robust density steering (DR-DS) problem is defined as follows.

Problem 1: For a given initial state x_0, find an admissible control sequence $\{\boldsymbol{u}_k\}_{k \in [N]} \in \pi$ that minimizes the DR cost functional (6) subject to the dynamics (1), noise ambiguity set $\mathcal{W}$ and DR-CVaR constraints (8), such that the terminal distributional uncertainty in the state satisfies $\mathbb{S}_N \subseteq \mathbb{S}_f$.

IV. PROBLEM REFORMULATION

We begin by first reformulating the dynamics (1) into a more amenable form for analysis. To this end, define the augmented state, control, and disturbance vectors $\boldsymbol{x} \triangleq [\boldsymbol{x}_0^\intercal, \ldots, \boldsymbol{x}_N^\intercal]^\intercal \in \mathbb{R}^{(N+1)n}, \boldsymbol{u} \triangleq [\boldsymbol{u}_0^\intercal, \ldots, \boldsymbol{u}_{N-1}^\intercal]^\intercal \in \mathbb{R}^{Nm}, \boldsymbol{w} \triangleq [\boldsymbol{w}_0^\intercal, \ldots, \boldsymbol{w}_{N-1}^\intercal]^\intercal \in \mathbb{R}^{Nd}$, respectively, which obey the augmented linear system

$$\boldsymbol{x} = \mathcal{A} x_0 + \mathcal{B} \boldsymbol{u} + \mathcal{D} \boldsymbol{w}, \tag{9}$$

for appropriate matrices $\mathcal{A}, \mathcal{B}, \mathcal{D}$ [4]. We consider the affine state feedback control law $\boldsymbol{u}_k = K_k \tilde{\boldsymbol{x}}_k + v_k$, where $v_k \in \mathbb{R}^m$ is the feed-forward control and $K_k \in \mathbb{R}^{m \times n}$ is the feedback gain. Defining the augmented feed-forward control $v \triangleq [v_0^\intercal, \ldots, v_{N-1}^\intercal]^\intercal \in \mathbb{R}^{Nm}$ and augmented feedback gain matrix $K \in \mathbb{R}^{Nm \times (N+1)n}$, and using the state decomposition in (4)-(5), the dynamics (9) become

$$\begin{aligned} \bar{x} &= \mathcal{A} x_0 + \mathcal{B} v, \\ \tilde{\boldsymbol{x}} &= (I - \mathcal{B}K)^{-1} \mathcal{D} \boldsymbol{w}. \end{aligned} \tag{10}$$

Since K is block lower-triangular and $\mathcal{B}$ is strictly block lower-triangular, it follows that that the matrix $I - \mathcal{B}K$ is invertible. Furthermore, following [17], we define the new decision variable $L \triangleq K(I - \mathcal{B}K)^{-1}$, from which it can be shown that $I + \mathcal{B}L = (I - \mathcal{B}K)^{-1}$, and the original gains can be recovered from $K = L(I + \mathcal{B}L)^{-1}$ following the same logic. As a result, the error state dynamics become

$$\tilde{\boldsymbol{x}} = (I + \mathcal{B}L) \mathcal{D} \boldsymbol{w}. \tag{11}$$

Given $\boldsymbol{w} \in \mathcal{W}$, it follows that the distributional uncertainty in the error state results from the linear transformation $\tilde{\mathbb{S}}_k = (\tilde{L}_k)_{\#} \mathbb{B}^{\|\cdot\|}_{\varepsilon}(\hat{\mathbb{P}})$, with $\tilde{L}_k \triangleq E_k(I + \mathcal{B}L)\mathcal{D}$, where $E_k \in \mathbb{R}^{n \times (N+1)n}$ is a matrix that isolates the kth state element from $\boldsymbol{x}$. To this end, we now state a result on the propagation of ambiguity sets via linear transformations [12].

Theorem 1: Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d)$, and consider the linear transformation defined by the matrix $A \in \mathbb{R}^{m \times d}$. Moreover, let $c : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be orthomonotone[1]. Then,

$$A_{\#} \mathbb{B}^c_{\varepsilon}(\mathbb{P}) \subseteq \mathbb{B}^{c \circ A^\dagger}_{\varepsilon}(A_{\#}\mathbb{P}). \tag{12}$$

Moreover, if the matrix A is full row-rank, then

$$A_{\#} \mathbb{B}^c_{\varepsilon}(\mathbb{P}) = \mathbb{B}^{c \circ A^\dagger}_{\varepsilon}(A_{\#}\mathbb{P}), \tag{13}$$

with $A^\dagger = A^\intercal (AA^\intercal)^{-1}$.

Since $\tilde{L}_k \in \mathbb{R}^{n \times Nd}$, where $Nd \gg n$, in most cases of interest, it is safe to assume that $\tilde{L}_k$ is full row-rank without loss of generality. Noting that the nominal state is simply a delta distribution in the probability space, the distributional uncertainty in the state at time step k becomes

$$\mathbb{S}_k = \delta_{\mathcal{A}x_0 + \mathcal{B}v} * \mathbb{B}^{\|\cdot\| \circ \tilde{L}_k^\dagger}_{\varepsilon} \big((\tilde{L}_k)_{\#} \hat{\mathbb{P}}_w \big), \tag{14}$$

defined on the support $\mathbb{R}^n$.

Remark 3: The interpretation of (14) is that the feedback gain L affects both the *shape* of the center distribution as well as the *size* of the ambiguity set, while the open-loop term v controls the *position* of the center distribution in $\mathcal{P}(\mathbb{R}^n)$. This is a direct generalization of the traditional CS literature, where the open-loop controls the mean state, while the feedback controls the covariance of the state.

In the next section, we tractably formulate the DR-CVaR constraints (8) using the exact ambiguity set (14) and techniques from DRO. For the sake of space, we defer the reader to [18] for detailed proofs of all results.

A. DR-CVaR Constraints

To make the CVaR constraints (8) tractable, we should use our knowledge of the ambiguity set $\mathbb{S}_k$, defined in terms of its Gaussian reference distribution $\hat{\mathbb{P}}_k = \mathcal{N}(0, \tilde{L}_k \Sigma_w \tilde{L}_k^\intercal)$, and transportation cost according to (14). The work in [9] tractably computes the DR-CVaR of piece-wise linear functions, while the work in [19] tractably computes the DR cost of expectations of general piece-wise quadratic functions. In both cases, however, it can be shown that the resulting convex programs are nonlinear in the feedback gain L, which makes them intractable from a computational standpoint. We thus leave it as an open problem to tractably formulate *joint* DR-CVaR constraints for a polyhedral constraint space with a nominal Gaussian distribution whose covariance is parameterized by the feedback gain decision variables.

Instead, we consider an alternative where we wish to enforce the DR-CVaR constraints for each *side* of the polytope along the planning horizon, that is,

$$\sup_{\mathbb{P}_k \in \mathbb{S}_k} \mathrm{CVaR}^{\mathbb{P}_k}_{1-\gamma_{jk}} (\alpha_j^\intercal x_k + \beta_j) \leq 0, \ \forall j \in [J], \ \forall k \in [N]. \tag{15}$$

In essence, at each time step, we split up the joint risk γ to individual risks γ_{jk} of violating the DR-CVaR constraints

[1] That is, $c(x_1 + x_2) \geq c(x_1)$ for all $x_1, x_2 \in \mathbb{R}^d$ satisfying $x_1^\intercal x_2 = 0$.

along each half space and for each time step. This now becomes the DR-CVaR of a *linear* function, which we will show is SDP representable and linear in the decision variables (v, L). First, however, we need to define the notion of a *Gelbrich* ambiguity set.

Definition 3: The Gelbrich ambiguity set of radius ε centered at a mean-covariance pair (μ, Σ) is given by

$$\mathcal{G}_\varepsilon(\mu,\Sigma) = \{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^d) : (\mathbb{E}_\mathbb{Q}[\xi], \mathrm{Cov}_\mathbb{Q}[\xi]) \in \mathcal{U}_\varepsilon(\mu,\Sigma)\}, \tag{16}$$

where $\mathcal{U}_\varepsilon(\mu,\Sigma)$ is an uncertainty set in the space of mean vectors and covariance matrices, defined as

$$\mathcal{U}_\varepsilon(\hat{\mu},\hat{\Sigma}) = \{(\mu,\Sigma) \in \mathbb{R}^d \times \mathbb{S}^d_+ : \mathbb{G}((\mu,\Sigma),(\hat{\mu},\hat{\Sigma})) \le \varepsilon\}, \tag{17}$$

where

$$\mathbb{G}((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)) \triangleq \|\hat{\mu}-\mu\|^2 + \mathrm{tr}\left[\hat{\Sigma} + \Sigma - 2\left(\hat{\Sigma}^{\frac{1}{2}}\Sigma\hat{\Sigma}^{\frac{1}{2}}\right)^{\frac{1}{2}}\right], \tag{18}$$

is the Gelbrich distance between two mean-covariance pairs.

Theorem 2 ([9]): If the nominal distribution $\hat{\mathbb{P}}$ has mean $\hat{\mu} \in \mathbb{R}^d$ and covariance matrix $\hat{\Sigma} \succeq 0$, then we have $\mathbb{B}^{\|\cdot\|}_\varepsilon(\hat{\mathbb{P}}) \subseteq \mathcal{G}_\varepsilon(\hat{\mu},\hat{\Sigma})$.

Since the Gelbrich ambiguity set constitutes an *outer* approximation of the associated Wasserstein ambiguity set (under the 2-norm transportation cost), satisfaction of Gelbrich DR-CVaR constraints implies satisfaction of Wasserstein DR-CVaR constraints. Using this idea, the next result provides a reformulation of the constraints (15).

Theorem 3: The individual DR-CVaR constraints (15) are satisfied if the following convex constraints are satisfied.

$$\beta_j + \alpha_j^\intercal \hat{\mu}_k(v) + \tau_{jk}\sqrt{\alpha_j^\intercal \hat{\Sigma}_k(L)\alpha_j} + \tilde{\varepsilon}_k(L)\|\alpha_j\| \le 0, \quad \forall j \in [J],\ \forall k \in [N]. \tag{19}$$

where $\hat{\mu}_k \triangleq \bar{x}_k(v) = E_k(\mathcal{A}x_0 + \mathcal{B}v)$ is the propagated mean of the nominal distribution, $\hat{\Sigma}_k \triangleq \tilde{L}_k\Sigma_w\tilde{L}_k^\intercal$ is the propagated covariance of the nominal distribution, $\tilde{\varepsilon}_k \triangleq \varepsilon(1+\tau_{jk}^2)^{1/2}\sigma_{\max}^2(\tilde{L}_k)$, and

$$\tau \triangleq \sup_{\mathbb{P}\in\mathcal{C}(\mu,\Sigma)} \mathrm{CVaR}^{\mathbb{P}}_{1-\gamma}\left(\frac{\alpha^\intercal(x-\mu)}{\sqrt{\alpha^\intercal\Sigma\alpha}}\right) = \sqrt{\frac{1-\gamma}{\gamma}}, \tag{20}$$

is the *standard CVaR risk coefficient*, where $\mathcal{C}(\mu,\Sigma)$ denotes the Chebyshev ambiguity set of all distributions in $\mathcal{S}$ with same mean μ and covariance Σ.

The constraints in (19) are convex, but nonlinear in the decision variable L. However, using Schur complement we can further reformulate these constraints as tractable second-order cone constraints (SOCC) and linear matrix inequalities (LMIs).

Corollary 1: The convex constraints (19) are equivalent to the following tractable constraints.

$$\beta_j + \alpha_j^\intercal\hat{\mu}_k(v) + \tau_{jk}\|\Sigma_w^{1/2}\mathcal{D}^\intercal(I+\mathcal{B}L)^\intercal E_k^\intercal\alpha_j\| + \varepsilon\rho_k\|\alpha_j\|\sqrt{1+\tau_{jk}^2} \le 0, \quad \forall j\ \forall k, \tag{21a}$$

$$\begin{bmatrix} I & E_k(I+\mathcal{B}L)\mathcal{D} \\ \mathcal{D}^\intercal(I+\mathcal{B}L)^\intercal E_k^\intercal & \rho_k I \end{bmatrix} \succeq 0 \quad \forall k, \tag{21b}$$

with respect to the decision variables $\{v, L, \rho_k\}$.
Along the same lines, in the next section, we reformulate the DR objective function (6) as a tractable convex program.

B. DR Objective Reformulation

Substituting the error dynamics (5) and feedback control $\tilde{\boldsymbol{u}} = K\tilde{\boldsymbol{x}} = L\mathcal{D}\boldsymbol{w}$ into the cost (6) yields

$$\begin{aligned}\mathcal{J} &= \beta\sum_{k=0}^{N-1}\|v_k\| + \max_{\mathbb{P}\in\mathcal{W}}\mathbb{E}_\mathbb{P}\left(\tilde{\boldsymbol{x}}^\intercal(\mathcal{Q}+K^\intercal\mathcal{R}K)\tilde{\boldsymbol{x}}\right)\\ &= \beta\sum_{k=0}^{N-1}\|E_k v\| + \max_{\mathbb{P}\in\mathbb{B}^{\|\cdot\|}_\varepsilon(\hat{\mathbb{P}})}\mathbb{E}_\mathbb{P}(\boldsymbol{w}^\intercal\Xi(L)\boldsymbol{w}),\end{aligned} \tag{22}$$

where $\mathcal{Q} \triangleq \mathrm{blkdiag}(Q_0,\ldots,Q_{N-1},0) \succeq 0, R \triangleq \mathrm{blkdiag}(R_0,\ldots,R_{N-1}) \succ 0$ are the augmented cost matrices, and $\Xi \triangleq \mathcal{D}^\intercal\left((I+\mathcal{B}L)\mathcal{Q}(I+\mathcal{B}L) + L^\intercal\mathcal{R}L\right)\mathcal{D} \succeq 0$. Thus, we aim to find the worst-case expected value of a quadratic form over the Wasserstein ambiguity set centered around the nominal distribution $\hat{\mathbb{P}} = \mathcal{N}(0,\Sigma_w)$. To this end, it can be shown [9] that this worst-case expectation is *equivalent* to the worst-case expectation with respect to the associated Gelbrich ambiguity set, provided that the nominal distribution is elliptical. The following result provides a reformulation of the DR cost.

Theorem 4: The DR quadratic cost in the objective function (22) is equivalent to the convex program

$$\min_{\lambda I \succ \Xi(L)} \lambda(\varepsilon^2 - \mathrm{tr}[\Sigma_w] + \lambda\mathrm{tr}[\Sigma_w(\lambda I - \Xi(L))^{-1}]). \tag{23}$$

Similar to the DR-CVaR constraints (19), the reformulated DR cost (23) is convex but nonlinear in the decision variables γ and L. Using Schur complements, we can reformulate (23) as the following SDP.

Corollary 2: The convex program (23) is equivalent to the semi-definite program

$$\min_{\substack{\lambda\ge 0\\ \Gamma,\Psi\succeq 0}} \quad \lambda(\varepsilon^2 - \mathrm{tr}[\Sigma_w]) + \mathrm{tr}[\Gamma], \tag{24a}$$

$$\text{s.t.} \quad \begin{bmatrix}\Gamma & \lambda\Sigma_w^{1/2}\\ \lambda\Sigma_w^{1/2} & \Psi\end{bmatrix} \succeq 0, \tag{24b}$$

$$\begin{bmatrix}\lambda I - \mathcal{D}^\intercal\tilde{M}(L)\mathcal{D} - \Psi & \mathcal{D}^\intercal L^\intercal\\ L\mathcal{D} & \tilde{\mathcal{R}}^{-1}\end{bmatrix} \succeq 0, \tag{24c}$$

where $\tilde{M} \triangleq \mathcal{Q} + M(L) + M^\intercal(L),\ M \triangleq \mathcal{Q}\mathcal{B}L$, and $\tilde{\mathcal{R}} \triangleq \mathcal{B}^\intercal\mathcal{Q}\mathcal{B} + \mathcal{R}$.
Lastly, in the next section, we reformulate the terminal constraints that require the terminal distributional uncertainty of the state to lie within a desired target ambiguity set $\mathbb{S}_f$.

C. Terminal Constraints

Using (14), the terminal propagated ambiguity set of the state is given by

$$\mathbb{S}_N = \mathbb{B}_\varepsilon^{\|\cdot\|\circ\tilde{L}_N^\dagger}(\hat{\mathbb{P}}_N), \quad \hat{\mathbb{P}}_N = \mathcal{N}(\bar{x}_N(v), \tilde{L}_N\Sigma_w\tilde{L}_N^\intercal). \quad (25)$$

To ensure the inclusion $\mathbb{S}_N \subseteq \mathbb{B}_\delta^{\|\cdot\|}(\hat{\mathbb{P}}_f)$, we first note that it can be shown [12] that

$$\mathbb{B}_\varepsilon^{\|\cdot\|\circ\tilde{L}^\dagger}(\hat{\mathbb{P}}) \subseteq \mathbb{B}_\varepsilon^{\sigma^2_{\min}(\tilde{L})\|\cdot\|}(\hat{\mathbb{P}}) = \mathbb{B}^{\|\cdot\|}_{\varepsilon\sigma^2_{\max}(\tilde{L})}(\hat{\mathbb{P}}). \quad (26)$$

Thus, since both $\hat{\mathbb{P}}_N$ and $\hat{\mathbb{P}}_f$ are normally distributed, it is sufficient to enforce the constraints

$$\bar{x}_N(v) = \mu_f, \quad \Sigma_{x_N}(L) \preceq \Sigma_f, \quad \varepsilon\sigma^2_{\max}(\tilde{L}_N) \leq \delta. \quad (27)$$

Remark 4: The first two constraints in (27) are equivalent to those in the traditional CS literature [5]. Indeed, the main goal of covariance control is to steer the covariance (and mean) of the state distribution to some desired terminal covariance, where the relaxation $\Sigma_{x_N} \preceq \Sigma_f$ is often introduced to make the terminal constraints tractable. In this context, however, the first two constraints align the center distributions of the terminal state, while the extra constraint in (27) can be interpreted as a way to robustify against distributionally uncertainty in the terminal state, providing an extra layer of safety guarantee against unknown disturbances. To this end, the terminal mean constraint in (27) is simply a linear constraint in v, given by

$$E_N(\mathcal{A}x_0 + \mathcal{B}v) - \mu_f = 0. \quad (28)$$

Second, the terminal covariance constraint in (27) can be written as the following LMI

$$\begin{bmatrix} \Sigma_f & E_N(I+\mathcal{B}L)\mathcal{D}\Sigma_w^{1/2} \\ \Sigma_w^{1/2}\mathcal{D}^\intercal(I+\mathcal{B}L)^\intercal E_N^\intercal & I \end{bmatrix} \succeq 0. \quad (29)$$

Lastly, noting that $\sigma_{\max}(L) = \|L\|$ and using the Schur complement, the terminal distributional uncertainty constraint can be written as the LMI

$$\begin{bmatrix} I & E_N(I+\mathcal{B}L)\mathcal{D} \\ \mathcal{D}^\intercal(I+\mathcal{B}L)^\intercal E_N^\intercal & (\delta/\varepsilon)I \end{bmatrix} \succeq 0. \quad (30)$$

In summary, combining all ingredients of Sections IV.A-IV.C, the DR-DS problem can be solved as the SDP

$$\min_{\substack{v,K \\ \rho_k,\lambda\geq 0 \\ \Gamma,\Psi\succeq 0}} \quad \beta\sum_{k=0}^{N-1}\|E_k v\| + \lambda(\varepsilon^2 - \mathrm{tr}[\Sigma_w]) + \mathrm{tr}[\Gamma]$$
$$\text{s.t.} \quad (21), (24b), (24c), (28), (29), (30).$$

V. NUMERICAL EXAMPLES

A. Double Integrator Path Planning

As a first example to showcase the proposed DR-DS framework, consider a 2D double integrator integrator with dynamics

$$A = \begin{bmatrix} I_2 & \Delta T I_2 \\ 0_2 & I_2 \end{bmatrix}, \quad B = \begin{bmatrix} \frac{\Delta T^2}{2}I_2 \\ \Delta T I_2 \end{bmatrix}, \quad D = 5\times 10^{-3}I_4,$$

with initial state $x_0 = [-1, 2, 0.1, -0.1]^\intercal$, and i.i.d. nominal disturbances drawn from $\hat{\mathbb{P}}_w = \mathcal{N}(0, I_4)$. The nominal target state distribution is $\hat{\mathbb{P}}_f = \mathcal{N}(0, (0.1/3)^2 I_4)$, and the desired target ambiguity set has radius $\delta = 0.05$. Lastly, the planning horizon has $N = 20$ time steps, $\Delta T = 0.3$, and we enforce probabilistic constraints with respect to the polytope defined by $\alpha_{1,[8:N]} = [-1, 0, 0, 0]^\intercal$, $\alpha_{2,[8:N]} = [1, 0, 0, 0]^\intercal$, and $b_{1,[8:N]} = b_{2,[8:N]} = -0.2$, which probabilistically enforces $|x| \leq 0.2$ in the terminal stage of planning, with probability $\gamma_{jk} = 0.05$ along each individual constraint. We compare the performance of the DR-DS control to that of the baseline CS solution with chance constraints [5]. The convex programs were all solved using the YALMIP optimization suite [20] with the MOSEK solver [21].

Firstly, we compare the optimal solutions subject to the nominal disturbances in Figure 1. Clearly, in the nominal

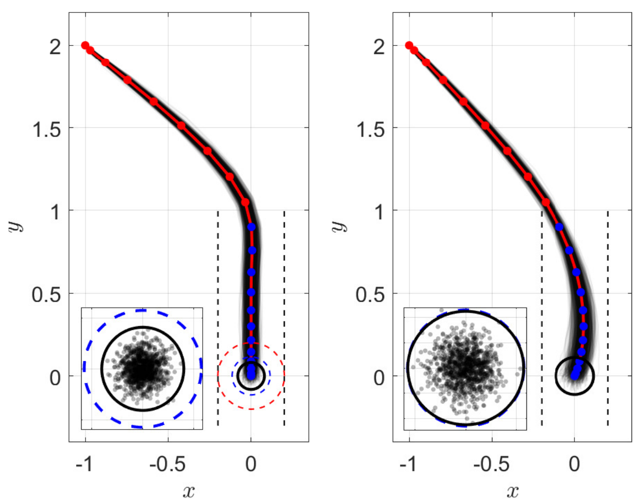

Fig. 1: Optimal trajectories for (left) DR-DS solution with $\varepsilon = 15$, and (right) baseline CS solution, subject to nominal disturbance $\mathbb{P}_w$.

case, when the disturbance is well-understood, both solutions are able to successfully steer to the desired terminal distribution and satisfy the constraints, since by construction, this is what CS is designed to do. Additionally, the empirical risk for the DR-CVaR constraints and (conservative) chance-constraints are both zero for 1,000 Monte-Carlo samples. Interestingly, however, note that the DR-DS solution steers to a smaller terminal covariance compared to that of CS. For reference, the red covariance ellipse in the left plot in Figure 1 is the *maximal* normal distribution in the target terminal ambiguity set with respect to the underlying structure of zero-mean Gaussian's, which is computed from $\mathbb{W}(\Sigma_f, \eta_f^2\Sigma_f) = \delta$, or equivalently, by using the Wasserstein distance between two normal distributions, as

$$\eta_f = 1 + \frac{\delta}{\sqrt{\mathrm{tr}(\Sigma_f)}}. \quad (31)$$

Thus, the DR-DS framework will steer the state distribution to $\Sigma_N \preceq \eta_f^2\Sigma_f$ for any disturbance $\mathbb{P}_w \in \mathbb{B}_\varepsilon^{\|\cdot\|}(\hat{\mathbb{P}}_w)$.

To illustrate this, Figure 2 shows the performance of the two methods when the noise distribution is now given by $\mathbb{P}_w = \mathcal{N}(0, \eta_w^2 I)$, where, as in (31), $\eta_w = 1 + \varepsilon/\sqrt{\mathrm{tr}(\Sigma_w)}$ is the maximal covariance in the ambiguity set $\mathbb{B}_\varepsilon^{\|\cdot\|}(\hat{\mathbb{P}}_w)$. When the true noise affecting the system is mis-characterized and not equivalent to the noise the system was designed to handle, the baseline CS is unable to steer to the terminal covariance nor satisfy the chance-constraints with the desired level of risk. The DR-DS solution, on the other hand, is agnostic to the noise distribution by design (within limits, of course), and is able to steer the state distribution to the terminal ambiguity set and still satisfies the CVaR constraints. Indeed, the empirical risk of constraint violation is 0.1% and 5.5%, respectively, for DR-DS and baseline CS.

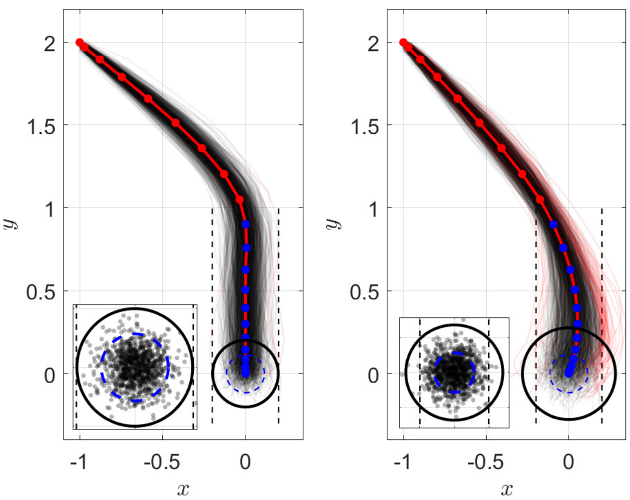

Fig. 2: Optimal trajectories for (left) DR-DS solution with $\varepsilon = 15$, and (right) baseline CS solution, subject to *maximal* disturbance $\mathbb{P}_w$ in disturbance ambiguity set $\mathbb{B}_\varepsilon^{\|\cdot\|}(\hat{\mathbb{P}}_w)$.

VI. CONCLUSION

In this work, we have developed a distributionally-robust density control method for steering the distributional uncertainty of the state of a linear dynamical system subject to imperfect knowledge of the disturbances affecting the system. Through characterizing the distributional uncertainty in the noise distribution via Wasserstein ambiguity sets, we are able to propagate the ambiguity set of the state through the LTI dynamics, and tractably formulate the DR objective function, DR-CVaR constraints, and terminal ambiguity set constraints as an SDP, which can be solved in polynomial time. We showcased the proposed methodology on a double integrator steering problem, illustrating safe planning under mis-characterized i.i.d. disturbances. Future work will aim to investigate tractable formulations of the DR-DS problem in data-driven settings, where the reference noise distribution is constructed from empirical samples.

VII. ACKNOWLEDGMENTS

This work has been supported by NASA University Leadership Initiative award 80NSSC20M0163 and ONR award N00014-18-1-2828. The article solely reflects the opinions and conclusions of its authors and not any NASA entity. We would also like to thank Dr. Daniel Kuhn for his discussion and helpful insights.

REFERENCES

[1] V. Renganathan, J. Pilipovsky, and P. Tsiotras, "Distributionally robust covariance steering with optimal risk allocation," in *American Control Conference*, San Diego, CA, May 31–June 2, 2023, pp. 2607–2614.

[2] A. F. Hotz and R. E. Skelton, "Covariance control theory," *International Journal of Control*, vol. 46, no. 1, pp. 13–32, 1987.

[3] E. Bakolas, "Finite-horizon separation-based covariance control for discrete-time stochastic linear systems," in *57th IEEE Conference on Decision and Control*, Miami Beach, FL, Dec 17–19, 2018, pp. 3299–3304.

[4] M. Goldshtein and P. Tsiotras, "Finite-horizon covariance control of linear time-varying systems," in *56th IEEE Conference on Decision and Control*, Melbourne, Australia, Dec 12–15 2017, pp. 3606–3611.

[5] K. Okamoto, M. Goldshtein, and P. Tsiotras, "Optimal covariance control for stochastic systems under chance constraints," *IEEE Control System Letters*, vol. 2, pp. 266–271, 2018.

[6] G. Rapakoulias and P. Tsiotras, "Discrete-time optimal covariance steering via semidefinite programming," in *62nd IEEE Conference on Decision and Control*, Marina Bay Sands, Singapore, Dec. 13–15, 2023.

[7] J. Pilipovsky and P. Tsiotras, "Computationally efficient chance constrained covariance control with output feedback," in *63rd IEEE Conference on Decision and Control*, Milan, Italy, Dec. 16–19, 2024, accepted.

[8] ——, "Data-driven covariance steering control design," in *62nd IEEE Conference on Decision and Control*, Marina Bay Sands, Singapore, Dec. 13–15, 2023, pp. 2610–2615.

[9] V. A. Nguyen, S. Shafiee, D. Filipović, and D. Kuhn, "Mean-covariance robust risk measurement," 2023, arXiv: 2112.09959.

[10] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations," *Mathematical Programming*, vol. 171, pp. 115 – 166, 2015.

[11] S. Shafieezadeh-Abadeh, L. Aolaritei, F. Dörfler, and D. Kuhn, "New perspectives on regularization and computation in optimal transport-based distributionally robust optimization," 2023, arXiv: 2303.03900.

[12] L. Aolaritei, N. Lanzetti, H. Chen, and F. Dörfler, "Distributional uncertainty propagation via optimal transport," 2023, arXiv: 2205.00343.

[13] L. Aolaritei, M. Fochesato, J. Lygeros, and F. Dörfler, "Wasserstein tube mpc with exact uncertainty propagation," in *62nd IEEE Conference on Decision and Control*, Marina Bay Sands, Singapore, Dec. 13–15, 2023, pp. 2036–2041.

[14] N. Fournier and A. Guillin, "On the rate of convergence in Wasserstein distance of the empirical measure," *Probability Theory and Related Fields*, vol. 162, no. 3-4, p. 707, Aug 2015.

[15] L. Blackmore, M. Ono, and B. C. Williams, "Chance-constrained optimal path planning with obstacles," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1080–1094, 2011.

[16] A. Nemirovski, "On safe tractable approximations of chance constraints," *European Journal of Operational Research*, vol. 219, no. 3, pp. 707–718, 2012.

[17] J. Ridderhof, J. Pilipovsky, and P. Tsiotras, "Chance-constrained covariance control for low-thrust minimum-fuel trajectory optimization," in *AAS/AIAA Astrodynamics Specialist Conference*, Lake Tahoe, CA, Aug 9–13 2020.

[18] J. Pilipovsky and P. Tsiotras, "Distributionally robust density control with Wasserstein ambiguity sets," 2024, arXiv: 2403.12378.

[19] D. Kuhn, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, "Wasserstein distributionally robust optimization: Theory and applications in machine learning," 2019, arXiv: 1908.08729.

[20] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in matlab," in *IEEE International Symposium on Computer Aided Control Systems Design*, Taipei, Taiwan, 2004, pp. 284–289.

[21] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019. [Online]. Available: http://docs.mosek.com/9.0/toolbox/index.html