

Distributionally Robust Control of Constrained Stochastic Systems

Bart P. G. Van Parys, Daniel Kuhn, Paul J. Goulart, *Member, IEEE*, and Manfred Morari, *Fellow, IEEE*

Abstract—We investigate the control of constrained stochastic linear systems when faced with limited information regarding the disturbance process, i.e., when only the first two moments of the disturbance distribution are known. We consider two types of *distributionally robust* constraints. In the first case, we require that the constraints hold with a given probability for all disturbance distributions sharing the known moments. These constraints are commonly referred to as *distributionally robust chance* constraints. In the second case, we impose conditional value-at-risk (CVaR) constraints to bound the expected constraint violation for all disturbance distributions consistent with the given moment information. Such constraints are referred to as *distributionally robust CVaR* constraints with second-order moment specifications. We propose a method for designing linear controllers for systems with such constraints that is both computationally tractable and practically meaningful for both finite and infinite horizon problems. We prove in the infinite horizon case that our design procedure produces the globally optimal linear output feedback controller for distributionally robust CVaR and chance constrained problems. The proposed methods are illustrated for a wind blade control design case study for which distributionally robust constraints constitute sensible design objectives.

Index Terms—Linear time-invariant (LTI), stochastic system.

I. INTRODUCTION

THE problem of finding a control policy for a dynamic system such that its state and inputs remain in a given constraint set, despite the uncertain nature of the system, is an important and well studied problem within the control literature. In this article, we focus on discrete-time linear time-invariant (DLTI) systems, with system dynamics

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{C}\mathbf{w}_t$$

where \mathbf{x}_t is the state of the system, and \mathbf{u}_t and \mathbf{w}_t are the input and disturbance to the system, respectively. In constrained

control problems, we are given a constraint set \mathbb{X} for which we would like to model the requirement that “ $\mathbf{x}_t \in \mathbb{X}$,” i.e., the state \mathbf{x}_t must remain in \mathbb{X} in *some sense*. There are several common approaches to putting the design requirement “ $\mathbf{x}_t \in \mathbb{X}$ ” on a mathematically sound footing. We first describe two standard methods for modeling such a constraint; the now classical worst-case approach and the more recent chance-constrained approach. Our notation is intentionally informal to start, with a more rigorous treatment deferred to later in the paper. We will propose two alternative approaches that overcome some of the deficiencies inherent in the two standard methods, particularly when faced with only a limited amount of information regarding the moments of the disturbance process \mathbf{w}_t .

Worst-Case Constraints: The problem of finding a control policy such that the state of an uncertain dynamical system remains in a given constraint set, for all possible disturbance realizations, is historically the most prevalent design goal within the control literature [4], [5], [23], [34]. This worst-case (or robust) formulation starts by assuming that the disturbance support is bounded and known, i.e., that the disturbance \mathbf{w}_t is restricted to be an element of a bounded set W . The constraint “ $\mathbf{x}_t \in \mathbb{X}$ ” is then interpreted as a condition that the state \mathbf{x}_t is an element of \mathbb{X} for all realizations of the disturbance process \mathbf{w}_t generated from W . Identification of an optimal control policy for such problems is computationally intractable in general, so significant research effort has focussed on the development of design methods that provide admissible, but possibly suboptimal, control policies; see [16], [23] and the references therein.

The worst-case formulation requires that the support of the disturbance process is completely known and, in the presence of state constraints, a bounded set. This assumption may be quite restrictive, e.g., in cases where the disturbances are normally distributed and hence the disturbance has unbounded support. This motivates the need for approaches that make no such assumption regarding the support of the disturbance.

Chance Constraints: Chance constraints require that the system’s state constraints hold only with a specified probability level [12], [13], [32]. The constraint “ $\mathbf{x}_t \in \mathbb{X}$ ” is then modeled as

$$\mathbb{P}^* \{ \mathbf{x}_t \in \mathbb{X} \} \geq 1 - \epsilon \quad (1)$$

where the probability measure \mathbb{P}^* is defined on the disturbance process \mathbf{w}_t and is assumed known. Although no boundedness assumption is required on the disturbance support $\text{supp}\{\mathbb{P}^*\}$, in contrast to the worst-case approach, chance constraints are arguably worse from a practical perspective since they require the availability of a probability measure over the disturbances. Unfortunately, verifying a chance constraint in the form (1) is

Manuscript received June 17, 2014; revised May 5, 2015; accepted May 6, 2015. Date of publication June 11, 2015; date of current version January 26, 2016. This work was supported by Marie Curie FP7-ITN project “Energy savings from smart operation of electrical, process and mechanical equipment—ENERGY-SMARTOPS,” Contract PITN-GA-2010-264940, by the Engineering and Physical Sciences Research Council (EP/I014640/1), and by the EU FP7 project EMBOCON (ICT-2009-4-248940). Recommended by Associate Editor C. M. Lagoa.

B. P. G. Van Parys and M. Morari are with the Automatic Control Laboratory, Swiss Federal Institute of Technology Zürich (ETHZ), 8092 Zürich, Switzerland (e-mail: vanparysbart@gmail.com; bart.vanparys@control.ee.ethz.ch; morari@control.ee.ethz.ch).

D. Kuhn is with the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland (e-mail: daniel.kuhn@epfl.ch).

P. J. Goulart is with the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, U.K. (e-mail: paul.goulart@eng.ox.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2015.2444134

intractable under generic distributions, i.e., checking (1) for a given state distribution \mathbf{x}_t is \mathcal{NP} -hard [25]. As a consequence, recent attention has shifted towards stochastic sampling methods, for which only probabilistic guarantees can typically be provided, e.g., that the chance constraint condition holds only with some level of confidence [8], [9].

In this paper we take an approach intermediate to these two extremes. Our goal is to provide a framework that addresses the constraint “ $\mathbf{x}_t \in \mathbb{X}$ ” using only partial information about the *true but unknown* disturbance probability measure \mathbb{P}^* , and without recourse to sampling. We briefly describe both of the new constraint models that will be introduced as alternatives to the worst-case and chance constrained problem formulations.

Distributionally Robust Chance Constraints: In many situations the disturbance distribution \mathbb{P}^* is unknown and must be estimated from historical data, and hence is uncertain. We therefore assume only that the distribution \mathbb{P}^* belongs to a set \mathcal{P} of distributions that share certain known structural properties, i.e. their first two moments. The distributionally robust counterpart [38] of the chance constraint (1) hence becomes

$$\forall \mathbb{P} \in \mathcal{P} : \quad \mathbb{P}\{\mathbf{x}_t \in \mathbb{X}\} \geq 1 - \epsilon. \quad (2)$$

The constraint (2) is referred to as a *distributionally robust chance constraint* [10], [38] on \mathbf{x}_t with a second moment specification. Such a constraint is a robust version of the classical chance constraint (1) in that it is insensitive to ambiguity in the disturbance distribution \mathbb{P}^* , at least with respect to its higher order moments. Alternatively in [2], [21] and [22], the ambiguity set \mathcal{P} contains all symmetric distributions sharing a unimodal structural property¹ with known rectangular support. One of the main advantages of these formulations over the classical chance constrained formulation is the fact that only partial information on the disturbance distribution \mathbb{P}^* is required.

Distributionally Robust Conditional Value-at-Risk (CVaR) Constraints: For chance constraints in the form (2), a natural additional goal is to guarantee that the expected constraint violation in the remaining ϵ percent of the cases is small. To model such a requirement, we will consider distributionally robust CVaR constraints [37]. Our general approach is to measure, via some loss function L , the severity of constraint violation in the ϵ percent of the worst cases and to keep the degree of loss by this measure small. A precise statement of this distributionally robust CVaR approach requires some additional notation, and we defer the formalities to Section I.

We will show in the remainder of the paper how both the distributionally robust chance constraint and worst-case CVaR constraint interpretations of the condition “ $\mathbf{x}_t \in \mathbb{X}$ ” constitute mathematically sound control design specifications. In particular, we will show that for both finite and infinite horizon optimal control problems with either constraint type, the resulting problems are both practically meaningful and computationally tractable. We stress that all of the numerical methods we present for solving such problems are deterministic, in contrast to the stochastic methods presented in [8], for which only probabilistic admissibility guarantees can be provided. Indeed, we will prove that for infinite horizon control problems our

design procedure produces the globally optimal linear output feedback controller for distributionally robust CVaR and chance constrained problems.

Outline: We provide a mathematically rigorous description of distributionally robust chance and CVaR constraints in the context of control design problems in Section I, which can be read as an extended introduction. In Sections II and III, we propose two control design problems, of finite and infinite horizon type respectively. We show that finding the globally optimal linear control policy in either case is a tractable problem when considered in conjunction with either of our alternative constraint descriptions for ellipsoidal sets \mathbb{X} . Additionally in the infinite horizon case, we offer exact tractable conditions on whether a linear system satisfies our alternative constraints for arbitrary polytopic sets \mathbb{X} . The latter of our proposed control problems is illustrated for a wind turbine blade control design case study in Section IV, for which an assumption of limited moment information on the disturbance is quite natural.

Notation and Definitions: We denote by \mathbb{I}_n the identity matrix in $\mathbb{R}^{n \times n}$ and by \mathbb{S}_+^n and \mathbb{S}_{++}^n the sets of all positive semidefinite and positive definite symmetric matrices in $\mathbb{R}^{n \times n}$, respectively. The diagonal concatenation of two matrices X and Y is denoted by $\text{diag}(X, Y)$. The Kronecker product of two matrices X and Y is denoted as $X \otimes Y$. For notational convenience, random vectors will be denoted in boldface, while their realizations will be denoted by the same symbols in normal font. For any random variable \mathbf{x} we introduce the following shorthand notation:

$$\begin{aligned} \mu_x &:= \mathbb{E}_{\mathbb{P}^*}\{\mathbf{x}\}, & \Sigma_x &:= \mathbb{E}_{\mathbb{P}^*}\{[\mathbf{x} - \mu_x] \cdot [\mathbf{x} - \mu_x]^\top\} \\ M_x &:= \begin{pmatrix} \Sigma_x + \mu_x \mu_x^\top & \mu_x \\ \mu_x^\top & 1 \end{pmatrix}. \end{aligned}$$

II. MATHEMATICAL PRELIMINARIES

Chance constraints are a popular means of modeling *soft constraints* on uncertain variables that need only to hold with a certain probability. Formally, the requirement that an n -dimensional random vector \mathbf{x} should be contained in a set $\mathbb{X} \subseteq \mathbb{R}^n$ with high probability is expressed as

$$\mathbb{Q}^*(\mathbf{x} \in \mathbb{X}) \geq 1 - \epsilon \quad (3)$$

where ϵ is a prescribed safety parameter that controls the level of acceptable constraint violations. The random vector \mathbf{x} is a surjective measurable function on the probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{Q}^*)$. The measure \mathbb{Q}^* denotes the true (but possibly unknown) probability measure on this space.

In the present section we will regard \mathbf{x} as a generic random vector and present a number of essential results relating to modeling the constraint (3). In the sequel, we will consider the specific situation of a variable representing the random state of a linear dynamical system that depends both on previous control inputs (actions) and exogenous disturbances (noise).

Chance constraints are often more practical than hard constraints, which can be viewed as degenerate chance constraints with $\epsilon = 0$ and which tend to encourage overly conservative decisions. More importantly, in linear dynamical systems hard state constraints typically become infeasible in the presence of unbounded (e.g., Gaussian) noise.

¹This unimodal property can be interpreted to mean that small disturbances are more likely than large disturbances, see also [27].

In spite of their conceptual appeal, chance constraints have not yet found wide application in optimization and control theory for a variety of reasons. On the one hand, the feasibility of a chance constraint can only be checked if the *true* distribution of the random vector \mathbf{x} (which is determined by the true probability measure \mathbb{Q}^*) is precisely known. In practice, however, almost invariably this distribution must be estimated from noisy data and is therefore itself subject to ambiguity. This is problematic because even small changes in the distribution can have a dramatic impact on the geometry and size of the set of inputs or actions consistent with the chance constraint [38]. Moreover, incorporating chance constraints into otherwise tractable optimization problems typically results in a non-convex problem, and consequently to computational intractability.

Finally, chance constraints of the type (3) bound the probability of constraint violation but do not impose any restrictions on the *degree* of any violations encountered. However, severe constraint violations, i.e., scenarios in which the system state strays far outside of \mathbb{X} , are often much more harmful than mild violations in which the state remains close to the boundary of \mathbb{X} . Chance constraints fail to distinguish between these two situations and provide no mechanism to penalize severe constraint violations relative to mild ones.

In order to address these deficiencies, we first require some terminology and notation. We will assume throughout that the set \mathbb{X} is characterised by the intersection of zero sublevel sets of finitely many convex functions $L_i : \mathbb{R}^n \rightarrow \mathbb{R}$, so that $\mathbb{X} := \{x \in \mathbb{R}^n \mid L_i(x) \leq 0, \forall i = 1, \dots, I\}$ and $\text{int } \mathbb{X} = \{x \in \mathbb{R}^n \mid L_i(x) < 0 \forall i = 1, \dots, I\}$. We will refer to the functions L_i as *loss functions*. We refer to (3) as an *individual* chance constraint if $I = 1$ and as a *joint* chance constraint if $I > 1$. Every joint chance constraint can easily be reduced to an individual chance constraint by re-expressing \mathbb{X} as $\{x \in \mathbb{R}^n \mid L^\alpha(x) \leq 0\}$, where the aggregate loss function

$$L^\alpha(x) := \max_{i=1, \dots, I} \alpha_i L_i(x) \quad (4)$$

remains convex in x and depends on a set of strictly positive scaling parameters $\alpha \in \mathbb{R}_{++}^n$. Note that the particular choice of α has no impact on the zero sublevel set of the function L^α , and consequently no impact on the set \mathbb{X} or the associated chance constraint (3). The reader may therefore regard α initially as a positive parameter that can be chosen arbitrarily. However, the flexibility to select α will be useful at a later stage either to control the tightness of a tractable approximation of the chance constraint (3), or to penalise the degree of constraint violation of statistical outliers in (3).

Throughout the paper we will exploit an interesting connection between chance constraints of the type (3) and quantile-based risk measures that are commonly used in economics. Given some measurable loss function $L^\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ and probability measure \mathbb{Q} on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, the so-called *Value-at-Risk* (VaR) at level ϵ with respect to \mathbb{Q} is defined as

$$\mathbb{Q}\text{-VaR}_\epsilon(L^\alpha(\mathbf{x})) := \inf \{ \gamma \in \mathbb{R} \mid \mathbb{Q}(L^\alpha(\mathbf{x}) > \gamma) \leq \epsilon \}.$$

We emphasize that the “value” at risk in our particular context typically relates to the degree of violation of some physical state constraint, and is unrelated to the loss of economic currency as in the usual interpretation in economics. In control applica-

tions, “violation” at risk might therefore be a more appropriate interpretation.

By construction, the VaR coincides with the $(1 - \epsilon)$ -quantile of the distribution of $L^\alpha(\mathbf{x})$. Moreover, the reader may easily verify that the chance constraint (3) can be reformulated as a constraint on the VaR at level ϵ of the aggregate loss function $L^\alpha(\mathbf{x})$, that is

$$\mathbb{Q}^*\text{-VaR}_\epsilon(L^\alpha(\mathbf{x})) \leq 0 \iff \mathbb{Q}^*(\mathbf{x} \in \mathbb{X}) \geq 1 - \epsilon. \quad (5)$$

A major deficiency of the VaR is its non-convexity in $L^\alpha(\mathbf{x})$. In fact, it is well known that the function $\mathbb{Q}\text{-VaR}_\epsilon(L^\alpha(\mathbf{x}))$ is generally non-convex in \mathbf{x} even for linear loss functions. A commonly employed alternative, convex, risk measure closely related to the VaR is the conditional value-at-risk:

Definition 1.1 (Conditional Value-at-Risk): For any measurable loss function $L^\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$, probability distribution \mathbb{Q} on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and tolerance $\epsilon \in (0, 1)$, the CVaR of the random loss $L^\alpha(\mathbf{x})$ at level ϵ with respect to \mathbb{Q} is defined as

$$\mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x})) := \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{Q}} \{ (L^\alpha(\mathbf{x}) - \beta)^+ \} \right\}. \quad (6)$$

Rockafellar and Uryasev [31] have shown that the set of optimal solutions for β in (6) is a closed interval whose left endpoint is given by $\mathbb{Q}\text{-VaR}_\epsilon(L^\alpha(\mathbf{x}))$. Moreover, it can be shown that if the random loss $L^\alpha(\mathbf{x})$ follows a continuous distribution, then CVaR coincides with the conditional expectation of $L^\alpha(\mathbf{x})$ above $\mathbb{Q}\text{-VaR}_\epsilon(L^\alpha(\mathbf{x}))$. This observation originally motivated the term *conditional value-at-risk*.

CVaR enjoys a number of practical advantages over VaR, since it is monotone, homogeneous and convex with respect to the loss function L^α . In addition, it represents a conservative (upper) approximation to VaR, and consequently a conservative means of approximating chance constraints. Indeed, it is easily shown that

$$\mathbb{Q}^*\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x})) \leq 0 \implies \mathbb{Q}^*(\mathbf{x} \in \mathbb{X}) \geq 1 - \epsilon. \quad (7)$$

Note that for convex loss functions the set of all random vectors \mathbf{x} satisfying the CVaR constraint in (7) is convex due to the convexity and monotonicity of CVaR.

In economic theory, CVaR traditionally measures an economic loss, hence the function L^α is specified *ab initio*. In control practice however, one is typically given a constraint set \mathbb{X} and is free to select any loss functions L_i compatible with \mathbb{X} , i.e., one can choose any L_i satisfying $\mathbb{X} = \{x \in \mathbb{R}^n \mid L^\alpha(x) \leq 0\}$. The choice of the positive weights α_i can then be used to indicate the relative importance of the individual loss functions L_i , i.e., the level of significance that the control designer attaches to the degree of violation of individual state constraints in the event that they occur.

CVaR constraints address two of the main shortcomings of chance constraints. First, unlike chance constraints, they lead to tractable convex optimization problems. Second, CVaR constraints impose a higher penalty on realizations of \mathbf{x} that materialize far outside of \mathbb{X} (i.e., with $L^\alpha(\mathbf{x}) \gg 0$) and therefore penalize severe constraint violations more aggressively than mild ones. In contrast, chance constraints impose uniform penalties on all constraint violations irrespective of their degree of infeasibility.

Unfortunately, checking the feasibility of CVaR constraints still requires precise knowledge of the *true* probability measure \mathbb{Q}^* . In practice, only limited information about \mathbb{Q}^* may be available, such as the support or some descriptive measures of the location and dispersion of random variables under \mathbb{Q}^* . Abstractly, we can represent the limited available information about \mathbb{Q}^* by a set \mathcal{Q} of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}^n))$ with the following properties: (i) It is known that $\mathbb{Q}^* \in \mathcal{Q}$, and (ii) \mathcal{Q} is the smallest set of probability distributions for which we can guarantee that $\mathbb{Q}^* \in \mathcal{Q}$. We will henceforth refer to \mathcal{Q} as an *ambiguity set*.

To immunize the chance constraint (3) against distributional ambiguity, we can require that it should hold for each probability measure in the ambiguity set. The resulting *distributionally robust chance constraint* can be represented as

$$\mathbb{Q}(x \in \mathbb{X}) \geq 1 - \epsilon \quad \forall \mathbb{Q} \in \mathcal{Q} \iff \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(x \in \mathbb{X}) \geq 1 - \epsilon. \quad (8)$$

Similarly, recalling that $\mathbb{X} = \{x \in \mathbb{R}^n \mid L^\alpha(x) \leq 0\}$ for any $\alpha \in \mathbb{R}_{++}$, we can immunize the CVaR constraint on the left hand side of (7) against distributional ambiguity. The resulting *distributionally robust CVaR constraint* takes the form

$$\begin{aligned} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(x)) &\leq 0 \quad \forall \mathbb{Q} \in \mathcal{Q} \\ \iff \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(x)) &\leq 0. \end{aligned} \quad (9)$$

As in the classical setting without distributional ambiguity, it can be shown that (9) provides a conservative approximation for (8); see [14], [38]. In other words

$$\sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(x)) \leq 0 \implies \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}(x \in \mathbb{X}) \geq 1 - \epsilon. \quad (10)$$

In order to further characterize the relation between the worst-case probability of constraint violation and worst-case CVaR, and to facilitate statements about computational tractability, we require some structural assumptions about the ambiguity set \mathcal{Q} and the loss functions $L_i(x)$, $i = 1, \dots, I$ defining the constraint set \mathbb{X} . We henceforth assume that the ambiguity set \mathcal{Q} contains all probability measures \mathbb{Q} under which the random variable x has a given mean value $\mu_x \in \mathbb{R}^n$ and a given covariance matrix $\Sigma_x \in \mathbb{S}_{++}^n$. Moreover, we require that each constraint function L_i is convex and quadratic, and is represented as $L_i(x) = x^\top E_i x + 2e_i^\top x + e_i^0$ for some $E_i \in \mathbb{S}_+^n$, $e_i \in \mathbb{R}^n$ and $e_i^0 \in \mathbb{R}$. This means that \mathbb{X} can be any intersection of half-spaces and generalized ellipsoids. The aggregate loss $L^\alpha(x)$ is then given by a pointwise maximum of a finite collection of quadratic functions.

We next recall some tractability and exactness results relating to the CVaR approximation:

Theorem I.2 (Tractability of Worst-Case CVaR [38]): Assume that the ambiguity set

$$\mathcal{Q} = \left\{ \mathbb{Q} \mid \int (x^\top, 1)^\top \cdot (x^\top, 1) d\mathbb{Q} = M_x \right\}$$

is based on mean and variance information and the loss function $L^\alpha(x)$ is given as in (4), where $L_i(x) = x^\top E_i x + 2e_i^\top x + e_i^0$ for some $E_i \in \mathbb{S}_+^n$, $e_i \in \mathbb{R}^n$ and $e_i^0 \in \mathbb{R}$. Then the worst-case

CVaR is equivalent to the following tractable semidefinite program (SDP):

$$\begin{aligned} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(x)) &= \inf_{\beta, X} \beta + \frac{1}{\epsilon} \text{Tr}\{M_x X\} \\ \text{s.t. } X &\in \mathbb{S}_+^{n+1}, \beta \in \mathbb{R} \\ X - \begin{pmatrix} \alpha_i E_i & \alpha_i e_i \\ \alpha_i e_i^\top & \alpha_i e_i^0 - \beta \end{pmatrix} &\succeq 0, \forall i \in \{1, \dots, I\}. \end{aligned} \quad (11)$$

An immediate consequence of Theorem I.2 is that it provides a tractable, convex and conservative means of approximating a distributionally robust chance constraint using LMI constraint by virtue of the relation (10). It has further been shown that such an approximation becomes *essentially exact* for a judicious choice of the scaling parameters $\alpha \in \mathbb{R}_{++}^n$; see [38, Thm. 3.6].

If there is only one loss function that is concentric with the distribution of x , then the SDP (11) admits an analytical solution. So far, this special case has not been considered in the literature. However, we will see in Section III that it is relevant for constrained infinite horizon problems.

Corollary I.3 (Concentric Distributions and Loss Functions): If \mathbb{X} constitutes a single ellipsoid centred at the origin (i.e., $I = 1$ and $e_1 = 0$), while the random vector x has mean $\mu_x = 0$ and variance Σ_x , then

$$\sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L_1(x)) = e_1^0 + \frac{1}{\epsilon} \text{Tr}\{\Sigma_x E_1\}. \quad (12)$$

admits an closed form exact reformulation.

Proof: See Appendix A. ■

As in the case of (11) in Theorem I.2, (12) in Corollary I.3 provides an immediate means of modeling a distributionally robust chance constraint via a single linear inequality constraint by virtue of the relation (10). In [38, Thm 2.2] it is shown that the implication in (10) is bidirectional under the conditions stated in Corollary I.3. Corollary I.3 hence offers an exact reformulation for distributionally robust chance constraints as well. In the next section, we will use these results to remodel chance constraints appearing in optimal control problems for linear systems.

III. FINITE HORIZON DISTRIBUTIONALLY ROBUST CONTROL PROBLEMS

We consider a DLTI system with n states, m control inputs, r outputs, d exogenous inputs or disturbances and r measurements

$$\begin{cases} x_{t+1} = Ax_t + Bu_t + Cw_t & \text{and } x_0 = x \\ y_t = Dx_t + Ew_t \end{cases} \quad (S)$$

where all matrices are of appropriate dimension and the disturbances w_t model both process noise (via the term Cw_t) and measurement noise (via Ew_t).

When addressing chance- or CVaR-constrained control of the uncertain system \mathcal{S} , we will require a more sophisticated abstract probability space $(\Omega, \mathcal{F}, \mathbb{P}^*)$ than that employed in the previous section. We will henceforward assume that the sample space Ω is sufficiently rich such that any (joint) distribution of all the random variables appearing in \mathcal{S} on the Cartesian product of their individual range spaces is induced by a probability measure on (Ω, \mathcal{F}) , and will denote by \mathcal{P}_0 the set of all such

probability measures.² We will assume that the *true* underlying probability measure $\mathbb{P}^* \in \mathcal{P}_0$ is imperfectly known. We use the notation \mathbb{P}^* for the underlying probability measure instead of \mathbb{Q}^* to remind the reader of the difference in probability space with respect to the last section.

Our goal is to design a finite-horizon control law for the system \mathcal{S} that minimizes an expected value quadratic cost, subject to an additional requirement that the state satisfies the constraint " $\mathbf{x}_t \in \mathbb{X}$ " in either a chance- or CVaR-constrained sense. The control inputs \mathbf{u}_t will be restricted to be $\mathcal{F}_t^y := \sigma(\mathbf{y}_0, \dots, \mathbf{y}_t)$ -measurable throughout. We wish to do this despite some ambiguity on the disturbance distribution. Specifically, we assume only that the following information is available about the disturbance process:

Assumption II.1 (Weak Sense Stationary Disturbances): We assume that in the DLTI system \mathcal{S} , the disturbance \mathbf{w}_t is a weak sense stationary (w.s.s.) white noise process with covariance matrix $\Sigma_{\mathbf{w}_t} =: \Sigma$ and mean $\mu_{\mathbf{w}_t} =: \mu$ for all $t \in \mathbb{N}_0$.

The w.s.s. assumption appears frequently in signal processing [30], but is less common in the control literature. In effect, it assumes that only the autocorrelation $R_{ww}(t) := \mathbb{E}_{\mathbb{P}^*} \{\mathbf{w}_i \cdot \mathbf{w}_{i-t}^\top\}$ is known, with $R_{ww}(0) = \Sigma + \mu\mu^\top$ and $R_{ww}(t) = \mu\mu^\top$ otherwise. Furthermore, knowing the first two moments of a w.s.s. process is, by merit of the Wiener-Khinchine Theorem, equivalent to knowing its power spectrum [30]. Estimating the spectral density of the disturbance \mathbf{w}_t , for example from historical data,³ is significantly easier in practice than determining the complete marginal distribution of \mathbf{w}_t with respect to \mathbb{P}^* .

The w.s.s. assumption implies that the only information available about the disturbance distribution is its autocorrelation function. Hence, the underlying probability measure \mathbb{P}^* is only known to be an element of the ambiguity set

$$\mathcal{P}_\infty := \left\{ \mathbb{P} \in \mathcal{P}_0 \mid \mathbb{E}_{\mathbb{P}} \left\{ \begin{pmatrix} \mathbf{w}_i^\top \\ 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{w}_j^\top \\ 1 \end{pmatrix}^\top \right\} = \begin{pmatrix} \Sigma \delta_{ij} + \mu\mu^\top & \mu \\ \mu^\top & 1 \end{pmatrix}, \quad \forall i, j \in \mathbb{N}_0 \right\}. \quad (13)$$

The set \mathcal{P}_∞ contains all probability measures consistent with the known moment information about the system disturbance. Notice that the measures in \mathcal{P}_∞ are defined for an infinite horizon. This will permit us to work with the same ambiguity set \mathcal{P}_∞ despite varying horizons in the finite horizon setting. When choosing a control policy for the system \mathcal{S} , we will require that it be distributionally robust with respect to the ambiguity set \mathcal{P}_∞ , in either a chance constrained or CVaR sense, for the constraint " $\mathbf{x}_t \in \mathbb{X}$." In order to achieve this control design objective, the notion of a distributionally robust constraint, introduced in Section I, is now used to formulate our control problem.

Control Constraints: We will consider distributionally robust constraints for the system \mathcal{S} enforced over a finite time horizon of length N , i.e.,

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0 \quad \forall t \in \{0, \dots, N-1\}. \quad (14)$$

²This means that we can think of Ω as the Cartesian product of all the random variables' range spaces, in which case \mathcal{F} is identified with the Borel σ -algebra on Ω , while each random variable reduces to a coordinate projection.

³The study of this problem is referred to as *spectral density estimation* in the signal processing community.

We assume that the parameter $\alpha \in \mathbb{R}_{++}^n$ is given, either as an attempt to approximate a distributionally robust chance constraint or as an indicator of the relative importance of the loss severity measures L_i .

Assumption II.2: An aggregated loss function $L^\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ with $\alpha \in \mathbb{R}_{++}^n$ for the distributionally robust CVaR constraints (14) is given as $L^\alpha(\mathbf{x}) = \max_{i \in \{1, \dots, I\}} [\alpha_i(\mathbf{x}^\top E_i \mathbf{x} + 2e_i^\top \mathbf{x} + e_i^0)]$ where $E_i \in \mathbb{S}_+^n$, $e_i \in \mathbb{R}^n$, $e_i^0 \in \mathbb{R}$.

Recall that the constraint set \mathbb{X} corresponds to the zero sub level set of the loss function L^α , and consequently is assumed to be a finite intersection of half-spaces and generalized ellipsoids by virtue of Assumption II.2.

For the system \mathcal{S} we define a causal control policy $\pi_N := \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}\}$, such that the control input selected at each time $t \in [0, \dots, N-1]$ is a function mapping prior measurements to actions, i.e. \mathbf{u}_t is \mathcal{F}_t^y -measurable, where we assume that the initial state $\mathbf{x}_0 = \mathbf{x}$ is known without any loss of generality.⁴ We denote the set of all such policies as Π_N . We wish to find, if it exists, a policy $\pi_N \in \Pi_N$ such that system \mathcal{S} satisfies the CVaR constraints (14) over a finite horizon. We refer to such a policy as *admissible* with respect to the system \mathcal{S} and the CVaR constraints (14).

Objective Function: Our aim is to find a causal control policy $\pi_N \in \Pi_N$ that is admissible with respect to the CVaR constraints while minimizing a given objective function J_N . We will assume here that the objective function $J_N : \mathbb{R}^n \times \Pi_N \rightarrow \mathbb{R}_+$ is a discounted sum of quadratic stage costs, i.e., that it is in the form

$$J_N(\mathbf{x}, \pi_N) := \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}} \left\{ \sum_{t=0}^{N-1} \beta^t [\mathbf{x}_t^\top Q \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t] + \beta^N \mathbf{x}_N^\top Q_f \mathbf{x}_N \right\} \quad (15)$$

where we refer to $\beta \in (0, 1]$ as the discount factor of the control cost. It is assumed that the objective function J_N is convex, i.e., $Q, Q_f \in \mathbb{S}_+$ and $R \in \mathbb{S}_{++}$. We are therefore interested in the solution to the optimal control problem

$$\begin{aligned} \inf_{\pi_N \in \Pi_N} \quad & J_N(\mathbf{x}, \pi_N) \\ \text{s.t.} \quad & \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + C\mathbf{w}_t, \quad \mathbf{x}_0 = \mathbf{x} \\ & \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0, \quad \forall t \in \{0, \dots, N-1\} \end{aligned} \quad (\mathcal{R}_N)$$

There are however no known methods of solving the above problem in its full generality. The hardness of the problem can be attributed to two observations; (i) optimizing directly over arbitrary measurable policies π_N in Π_N where Π_N is infinite dimensional is out the question; and (ii) distributionally robust constraints such as (14), even for convex loss functions L , are hard to deal with directly when $\mathbf{x}_t \in \mathcal{F}_t^y$ is a general non-linear function of the past measurements $(\mathbf{y}_0, \dots, \mathbf{y}_t)$. Hence, in what follows we restrict attention to control policies that are affine in the past disturbances as in [15]. Restricted policies of this type

⁴In the case that the initial state $\mathbf{x}_0 = \mathbf{x}$ is itself uncertain, one can always add an additional leading state $\mathbf{x}_{-1} = 0$ and a state update equation $\mathbf{x}_0 = A\mathbf{x}_{-1} + \mathbf{w}_{-1}$, where \mathbf{w}_{-1} equals \mathbf{x} in distribution.

are well known in the operations research and control community, where they are commonly referred to either as *linear decision rules* [3] or *affine feedback policies* [16]. Although such policies are typically suboptimal, recent research effort has focussed on providing suboptimality bounds when applied to systems with worst-case constraints [17], [18], [28].

Denote by $\mathbf{x} := (\mathbf{x}_0^\top, \dots, \mathbf{x}_N^\top)^\top$, $\mathbf{u} := (\mathbf{u}_0^\top, \dots, \mathbf{u}_{N-1}^\top)^\top$ and $\mathbf{y} := (\mathbf{y}_0^\top, \dots, \mathbf{y}_{N-1}^\top)^\top$ the collection of states, inputs and measurements, respectively, over the given finite horizon. Similarly define a vector of disturbances as

$$\mathbf{w} := (1, \mathbf{w}_0^\top, \dots, \mathbf{w}_{N-1}^\top)^\top \quad (16)$$

augmented with a leading one. This leading term is included for notational convenience so that any affine function of $(\mathbf{w}_0, \dots, \mathbf{w}_{N-1})$ can be written as $X\mathbf{w}$ for some matrix X with appropriate dimensions. Because of the w.s.s. condition on the disturbance process in Assumption II.1, we have that $\mathbb{E}_{\mathbb{P}^*}\{\mathbf{w} \cdot \mathbf{w}^\top\} = M_w \in \mathbb{S}_{++}^{Nd+1}$ with

$$M_w := (1, \mu^\top, \dots, \mu^\top)^\top (1, \mu^\top, \dots, \mu^\top) + \text{diag}(0, \mathbb{I}_N \otimes \Sigma).$$

The dynamics of the linear system \mathcal{S} over the finite horizon N can then be written as

$$\mathbf{x} = \mathcal{B}\mathbf{u} + \mathcal{C}\mathbf{w}, \quad \mathbf{y} = \mathcal{D}\mathbf{u} + \mathcal{E}\mathbf{w} \quad (17)$$

for some matrices $(\mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E})$ that can be derived from the system matrices and initial state $\mathbf{x}_0 = x$; see Appendix A. Note in particular that the leading one in (16) means that the term $\mathcal{C}\mathbf{w}$ is an affine function of both the disturbances and the initial state $\mathbf{x}_0 = x$. Our approach will be to restrict \mathbf{u} to be affine in the past disturbances, i.e., $\mathbf{u} = U\mathbf{w}$ for some causal feedback matrix $U \in \mathcal{N}$.

The set of causal policies \mathcal{N} must ensure that the resulting feedback policy \mathbf{u}_t is \mathcal{F}_t^y -measurable, i.e., that the feedback policy \mathbf{u}_t depends only on the initial state x and observed outputs $[\mathbf{y}_0, \dots, \mathbf{y}_t]$. This can be achieved by a reparametrization of the feedback policy $\mathbf{u} = \tilde{U}\boldsymbol{\eta}$ as an affine function of the *purified observations* $\boldsymbol{\eta} = (\mathcal{D}\mathcal{C} + \mathcal{E})\mathbf{w}$ as discussed in [3, §14.4.2]. The causality set can then be defined as

$$\mathcal{N} := \left\{ U \in \mathbb{R}^{N_x \times N_w} \left| U = \begin{pmatrix} \star & 0 & 0 & 0 \\ \star & \star & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \star & \star & \dots & \star \end{pmatrix} (\mathcal{D}\mathcal{C} + \mathcal{E}) \right. \right\}$$

which ensures that \mathbf{u}_t is \mathcal{F}_t^y -measurable. The \star indicates an arbitrary non-zero element in the matrix. Assume we have such an affine policy $\mathbf{u} = U\mathbf{w}$, then the cost of this policy according to the cost function (15) is

$$\tilde{J}_N(x, U) := \text{Tr} \{ U^\top (J_u + \mathcal{B}J_x\mathcal{B}) U M_w + 2\mathcal{C}J_x\mathcal{B}U M_w + \mathcal{C}^\top J_x\mathcal{C} M_w \}$$

where $J_x := \text{diag}(\text{diag}(\beta^0, \dots, \beta^{N-1}) \otimes Q, \beta^N Q_f)$ and $J_u := \text{diag}(\beta^0, \dots, \beta^{N-1}) \otimes R$. Note that $\tilde{J}_N(x, U)$ is convex quadratic in U since $\text{diag}(Q, R) \in \mathbb{S}_+$. We are now ready to state the main result of this section, which shows that finding the best affine control policy for problem \mathcal{R}_N can be reformulated as a tractable convex optimization problem.

Theorem II.3 (CVaR Constrained Control): The best admissible affine control policy of problem \mathcal{R}_N , i.e., a solution to the restricted problem

$$\begin{aligned} \inf_{U \in \mathcal{N}} \quad & \tilde{J}_N(x, U) \\ \text{s.t.} \quad & \mathbf{x} = \mathcal{B}\mathbf{u} + \mathcal{C}\mathbf{w}, \quad \mathbf{u} = U\mathbf{w} \\ & \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0, \quad \forall t \in \{0, \dots, N-1\} \end{aligned} \quad (\tilde{\mathcal{R}}_N)$$

where the loss function $L: \mathbb{R}^n \times \mathbb{R}_{++}^I \rightarrow \mathbb{R}$ satisfies Assumption II.2, can be found as a solution to the SDP

$$\begin{aligned} \inf \quad & \tilde{J}_N(x, U) \\ \text{s.t.} \quad & \left. \begin{aligned} & U \in \mathcal{N}, \beta_t \in \mathbb{R}, X_t \in \mathbb{S}_+^{Nd+2}, P_t^i \in \mathbb{S}_+^{Nd+1} \\ & \beta_t + \frac{1}{\epsilon} \text{Tr}\{M_w X_t\} \leq 0, \\ & X_t - \begin{pmatrix} \alpha_i P_t^i & \alpha_i (\mathcal{B}_t U + \mathcal{C}_t)^\top e_i \\ e_i^\top (\mathcal{B}_t U + \mathcal{C}_t) \alpha_i & \alpha_i e_i^0 - \beta_t \end{pmatrix} \succeq 0, \\ & \begin{pmatrix} P_t^i & (\mathcal{B}_t U + \mathcal{C}_t)^\top E_i^{\frac{1}{2}} \\ E_i^{\frac{1}{2}} (\mathcal{B}_t U + \mathcal{C}_t) & \mathbb{I}_n \end{pmatrix} \succeq 0, \end{aligned} \right\} \forall t, i \end{aligned} \quad (18)$$

where $\mathcal{B} =: (\mathcal{B}_0^\top, \dots, \mathcal{B}_{N-1}^\top)^\top$ and $\mathcal{C} =: (\mathcal{C}_0^\top, \dots, \mathcal{C}_{N-1}^\top)^\top$.

Proof: See Appendix A. \blacksquare

It has further been shown that such an approximation becomes *essentially exact* for a judicious choice of the scaling parameters $\alpha \in \mathbb{R}_{++}^n$; see [38, Thm. 3.6]. We remark again that [38, Thm. 3.6] ensures that there exists some $\alpha \in \mathbb{R}_{++}$ such that the constraints (14) reduce to a distributionally robust chance constraint for $\mathbb{X} = \{x \mid L^\alpha(x) < 0\}$, whenever \mathbf{x}_t is an affine function of the disturbances. In principle one could therefore identify such a parameter vector α to recover an exact representation of a robust chance constraint in the problem $\tilde{\mathcal{R}}_N$. However, simultaneous optimization over both U and α in (18) would result in a non-convex bi-affine optimization problem, and such problems are known to be \mathcal{NP} -hard in general.

While the result [38, Thm. 3.6] is only existential in nature, i.e., it is true for some unknown $\alpha \in \mathbb{R}_{++}$, the equivalence between chance constraints and CVaR constraints when \mathbb{X} is a simple ellipsoid or $I = 1$ is guaranteed. This result enables us to formulate the following corollary to Theorem II.3.

Corollary II.4 (Chance Constrained Control): The best admissible affine control policy of the restricted problem

$$\begin{aligned} \inf_{U \in \mathcal{N}} \quad & \tilde{J}_N(x, U) \\ \text{s.t.} \quad & \mathbf{x} = \mathcal{B}\mathbf{u} + \mathcal{C}\mathbf{w}, \quad \mathbf{u} = U\mathbf{w} \\ & \inf_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\{\mathbf{x}_t \in \mathbb{X}\} \geq 1 - \epsilon, \quad \forall t \in \{0, \dots, N-1\} \end{aligned}$$

where the constraint set $\mathbb{X} = \{x \mid x^\top E_1 x + 2e_1^\top x + e_1^0 \leq 0\}$ is a single ellipsoid, can be found as a solution of the SDP (18) with $I = 1$ and $\alpha_1 = 1$.

IV. INFINITE HORIZON DISTRIBUTIONALLY ROBUST CONTROL PROBLEMS

Infinite horizon control problems lend themselves to applications in which transient behaviour is of lesser importance, but in which we are interested in steady state behaviour. In Section IV we present a numerical example of such a problem in the context of wind turbine blade control. The problem setting is similar to the one presented in Section II, in that we again consider the DLTI system \mathcal{S} where the disturbance input process \mathbf{w}_t satisfies Assumption II.1. In addition, we assume that the disturbance \mathbf{w}_t has zero mean $\mu_{\mathbf{w}_t} = \mu = 0$, and a zero initial condition $\mathbf{x}_0 = 0$ reflects our indifference towards transient behaviour.

Hence in this infinite horizon setting, we consider the following optimal control problem:

$$\begin{aligned} \inf_{\pi \in \Pi_\infty} \quad & \sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbb{P}} \left\{ \sum_{t=0}^{N-1} [\mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t] \right\} \\ \text{s.t.} \quad & \mathbf{x}_{t+1} = \mathbf{A} \mathbf{x}_t + \mathbf{B} \mathbf{u}_t + \mathbf{C} \mathbf{w}_t \\ & \sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0 \end{aligned} \quad (\mathcal{R}_\infty)$$

where the loss function $L^\alpha(x)$ satisfies again Assumption II.2. We assume throughout this section that the pairs $(Q^{1/2}, A)$ and (C, A) are observable and that the pair (A, B) is stabilizable, which is sufficient to guarantee the existence of linear time-invariant exponentially stabilizing control policies. The feedback policies π are restricted to Π_∞ , where Π_∞ is the set of all linear time-invariant and causal (\mathcal{F}_t^y -measurable) feedback policies. We restricted attention to linear control strategies for the same reasons mentioned in Section II. It is also well known that such a restriction causes no loss of optimality when the distributionally robust constraint in \mathcal{R}_∞ is disregarded [20]. Indeed, the classical linear quadratic Gaussian (LQG) controller is optimal for the unconstrained version of problem \mathcal{R}_∞ .

The cost function in \mathcal{R}_∞ is the infinite horizon limit of the stage cost function in (15) for system \mathcal{S} , with no discounting or terminal cost. By omitting the discounting factor, the cost of a control law π becomes independent of the initial condition \mathbf{x}_0 reflecting an indifference towards the cost of transient behavior. The design goal in this case reduces to minimizing the average stage cost, so that the objective function becomes

$$\begin{aligned} J_\infty(\pi) &:= \sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathbb{P}} \left\{ \sum_{t=0}^{N-1} [\mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t] \right\} \\ &= \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}_{\mathbb{P}} \{ \mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t \} \quad \forall \mathbb{P} \in \mathcal{P}_\infty. \end{aligned}$$

where the equality follows from the fact that the expectation of a quadratic cost is independent of $\mathbb{P} \in \mathcal{P}_\infty$ for linear control policies.

The robust constraint in \mathcal{R}_∞ can be seen as a distributionally robust version of the nominal requirement

$$\limsup_{t \rightarrow \infty} \mathbb{P}^*\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0. \quad (19)$$

This constraint expresses the design requirement that $\mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0$ when t tends to infinity. If a steady

state distribution exists, the constraint (19) can be read as a constraint on the steady state distribution of $\{\mathbf{x}_t\}$, i.e.,

$$\mathbb{P}^*\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_\infty)) \leq 0$$

where $\mathbf{x}_t \rightarrow \mathbf{x}_\infty$ in distribution for $t \rightarrow \infty$. However, Assumption II.1 is not sufficient to guarantee that \mathbf{x}_t converges in distribution to a steady state \mathbf{x}_∞ , hence we cannot treat (19) as a steady state constraint in general. Nevertheless, we observe that, although $\{\mathbf{x}_t\}$ need not converge in distribution, its first two moments are known to converge whenever π is a strictly stabilizing linear control law:

Theorem III.1 (Steady State Behaviour [20, Theorem 6.23]): Let the discrete-time stochastic process \mathbf{x}_t be the solution of the stochastic difference equation $\mathbf{x}_{t+1} = \bar{\mathbf{A}}\mathbf{x}_t + \bar{\mathbf{C}}\mathbf{w}_t$, where $\bar{\mathbf{A}} \in \mathbb{R}^{n \times n}$, $\bar{\mathbf{C}} \in \mathbb{R}^{n \times d}$ and \mathbf{w}_t has zero mean and satisfies Assumption II.1. Define the covariance matrix

$$C_{xx}(t) := \mathbb{E}_{\mathbb{P}^*} \left\{ [\mathbf{x}_t - \mathbb{E}_{\mathbb{P}^*} \{\mathbf{x}_t\}] \cdot [\mathbf{x}_t - \mathbb{E}_{\mathbb{P}^*} \{\mathbf{x}_t\}]^\top \right\}.$$

If $\bar{\mathbf{A}}$ is asymptotically stable then the asymptotic variance matrix $P_\infty := \lim_{t \rightarrow \infty} C_{xx}(t)$ exists and is the unique solution of the discrete Lyapunov equation $P_\infty = \lim_{t \rightarrow \infty} C_{xx}(t) = \bar{\mathbf{A}}P_\infty\bar{\mathbf{A}}^\top + \bar{\mathbf{C}}\bar{\mathbf{C}}^\top$.

Despite the possible lack of convergence in distribution of \mathbf{x}_t , Theorem III.1 will enable us to represent the distributionally robust constraint of problem \mathcal{R}_∞

$$\limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0, \quad \forall \mathbb{P} \in \mathcal{P}_\infty \quad (20)$$

as a tractable constraint on the linear control law π , provided that we can identify a dynamic counterpart to Theorem I.2.

Note that a direct application of Theorem I.2 to the constraint (20) is problematic. If one assumes momentarily that \mathbf{x}_∞ exists as done for instance in [32], then (20) could be reformulated as a distributionally robust constraint in the form

$$\mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_\infty)) \leq 0, \quad \forall \mathbb{Q} \in \mathcal{Q} \quad (21)$$

where $\mathcal{Q} := \{\mathbb{Q} \mid \mathbb{E}_{\mathbb{Q}}\{\mathbf{x}_\infty\} = 0, \mathbb{E}_{\mathbb{Q}}\{\mathbf{x}_\infty \cdot \mathbf{x}_\infty^\top\} = P_\infty\}$ a set of measures on the measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. One could then apply Theorem I.2 to produce a linear matrix inequality (LMI) representation of the constraint (21). However, not every measure in \mathcal{Q} is necessarily a steady state distribution obtainable as a limit distribution of $\{\mathbf{x}_t\}$. Hence, even if \mathbf{x}_∞ exists, a replacement of the infinite horizon constraint (20) with (21) is seemingly conservative.

However, in the finite-horizon case we have the following result:

Lemma III.2: Suppose that $\mathbf{x}_0 = 0$ and each \mathbf{x}_t is a linear function of $(\mathbf{w}_0, \dots, \mathbf{w}_{t-1})$ resulting from some linear control policy $\pi \in \Pi_\infty$. Define $\mathcal{Q}_t := \{\mathbb{Q} \mid \mathbb{E}_{\mathbb{Q}}\{\mathbf{x}\} = 0, \mathbb{E}_{\mathbb{Q}}\{\mathbf{x} \cdot \mathbf{x}^\top\} = C_{xx}(t)\}$ for each $t \in \mathbb{N}$. Then

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) = \sup_{\mathbb{Q} \in \mathcal{Q}_t} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}))$$

Proof: To prove the claim, according to definition (6) of the CVaR, it suffices to show that $\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}}\{g(\mathbf{x}_t)\} = \sup_{\mathbb{Q} \in \mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\{g(\mathbf{x})\}$ holds for any measurable function $g: \mathbb{R}^n \rightarrow \mathbb{R}$. Because \mathbf{x}_t is linear in the disturbances $\mathbf{x}_t = R \cdot (\mathbf{w}_0, \dots, \mathbf{w}_{t-1})$ for some $R \in \mathbb{R}^{n \times td}$ as $\mathbf{x}_0 = 0$, we can write

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}}\{g(R \cdot (\mathbf{w}_0, \dots, \mathbf{w}_{t-1}))\} = \sup_{\mathbb{Q} \in \mathcal{Q}_t} \mathbb{E}_{\mathbb{Q}}\{g(\mathbf{x})\}$$

The last equality can easily be shown using the general projection property [35, Theorem 1]. ■

If one implements a stabilizing linear control policy $\pi \in \Pi_\infty$ such that $P_\infty = \lim_{t \rightarrow \infty} Cx(t, t)$ can be shown to exist by application of Theorem III.1, then it follows in the limit from Lemma III.2 that

$$\limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) = \sup_{\mathbb{Q} \in \mathcal{Q}_\infty} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x})) \quad (22)$$

for $\mathcal{Q}_\infty = \{\mathbb{Q} \mid \mathbb{E}_{\mathbb{Q}}\{\mathbf{x}\} = 0, \mathbb{E}_{\mathbb{Q}}\{\mathbf{x} \cdot \mathbf{x}^\top\} = P_\infty\}$, since the worst-case CVaR is continuous in its moment information. However, the preceding worst-case CVaR bound, although tractable as indicated by Theorem I.2, could potentially lead to a conservative reformulation of the constraint of interest (20) since

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq \limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \quad (23)$$

and consequently

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq \sup_{\mathbb{Q} \in \mathcal{Q}_\infty} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x})).$$

This conservatism follows from the possibility that for each time t the distributions attaining the worst-case bound of the right-hand side of (23) may depend on t . In other words, the worst-case bound of the right-hand side is not obviously obtainable as the limit when t tends to infinity for some fixed distribution $\mathbb{P} \in \mathcal{P}_\infty$, similar as the situation discussed for condition (21). Fortunately, we can show that this is in fact not the case and that no conservatism is incurred.

Lemma III.3: Let the discrete-time stochastic process \mathbf{x}_t be the solution of the stochastic difference equation $\mathbf{x}_{t+1} = \bar{A}\mathbf{x}_t + \bar{C}\mathbf{w}_t$, where $\bar{A} \in \mathbb{R}^{n \times n}$, $\bar{C} \in \mathbb{R}^{n \times d}$ and \mathbf{w}_t has zero mean and satisfies Assumption II.1. If \bar{A} is asymptotically stable then

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) = \sup_{\mathbb{Q} \in \mathcal{Q}_\infty} \mathbb{Q}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x})).$$

Proof: See Appendix A. ■

The equivalence (12) in view of Lemma III.3 now provides a probabilistic interpretation to what otherwise could be considered an *ad hoc* covariance constraints

$$\lim_{t \rightarrow \infty} \text{Tr} \left\{ E^{\frac{1}{2}} \mathbb{E} \{ \mathbf{x}_t \mathbf{x}_t^\top \} E^{\frac{1}{2}} \right\} \leq \epsilon$$

as for instance discussed in [36]. Indeed, constraining the variance to be bounded using a trace norm can now be read as a distributionally robust probabilistic constraint on the state satisfying an ellipsoidal state constraint $\mathbb{X} = \{x \mid x^\top E x \leq 1\}$. In general we have the following counterpart to Theorem I.2.

Theorem III.4 (Tractability of Worst-Case CVaR for Linear Systems): Let the discrete-time stochastic process \mathbf{x}_t be the solution of the stochastic difference equation $\mathbf{x}_{t+1} = \bar{A}\mathbf{x}_t +$

$\bar{C}\mathbf{w}_t$, where $\bar{A} \in \mathbb{R}^{n \times n}$, $\bar{C} \in \mathbb{R}^{n \times d}$ and \mathbf{w}_t has zero mean and satisfies Assumption II.1. If \bar{A} is asymptotically stable then

$$\begin{aligned} \forall \mathbb{P} \in \mathcal{P}_\infty : \quad & \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq 0 \\ & \Updownarrow \\ \inf_{\beta, X} \quad & \beta + \frac{1}{\epsilon} \text{Tr} \{ \text{diag}(P_\infty, 1) \cdot X \} \leq 0 \\ \text{s.t.} \quad & X \in \mathbb{S}_+^{n+1}, \beta \in \mathbb{R}, \forall i \in \{1, \dots, I\} : \\ & X - \begin{pmatrix} \alpha_i E_i & \alpha_i e_i \\ \alpha_i e_i^\top & \alpha_i e_i^0 - \beta \end{pmatrix} \succeq 0 \end{aligned}$$

where $P_\infty = \bar{A}P_\infty\bar{A}^\top + \bar{C}\bar{C}^\top$ the stationary variance of the state.

Proof: The theorem is a direct consequence of Lemma III.3 combined with Theorem I.2. ■

We remark that the condition in Theorem III.4 offers an exact condition for constraints of the type (20) to hold under the disturbance Assumption II.1. When a tighter condition is required, one must either resort to nonlinear control laws π or assume additional information regarding the disturbance process. We next show that the equivalence (12) also implies that the optimal linear feedback law for problem \mathcal{R}_∞ has an order which equals the number of states n of system S , and is the combination of a Kalman filter and a static feedback gain.

Theorem III.5 (Optimal Linear Feedback Law): The optimal linear feedback law π^* of problem \mathcal{R}_∞ consists of a linear estimator-controller pair (S, K) and hence is of the form

$$\pi^* : \begin{cases} \hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t + B\mathbf{u}_t + S(\mathbf{y}_{t+1} - C(A\hat{\mathbf{x}}_t + B\mathbf{u}_t)) \\ \mathbf{u}_t = K\hat{\mathbf{x}}_t \end{cases} \quad (24)$$

with $S := YD^\top(DYD^\top + EE^\top)^{-1}$. The matrix Y is the unique positive definite solution of the discrete algebraic Riccati equation

$$Y = A(Y - YD^\top(DYD^\top + EE^\top)^{-1}DY)A^\top + CC^\top$$

which can be solved efficiently [1]. The static feedback matrix is given by $K = Z^*(P^*)^{-1}$, where $P^* \in \mathbb{S}_{++}^n$ and $Z^* \in \mathbb{R}^{m \times n}$ can be found as the optimal solution of the SDP

$$\begin{aligned} \inf \quad & \text{Tr} Q(\Sigma + P) + \text{Tr} R X \\ \text{s.t.} \quad & P \in \mathbb{S}_+^n, Z \in \mathbb{R}^{m \times n}, X \in \mathbb{S}_+^m \\ & \begin{pmatrix} X & Z \\ Z^\top & P \end{pmatrix} \succeq 0, e^0 + \frac{1}{\epsilon} \text{Tr} \{ E_0(\Sigma + P) \} \leq 0 \\ & \begin{pmatrix} P - APA^\top - BZA^\top - AZ^\top B^\top - \tilde{W} & BZ \\ Z^\top B^\top & P \end{pmatrix} \succeq 0 \end{aligned} \quad (25)$$

where $\tilde{W} := YD^\top(DYD^\top + EE^\top)^{-1}DY$ and $\Sigma = Y - \tilde{W}$. Since (24) can be decomposed into a Kalman estimator S and state feedback controller K , problem \mathcal{R}_∞ satisfies a separation or certainty equivalence principle.

Proof: See Appendix A. ■

The Kalman filter in Theorem III.5 depends only on the process and measurement noise characteristics and is independent of the distributionally robust constraint (20) and cost function J_∞ . Finding the optimal static feedback gain K requires only the solution of the tractable convex problem (25).

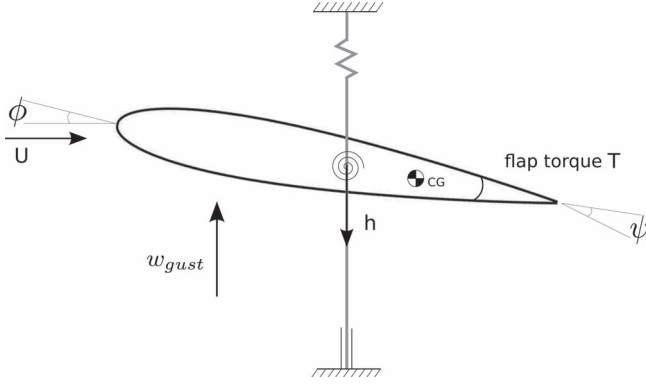


Fig. 1. Geometry of the 2-DOF structural model. The overall model is linear continuous time invariant and has a modest size of 13 states, one endogenous and exogenous input T and w_{gust} , respectively.

V. WIND TURBINE BLADE CONTROL DESIGN PROBLEM

To illustrate the method introduced in the preceding section, we consider a wind turbine control problem similar to the one introduced in [26]. As the size of wind turbines is increased for larger energy capture, they are subject to greater risks of fatigue failure and extreme loading events. Therefore, most large wind turbines today are equipped with pitch control for speed regulation, which can also be used for load alleviation.

However, these pitch actuators are slow and limited by the inertia of the blades. Hence, as in [26], we assume that the blades are equipped with an actively controlled flap. The control objective is to minimize actuation energy while keeping some measure of blade loading within specified bounds. The disturbance acting on the turbine blades is mostly due to atmospheric turbulence, for which little more than the frequency spectrum is known [11]. According to the standard design reference [24], atmospheric turbulence is typically treated as a Gaussian stochastic process defined by a standardized velocity spectrum. We follow the standard atmospheric turbulence model provided in [24], modulo the normality assumption which is not well supported in reality. Hence, this is a natural setting in which the ideas developed in this paper are of practical interest.

An aerofoil section with flap can be modeled using a simple two degree of freedom (2-DOF) plunge-pitch aerofoil, restrained by a pair of springs as shown in Fig. 1. The two dimensional aerofoil represents a cross section of one of the flexible wind turbine blades. For small elastic deformations and under the assumption of potential flow, we can use classical methods provided by [33] to describe the behaviour of our simple 2-DOF plunge-pitch aerofoil with a simple linear model. The modeling technique used here is by no means the only one possible, but results in a modest size plant of only ten states. An alternative technique using classical vortex-panel methods [19] to get higher fidelity, but still linear, models is presented in [26]. We note that the methods described in this paper are not limited by the modest size of our control model, as indicated in the follow up work [29] where a high fidelity model is considered.

Since the disturbance modeling is important to our approach, we discuss it in slightly more detail in the next subsection.

A. Disturbance Model

The majority of the disturbance acting upon the wind turbine blades is a direct result of atmospheric turbulence. Most

commonly, atmospheric turbulence is represented as the convolution of (Gaussian) white noise through a linear time-invariant (LTI) shaping filter, usually referred to as a *von Kármán* filter, see [11], [24]. Hence

$$w_{gust} := \mathcal{H} \cdot n_1$$

where n_1 is Gaussian white noise and \mathcal{H} the *von Kármán* filter, which we choose to be a proper stable rational filter as in [11] with state space representation

$$\begin{bmatrix} -7.701 & -7.008 & -1.404 & | & 1 \\ 1 & 0 & 0 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ \hline 1.447 & 7.022 & 1.533 & | & 0 \end{bmatrix}. \quad (\mathcal{H})$$

It is clear that the Gaussian assumption made on w_{gust} is unlikely to be fulfilled in practice, hence we assume only that n_1 is a scalar white w.s.s. noise process, i.e., $\mathbb{E}\{n_1^2(t)\} = 1$ and thus not necessarily Gaussian. Hence, in practice we need only estimate the power spectrum of the atmospheric turbulence w_{gust} , e.g., from historical data.

The overall system of the wind turbine blade model with additional flap and disturbance filter is a linear continuous time invariant system with 13 states, 10 states for the 2-DOF airfoil model and 3 states for the turbulence model. The overall model has one endogenous input T and one exogenous input n_1 . We assume that the states ϕ and h representing the pitch and plunge, see Fig. 1, are measured with negligible measurement noise, i.e.,

$$y = \begin{pmatrix} \phi \\ h \end{pmatrix} + \delta n_2$$

where n_2 is a zero mean white noise signal with unit covariance matrix, uncorrelated with n_1 . To fit in the framework provided in the paper, we discretize the continuous time model using the zero order hold method at sampling frequency $f_s = 100$ Hz which captures most of the salient plant dynamics for the model parameters we have selected.

B. Numerical Results

A natural control design criterion in this setting is to ensure that the vector $(\dot{\alpha}, \dot{h})$ is kept small in order to bound the fatigue stress, usually caused by high variance dynamic loads. In addition we would like extreme static loading events to be rare, corresponding to the requirement that the deformation vector (α, h) remains close to zero. We express these two design criteria, respectively, as

$$\inf_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P} \left\{ (\dot{\phi}(t), \dot{h}(t)) \in \mathcal{B}_2[55] \right\} \geq 1 - \epsilon \quad (26)$$

$$\inf_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P} \left\{ (\phi(t), h(t)) \in \mathcal{B}_2[6] \right\} \geq 1 - \epsilon \quad (27)$$

where $\epsilon = 0.1$, and $\mathcal{B}_n[r]$ denotes a closed ball in \mathbb{R}^n of radius r around the origin. The natural control objective in this setting is to minimize the expected actuation power usage. We express this by taking as a cost function

$$J(\pi) = \limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}} \left\{ \dot{\psi}^2(t) \right\}$$

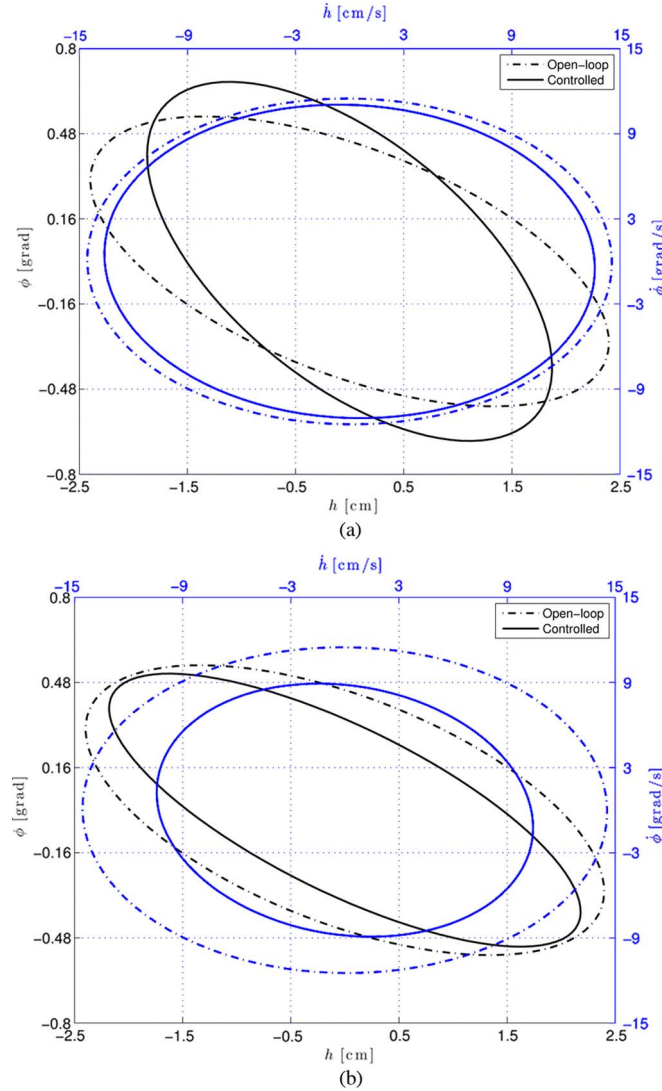


Fig. 2. (a) shows the variance of the vectors (ϕ, \mathbf{h}) and $(\dot{\phi}, \dot{\mathbf{h}})$ when uncontrolled and with the optimal controller according to Section III, as the sets $\{x \in \mathbb{R}^2 \mid x^\top \Sigma^{-1} x - 1 \leq 0\}$ with Σ the respective covariance matrix. Similarly, (b) shows the variance of the vectors (ϕ, \mathbf{h}) and $(\dot{\phi}, \dot{\mathbf{h}})$ when uncontrolled, and with the standard LQR controller $K_{\text{LQR}}(0.1)$.

which must be minimized subject to the fatigue and loading constraints (26) and (27) respectively. Using the method described in Section III, the optimal linear time invariant controller can be computed efficiently. Although it should be noted that in Theorem III.5 only one probability constraint is considered, the generalisation to the case of finitely many constraints of type (20) is straightforward and omitted here. The difference between the variance of the vectors (ϕ, \mathbf{h}) and $(\dot{\phi}, \dot{\mathbf{h}})$, when uncontrolled or controlled with the synthesized controller K^* , is visualized in Fig. 2(a).

We compare this controller to the standard \mathcal{H}_2 -optimal controller found by tuning the cost function

$$\limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{E}_{\mathbb{P}} \left\{ \gamma \dot{\mathbf{p}}^2(t) + \phi^2(t) + \dot{\phi}^2(t) + \mathbf{h}^2(t) + \dot{\mathbf{h}}^2(t) \right\}$$

which weighs the actuation energy versus the size of the states $(\phi, \dot{\phi}, \mathbf{h}, \dot{\mathbf{h}})$, according to the tuning factor γ . A naïve method

TABLE I
NUMERICAL RESULTS FOR THE WIND TURBINE BLADE CONTROL PROBLEM. THE THIRD AND FOURTH COLUMN SHOW THE WORST-CASE PROBABILITY THAT $(\phi, \mathbf{h}) \notin \mathcal{B}_2[6]$ AND $(\dot{\phi}, \dot{\mathbf{h}}) \notin \mathcal{B}_2[55]$, RESPECTIVELY

Control	J	$(\phi, \mathbf{h}) \notin \mathcal{B}_2[6]$	$(\dot{\phi}, \dot{\mathbf{h}}) \notin \mathcal{B}_2[55]$
Uncontrolled	0	0.16	0.12
K^*	82	0.10	0.10
$K_{\text{LQR}}(0.43)$	82	0.16	0.09
$K_{\text{LQR}}(0.1)$	425	0.15	0.07
$K_{\text{LQR}}(3.2 \times 10^{-3})$	3730	0.10	0.05

of designing a controller is to tune γ such that the closed loop system satisfies the fatigue (26) and loading (27) constraints.

We compare in Table I the cost of the optimal controller K^* and three naïvely tuned controllers $K_{\text{LQR}}(\gamma_i)$. First it is noted that when uncontrolled, the control cost is zero. However, since $\epsilon = 0.1$ both design specifications (26) and (27) are violated. The optimal controller K^* has satisfied (26) and (27) exactly with no conservatism and relatively low cost. The LQR controller $K_{\text{LQR}}(0.43)$ has the same cost as K^* but does not satisfy the constraints. The other LQR controllers either violate one of the constraints or have a massive cost compared to K^* . The difference between the variance of the vectors (ϕ, \mathbf{h}) and $(\dot{\phi}, \dot{\mathbf{h}})$, when uncontrolled or controlled with the controller $K_{\text{LQR}}(0.1)$, is visualized in Fig. 2(b).

It can be seen from this example that the methodology of Section III provides an easy procedure to design controllers that handle constraints of the type (26) and (27). Again we point out that, by dropping the Gaussian assumption on the stochastic process $(\mathbf{n}_1, \mathbf{n}_2)$, an assumption which in reality can not be justified anyway, the distributionally robust constraint formulation both makes practical sense and leads to a computationally tractable formulation.

VI. CONCLUSION

We investigate constrained control problems for stochastic linear systems when faced with the problem of only having limited information regarding the disturbance process, i.e., knowing only the first two moments of the disturbance distribution. We propose the use of distributionally robust chance and CVaR constraints to express constraint specifications when faced with distributional ambiguity.

These distributionally robust constrained formulations are subsequently used as control design specifications in both a finite horizon optimal control problem, and in an average cost optimal infinite horizon control problem.

We argue that these types of constraint formulations are practically meaningful and computationally tractable in the proposed finite and infinite horizon control design problems. We show that our results are tight, i.e., completely non-conservative given the information available. A natural conclusion is that substantially more structural information (e.g., higher order moments, continuity, support, etc.) about the uncertainty would be required to improve upon the results stated in the paper. The efficacy of the proposed formulation is illustrated for a wind turbine blade control design case study where flexibility issues play an important role and in which the distributionally robust framework makes practical sense.

APPENDIX

Define the matrices $\mathcal{B} \in \mathbb{R}^{N_x \times N_u}$, $\mathcal{C} \in \mathbb{R}^{N_x \times N_w}$, $\mathcal{D} \in \mathbb{R}^{N_y \times N_u}$ and $\mathcal{E} \in \mathbb{R}^{N_y \times N_w}$ as follows:

$$\begin{aligned} \mathcal{B} &:= \begin{pmatrix} \mathcal{B}_0 \\ \mathcal{B}_1 \\ \mathcal{B}_2 \\ \vdots \\ \mathcal{B}_N \end{pmatrix} = \begin{pmatrix} 0 & & & & \\ B & 0 & & & \\ AB & B & 0 & & \\ \vdots & & \ddots & \ddots & \\ \vdots & & & B & 0 \\ A^{N-1}B & A^{N-2}B & \dots & AB & B \end{pmatrix} \\ \mathcal{D} &:= \begin{pmatrix} \mathcal{D}_0 \\ \mathcal{D}_1 \\ \mathcal{D}_2 \\ \vdots \\ \mathcal{D}_{N-1} \end{pmatrix} = \begin{pmatrix} 0 & & & & \\ D & 0 & & & \\ & D & 0 & & \\ & & \ddots & \ddots & \\ & & & D & 0 \end{pmatrix} \\ \mathcal{C} &:= \begin{pmatrix} \mathcal{C}_0 \\ \mathcal{C}_1 \\ \mathcal{C}_2 \\ \vdots \\ \mathcal{C}_N \end{pmatrix} = \begin{pmatrix} x_0 & & & & \\ Ax_0 & C & & & \\ A^2x_0 & AC & C & & \\ \vdots & & \ddots & \ddots & \\ A^Nx_0 & A^{N-1}C & \dots & AC & C \end{pmatrix} \\ \mathcal{E} &:= \begin{pmatrix} \mathcal{E}_0 \\ \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_{N-1} \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ E & & & & \\ & E & & & \\ & & \ddots & & \\ & & & E & \end{pmatrix} \end{aligned}$$

where x_0 is the initial state of system \mathcal{S} , and $N_x := (N+1)n$, $N_u := Nm$, $N_w := Nd+1$ and $N_y = rN$.

Proof of Corollary I.3: For $I = 1$, $\alpha_1 = 1$ and $e_1 = \mu_x = 0$ Theorem I.2 implies

$$\begin{aligned} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L_1(\mathbf{x})) &= \inf_{\beta + \frac{1}{\epsilon} (\text{Tr}\{\Sigma_x Y\} + y_0)} \\ \text{s.t. } Y &\in \mathbb{S}_+^n, y \in \mathbb{R}^n, y_0 \in \mathbb{R}_+, \beta \in \mathbb{R} \\ \begin{pmatrix} Y & y \\ y^\top & y_0 \end{pmatrix} &\succeq 0, \begin{pmatrix} Y - E_1 & y \\ y^\top & y_0 - e_1^0 + \beta \end{pmatrix} \succeq 0. \end{aligned} \quad (28)$$

As $Y = E_1$, $y = 0$, $y_0 = 0$ and $\beta = e_1^0$ is feasible in (28), it is clear that the worst-case CVaR is bounded above by $e_1^0 + (1/\epsilon)\text{Tr}\{\Sigma_x E_1\}$. To prove the converse inequality, we let $(Y^*, y^*, y_0^*, \beta^*)$ be an optimal solution of (28). Then, we find

$$\begin{aligned} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{Q}\text{-CVaR}_\epsilon(L_1(\mathbf{x})) &= \beta^* + \frac{1}{\epsilon} (\text{Tr}\{\Sigma_x Y^*\} + y_0^*) \\ &\geq \beta^* + \frac{1}{\epsilon} \left(\text{Tr}\{\Sigma_x E_1\} + (e_1^0 - \beta^*)^+ \right) \\ &\geq e_1^0 + \frac{1}{\epsilon} \text{Tr}\{\Sigma_x E_1\} \end{aligned}$$

where the first inequality exploits the feasibility of $(Y^*, y^*, y_0^*, \beta^*)$ in (28), and the second inequality exploits the fact that $y_0^* \geq (e_1^0 - \beta^*)^+$ and $\epsilon \in (0, 1)$.

Proof of Theorem II.3: The proof follows by applying the tractability result in Theorem I.2 to the constraints

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L(\overbrace{(\mathcal{B}_t U + \mathcal{C}_t) \mathbf{w}}^{\mathbf{x}_t}; \alpha)) \leq 0.$$

Explicitly writing out the quadratic form in the preceding inequality as

$$\sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon \left(\max_i \alpha_i \mathbf{w}^\top (\mathcal{B}_t U + \mathcal{C}_t)^\top E_i (\mathcal{B}_t U + \mathcal{C}_t) \mathbf{w} + 2\alpha_i e_i^\top (\mathcal{B}_t U + \mathcal{C}_t) \mathbf{w} + \alpha_i e_i^0 \right) \leq 0$$

yields a matrix inequality with quadratic terms in the variable $U : \exists \beta_t \in \mathbb{R}, X_t \in \mathbb{S}_+^{Nd+2} : \beta_t + (1/\epsilon)\text{Tr}\{M_w X_t\} \leq 0$ and

$$X_t \succeq \begin{pmatrix} \alpha_i (\mathcal{B}_t U + \mathcal{C}_t)^\top E_i (\mathcal{B}_t U + \mathcal{C}_t) & (\mathcal{B}_t U + \mathcal{C}_t)^\top e_i \alpha_i \\ \alpha_i e_i^\top (\mathcal{B}_t U + \mathcal{C}_t) & e_i^0 \alpha_i - \beta_t \end{pmatrix}, \quad \forall i \in \{1, \dots, I\}.$$

The final result claimed in the theorem is then found by applying Schur complements, and rewriting the quadratic matrix inequality as two LMIs using the additional variables $P_t^i \in \mathbb{S}_+^{Nd+1}$.

Proof of Lemma III.3: Recalling (22), it is sufficient to prove that the inequality (23) actually holds with equality. In other words, it suffices to show that

$$\limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq \sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)). \quad (29)$$

Choose any $\delta > 0$ and $N' \in \mathbb{N}$. From the definitions of the limit superior and the supremum appearing in the left-hand side of (29), there exists some time instance $N \geq N' > 0$ and some probability measure $\tilde{\mathbb{P}}$ such that the left-hand side of (29) is bounded by

$$\limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \leq \tilde{\mathbb{P}}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_N)) + \delta. \quad (30)$$

Consider now the subsequence $\{\mathbf{x}_{kN}\}_{k=0}^\infty$, i.e., the subsequence obtained by taking every N th element of the sequence $\{\mathbf{x}_t\}$ beginning from \mathbf{x}_0 . The elements of this subsequence are related by

$$\mathbf{x}_{(k+1)N} = \bar{A}^N \mathbf{x}_{kN} + [\bar{A}^{N-1} \bar{C}, \dots, \bar{A} \bar{C}, \bar{C}] \Delta \mathbf{w}_k, \quad k \in \mathbb{N}$$

where each $\Delta \mathbf{w}_k := (\mathbf{w}_{kN}; \dots; \mathbf{w}_{(k+1)N-1})$ is a collection of disturbances N steps long.

Consider the distribution of $\Delta \mathbf{w}_0$ under the measure $\tilde{\mathbb{P}}$, i.e., the marginal distribution of the first N elements of the disturbance sequence $\{\mathbf{w}_t\}$ under $\tilde{\mathbb{P}}$. Construct a probability measure $\mathbb{P}' \in \mathcal{P}_\infty$ such that the subsequences $\Delta \mathbf{w}_k$ are independent and identically distributed (i.i.d.) and such that $\Delta \mathbf{w}_0$ has the same distribution under both \mathbb{P}' and $\tilde{\mathbb{P}}$. The marginal distribution \mathbf{x}_N will then likewise be the same under both \mathbb{P}' and $\tilde{\mathbb{P}}$. Consequently, we must also have

$$\tilde{\mathbb{P}}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_N)) = \mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_N)).$$

The loss function L^α satisfies $L^\alpha(x+y) \geq L^\alpha(x) + \partial L^\alpha(x)^\top y$ with ∂L^α a subgradient of L^α . Since CVaR is monotone, we obtain the inequality

$$\begin{aligned} \mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_{(k+1)N})) &= \mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha(\bar{A}^N \mathbf{x}_{kN} + [\bar{A}^{N-1} \bar{C}, \dots, \bar{A} \bar{C}, \bar{C}] \Delta \mathbf{w}_k)) \\ &\geq \mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha([\bar{A}^{N-1} \bar{C}, \dots, \bar{A} \bar{C}, \bar{C}] \Delta \mathbf{w}_k) \\ &\quad + \partial L^\alpha([\bar{A}^{N-1} \bar{C}, \dots, \bar{A} \bar{C}, \bar{C}] \Delta \mathbf{w}_k)^\top \\ &\quad \times \bar{A}^N \mathbf{x}_{kN}). \end{aligned}$$

From the definition of the CVaR given in (6) and the inequality $(a+b)^+ \geq (a)^+ - |b|$ we can then conclude that $\mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_{(k+1)N}))$ is larger or equal to

$$\mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha([\bar{A}^{N-1}\bar{C}, \dots, \bar{A}\bar{C}, \bar{C}]\Delta\mathbf{w}_k)) - \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}'}\left\{\left|\partial L^\alpha([\bar{A}^{N-1}\bar{C}, \dots, \bar{A}\bar{C}, \bar{C}]\Delta\mathbf{w}_k)^\top \bar{A}^N \mathbf{x}_{kN}\right|\right\}.$$

For every $\delta > 0$ there exists p such that $|x| \leq \epsilon\delta + px^2$. Since the disturbance subsequences $\Delta\mathbf{w}_k$ are assumed i.i.d. with $\mathbf{x}_0 = 0$, the random variable \mathbf{x}_N has the same distribution as $[\bar{A}^{N-1}\bar{C}, \dots, \bar{C}]\Delta\mathbf{w}_k$ for all k . Using the fact that \mathbf{x}_{kN} is independent from $\Delta\mathbf{w}_k$ we then obtain

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}'}\left\{\left|\partial L^\alpha([\bar{A}^{N-1}\bar{C}, \dots, \bar{A}\bar{C}, \bar{C}]\Delta\mathbf{w}_k)^\top (\bar{A}^N)^\top \mathbf{x}_{kN}\right|\right\} - \epsilon\delta \\ & \leq p\mathbb{E}_{\mathbb{P}'}\left\{\left(\partial L^\alpha(\mathbf{x}_N)^\top \bar{A}^N \mathbf{x}_{kN}\right)^\top \cdot \left(\partial L^\alpha(\mathbf{x}_N)^\top \bar{A}^N \mathbf{x}_{kN}\right)\right\} \\ & \leq p\text{Tr}\left\{\mathbb{E}_{\mathbb{P}'}\left\{\partial L^\alpha(\mathbf{x}_N)\partial L^\alpha(\mathbf{x}_N)^\top\right\} \right. \\ & \quad \left. \times \bar{A}^N \mathbb{E}_{\mathbb{P}'}\left\{\mathbf{x}_{kN}\mathbf{x}_{kN}^\top\right\} (\bar{A}^N)^\top\right\} \end{aligned}$$

using additionally that $C_{xx}(kN) \preceq P_\infty$ for all k . Since N' could be chosen arbitrarily large, we now assume that $N \geq N' > 0$ is large enough that

$$p\text{Tr}\left\{\mathbb{E}_{\mathbb{P}'}\left\{\partial L^\alpha(\mathbf{x}_N)\partial L^\alpha(\mathbf{x}_N)^\top\right\} \bar{A}^N P_\infty (\bar{A}^N)^\top\right\} \leq \epsilon\delta$$

which is always possible when \bar{A} is asymptotically stable and because

$$\mathbb{E}_{\mathbb{P}'}\left\{\partial L^\alpha(\mathbf{x}_N)\partial L^\alpha(\mathbf{x}_N)^\top\right\} \preceq E \mathbb{E}_{\mathbb{P}'}\left\{\mathbf{x}_N \cdot \mathbf{x}_N^\top\right\} E^\top + e \preceq \infty$$

for any $E \succeq \alpha_i E_i$ and $e \succeq \alpha_i e_i \cdot e_i^\top$ for all $i \in \{1, \dots, I\}$ is bounded from above uniformly in N . It then follows that:

$$\begin{aligned} & \mathbb{P}'\text{-CVaR}_\epsilon(L_\alpha(\mathbf{x}_{(k+1)N})) + 2\delta \\ & \geq \mathbb{P}'\text{-CVaR}_\epsilon(L_\alpha(\mathbf{x}_N)) \\ & = \tilde{\mathbb{P}}\text{-CVaR}_\epsilon(L_\alpha(\mathbf{x}_N)), \quad \forall k. \end{aligned} \quad (31)$$

Combining the preceding inequalities (30) and (31), we obtain the inequality

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}_\infty} \mathbb{P}\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) \\ & \leq \limsup_{t \rightarrow \infty} \mathbb{P}'\text{-CVaR}_\epsilon(L^\alpha(\mathbf{x}_t)) + 3\delta. \end{aligned}$$

Since $\delta > 0$ could be chosen to be arbitrarily small, (29) immediately follows and the proof is complete.

Proof of Theorem III.5: We have according to Theorem III.3 and Corollary I.3 the equivalences $\sup_{\mathbb{P} \in \mathcal{P}_\infty} \limsup_{t \rightarrow \infty} \mathbb{P}\text{-CVaR}_\epsilon(L_1(\mathbf{x}_t)) \leq 0$

$$\begin{aligned} & \iff \sup_{\mathbb{Q} \in \mathcal{Q}_\infty} \mathbb{Q}\text{-CVaR}_\epsilon(L_1(\mathbf{x})) \leq 0. \\ & \iff \limsup_{t \rightarrow \infty} \text{Tr}\left\{E_0^{\frac{1}{2}} \mathbb{E}\left\{\mathbf{x}_t \mathbf{x}_t^\top\right\} E_0^{\frac{1}{2}}\right\} \leq -e_0 \epsilon. \end{aligned}$$

when the closed loop system is stable. Notice that this without loss of generality as an unstable system would yield an unbounded cost J_∞ and hence can be discarded.

The following inequalities:

$$\begin{aligned} & \liminf_{t \rightarrow \infty} \mathbb{E}_{\mathbb{P}^*}\left\{\mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t\right\} \leq J_\infty(\pi) \\ & J_\infty(\pi) \leq \limsup_{t \rightarrow \infty} \mathbb{E}_{\mathbb{P}^*}\left\{\mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t\right\}, \quad \forall \mathbb{P} \in \mathcal{P}_\infty \end{aligned}$$

follow immediately from the definition of limit inferior and limit superior, respectively. The objective function now can be written in the form of a standard \mathcal{H}_2 -problem

$$J_{\text{IQR}} = \lim_{t \rightarrow \infty} \text{Tr}\left\{Q^{\frac{1}{2}} \mathbb{E}_{\mathbb{P}^*}\left\{\mathbf{x}_t \mathbf{x}_t^\top\right\} Q^{\frac{1}{2}} + R^{\frac{1}{2}} \mathbb{E}_{\mathbb{P}^*}\left\{\mathbf{u}_t \mathbf{u}_t^\top\right\} R^{\frac{1}{2}}\right\}$$

using the fact that the expectation operator is linear and the trace identity $\text{Tr}\{AB\} = \text{Tr}\{BA\}$ and $\mathbb{E}_{\mathbb{P}}\{\mathbf{x}_t \mathbf{x}_t^\top\}$ converges for $t \rightarrow \infty$. Hence, when restricted to linear control strategies, problem \mathcal{R}_∞ reduces to

$$\begin{aligned} & \inf_{\pi} \quad \lim_{t \rightarrow \infty} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{x}_t^\top \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t\right\} \\ & \text{s.t.} \quad \mathbf{x}_{t+1} = \mathbf{A} \mathbf{x}_t + \mathbf{B} \mathbf{u}_t + \mathbf{C} \mathbf{w}_t \\ & \quad \lim_{t \rightarrow \infty} \text{Tr}\left\{E^{\frac{1}{2}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{x}_t \mathbf{x}_t^\top\right\} E^{\frac{1}{2}}\right\} \leq -e_0 \epsilon. \end{aligned}$$

However, the last problem is an instance of a standard multi-criterion \mathcal{H}_2 -problem, see [6, §12.2.1]. The fact that the optimal control law is of the form (24) is a result of the fact that it solves an \mathcal{H}_2 -problem with a different cost measure, i.e., there exists an unconstrained \mathcal{H}_2 -problem with state and input penalty matrices $\tilde{\mathbf{Q}}, \tilde{\mathbf{R}}$ for which the solution satisfies the omitted trace constraint [6, §6.5.1]. The fact that K can be found as the solution to an SDP can be found in [7], and essentially follows from standard LMI manipulations.

ACKNOWLEDGMENT

The authors would like to thank B. F. Ng and R. Palacios for introducing the gust load alleviation control problem and providing them with an appropriate model.

REFERENCES

- [1] W. F. Arnold, III and A. J. Laub, "Generalized eigenproblem algorithms and software for algebraic Riccati equations," *Proc. IEEE*, vol. 72, no. 12, pp. 1746–1754, 1984.
- [2] B. R. Barmish and C. M. Lagoa, "The uniform distribution: A rigorous justification for its use in robustness analysis," *Math. Control, Signals, Syst.*, vol. 10, no. 3, pp. 203–222, 1997.
- [3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [4] D. P. Bertsekas and I. B. Rhodes, "On the minimax reachability of target sets and target tubes," *Automatica*, vol. 7, no. 2, pp. 233–247, 1971.
- [5] F. Blanchini, "Set invariance in control," *Automatica*, vol. 35, no. 11, pp. 1747–1767, 1999.
- [6] S. P. Boyd and C. H. Barratt, *Linear Controller Design: Limits of Performance*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1991.
- [7] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA, USA: SIAM, 1994.
- [8] G. C. Calafiore, "Random convex programs," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 2010.
- [9] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, 2006.
- [10] G. C. Calafiore and L. El Ghaoui, "On distributionally robust chance-constrained linear programs," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1–22, 2006.

- [11] C. W. Campbell, "Monte Carlo turbulence simulation using rational approximations to von Kármán spectra," *AIAA J.*, vol. 24, no. 1, pp. 62–66, 1986.
- [12] M. Cannon, B. Kouvaritakis, and X. Wu, "Probabilistic constrained MPC for multiplicative and additive stochastic uncertainty," *IEEE Trans. Autom. Control*, vol. 54, no. 7, pp. 1626–1632, 2009.
- [13] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Manag. Sci.*, vol. 6, no. 1, pp. 73–79, 1959.
- [14] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From CVaR to uncertainty set: Implications in joint chance-constrained optimization," *Oper. Res.*, vol. 58, no. 2, pp. 470–485, 2010.
- [15] P. J. Goulart and E. C. Kerrigan, "Output feedback receding horizon control of constrained systems," *Int. J. Control*, vol. 80, no. 1, pp. 8–20, 2007.
- [16] P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, "Optimization over state feedback policies for robust control with constraints," *Automatica*, vol. 42, no. 4, pp. 523–533, 2006.
- [17] M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn, "An efficient method to estimate the suboptimality of affine controllers," *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2841–2853, 2011.
- [18] D. A. Iancu, M. Sharma, and M. Sviridenko, "Supermodularity and affine policies in dynamic robust optimization," *Oper. Res.*, 2013. [Online]. Available: http://web.stanford.edu/~daniiancu/Papers/Journal/supermod_robust.pdf
- [19] J. Katz and A. Plotkin, *Low-Speed Aerodynamics*, vol. 13, ser. *Cambridge Aerospace Series*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [20] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, vol. 172. New York, NY, USA: Wiley-Interscience, 1972.
- [21] C. M. Lagoa and B. R. Barmish, "Distributionally robust Monte Carlo simulation: A tutorial survey," in *Proc. 15th IFAC World Congr.*, 2002, pp. 358–358.
- [22] C. M. Lagoa, P. S. Shcherbakov, and B. R. Barmish, "Probabilistic enhancement of classical robustness margins: The unrectangularity concept," *Syst. Control Lett.*, vol. 35, no. 1, pp. 31–43, 1998.
- [23] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
- [24] "U.S. Military Handbook," MIL-HDBK-1797, 1997.
- [25] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2006.
- [26] B. F. Ng, R. Palacios, J. M. R. Graham, and E. C. Kerrigan, "Robust control synthesis for gust load alleviation from large aeroelastic models with relaxation of spatial discretisation," in *Proc. EWEA*, 2012. [Online]. Available: http://www.researchgate.net/publication/261948088_Robust_Control_Synthesis_for_Gust_Load_Alleviation_from_Large_Aeroelastic_Models_with_Relaxation_of_Spatial_Discretisation
- [27] B. P. Van Parys, P. J. Goulart, and D. Kuhn, "Generalized gauss inequalities via semidefinite programming," *Math. Programming*, 2015. [Online]. Available: http://www.optimization-online.org/DB_FILE/2014/01/4187.pdf
- [28] B. P. Van Parys, P. J. Goulart, and M. Morari, "Infinite horizon performance bounds for uncertain constrained systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2803–2817, Nov. 2013.
- [29] B. P. Van Parys, B. F. Ng, P. J. Goulart, and R. Palacios, "Optimal control for load alleviation in wind turbines," in *Proc. 32nd ASME Wind Energy Symp.*, 2014. [Online]. Available: <http://arc.aiaa.org/doi/abs/10.2514/6.2014-1222>
- [30] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.
- [31] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, pp. 21–42, 2000.
- [32] G. Schildbach, P. J. Goulart, and M. Morari, "Linear controller design for chance constrained systems," *Automatica*, vol. 51, pp. 276–284, Jan. 2015.
- [33] T. Theodorsen, "General theory of aerodynamic instability and the mechanism of flutter," National Advisory Committee Aeronautics, Cranfield, U.K., TR-496, 1936.
- [34] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [35] Y. L. Yu, Y. Li, D. Schuurmans, and C. Szepesvári, "A general projection property for distribution families," in *Proc. 23rd Annu. Conf. Adv. Neural Inform. Process. Syst.*, 2009, pp. 2232–2240.
- [36] G. Zhu, M. A. Rotea, and R. Skelton, "A convergent algorithm for the output covariance constraint control problem," *SIAM J. Control Optim.*, vol. 35, no. 1, pp. 341–361, 1997.
- [37] S. Zhu and M. Fukushima, "Worst-case conditional value-at-risk with application to robust portfolio management," *Oper. Res.*, vol. 57, no. 5, pp. 1155–1168, 2009.
- [38] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, vol. 137, no. 1–2, pp. 167–198, 2013.



Bart P. G. Van Parys was born in Leuven, Belgium. He received the B.S. degree in electrical engineering and the M.S. degree in applied and engineering mathematics from the University of Leuven, Leuven, Belgium, and is currently pursuing the Ph.D. degree at the Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland.

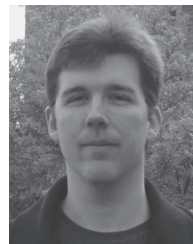
His current research interests are situated around constrained control of uncertain systems. In particular the interplay between constrained control and convex optimization is a recurring theme throughout his work. He also has a keen interest in infinite dimensional optimization problems and their impact on probability and decision theory.



Daniel Kuhn received the M.Sc. degree in theoretical physics from ETH Zurich, Zurich, Switzerland and the Ph.D. degree in economics from University of St. Gallen, St. Gallen, Switzerland.

He is an Associate Professor at the College of Management of Technology, EPFL, Lausanne, Switzerland, where he holds the Chair of Risk Analytics and Optimization (RAO). Before joining EPFL, Daniel Kuhn was a faculty member in the Department of Computing at Imperial College London (2007–2013) and a postdoctoral research associate in the Department of Management Science and Engineering at Stanford University (2005–2006). He serves on the editorial boards of several academic journals including *Energy Systems*, *Operations Research*, *Management Science*, *Mathematical Programming*, and *Operations Research Letters*. His current research interests are focused on the modeling of uncertainty, the development of efficient computational methods for the solution of stochastic and robust optimization problems and the design of approximation schemes that ensure their computational tractability. This work is primarily application-driven, the main application areas being energy systems, operations management and engineering.

Dr. Kuhn is a member of the Steering Committee, Stochastic Programming Society (COSP).



Paul J. Goulart (M'04) received the B.Sc. and M.Sc. degrees in aeronautics and astronautics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, and the Ph.D. degree in control engineering from the University of Cambridge, Cambridge, U.K., in 2007.

From 2007 to 2011, he was a Lecturer in control systems in the Department of Aeronautics, Imperial College London, and from 2011 to 2014 a Senior Researcher in the Automatic Control Laboratory, ETH Zurich. He is currently an Associate Professor in the Department of Engineering Science and Tutorial Fellow, St. Edmund Hall, University of Oxford, Oxford, U.K. His research interests include model predictive control, robust optimization, and control of fluid flows.

Dr. Goulart was selected as a Gates Scholar at the University of Cambridge.



Manfred Morari (F'05) received the diploma from ETH Zurich and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, both in chemical engineering.

He was head of the Department of Information Technology and Electrical Engineering, ETH Zurich, from 2009 to 2012 and head of the Automatic Control Laboratory from 1994 to 2008. Before that he was the McCollum-Corcoran Professor of Chemical Engineering and Executive Officer for Control and Dynamical Systems at the California Institute of Technology, Pasadena. From 1977 to 1983 he was on the faculty of the University of Wisconsin, Madison. His interests are in hybrid systems and the control of biomedical systems. Morari's research is internationally recognized. The analysis techniques and software developed in his group are used in universities and industry throughout the world.

Dr. Morari received the Eckman Award, Ragazzini Award and Bellman Control Heritage Award from the American Automatic Control Council; the Colburn Award, Professional Progress Award and CAST Division Award from the American Institute of Chemical Engineers; the Control Systems Technical Field Award and the Bode Lecture Prize from IEEE. He is a Fellow of AIChE and IFAC. In 1993 he was elected to the U.S. National Academy of Engineering. Manfred Morari served on the technical advisory boards of several major corporations.