Automatica xxx (xxxx) xxx

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Ambiguity tube MPC☆

Fan Wu [a], Mario Eduardo Villanueva [a,b], Boris Houska [a,*]

[a] ShanghaiTech University, China
[b] IMT School for Advanced Studies Lucca, Italy

ARTICLE INFO

Article history:
Received 2 March 2021
Received in revised form 5 May 2022
Accepted 21 August 2022
Available online xxxx

Keywords:
Stochastic control
Model predictive control
Stability analysis
Markov processes
Martingale theory

ABSTRACT

This paper is about a class of distributionally robust model predictive controllers (MPC) for nonlinear stochastic processes, which evaluate risk and control performance measures by propagating ambiguity sets in the space of state probability measures. A framework for formulating such ambiguity tube MPC controllers is presented using methods from the field of optimal transport theory. Moreover, an analysis technique based on supermartingales is proposed, leading to stochastic stability results for a large class of distributionally robust controllers. In this context, we also discuss how to construct terminal cost functions for stochastic and distributionally robust MPC that ensure closed-loop stability and asymptotic convergence to robust invariant sets. The corresponding theoretical developments are illustrated by tutorial-style examples and a numerical case study.

1. Introduction

Traditional robust MPC formulations that systematically take model uncertainties and external disturbances into account can be categorized into two classes. The first class of robust MPC controllers are based on min–max (Houska & Villanueva, 2019; Rawlings, Mayne, & Diehl, 2018) or tube-based MPC formulations (Langson, Chryssochoos, Raković, & Mayne, 2004; Mayne, Seron, & Raković, 2005; Raković, Kouvaritakis, Findeisen, & Cannon, 2012), which typically assume that worst-case bounds on the uncertainty are available. This is in contrast to the second class of optimization based robust controllers, namely, stochastic MPC controllers (Kouvaritakis & Cannon, 2015; Mesbah, 2016), which assume that the probability distribution of the external disturbance is known. The main practical difference between these formulations is that most stochastic MPC controllers attempt to either bound or penalize the probability of a constraint violation, but, in contrast to min–max MPC formulations, conservative worst-case constraints are not enforced.

In terms of recent developments in the field of robust MPC, several attempts have been made to unify the above classes by considering distributionally robust MPC controllers (Van Parys, Kuhn, Goulart, & Morari, 2016). Here, one assumes that the uncertainty is stochastic, but the associated probability distribution is only known to be in a given ambiguity set. Thus, in the most general setting, distributionally robust MPC formulations contain both traditional stochastic MPC as well as min–max MPC as special cases: in the context of stochastic MPC the ambiguity set is a singleton whereas min–max MPC is based on ambiguity sets that contain all uncertainty distributions with a given bounded support. Notice that modern distributionally robust MPC formulations are often formulated by using risk measures (Sopasakis, Herceg, Bemporad, & Patrinos, 2019). This trend is motivated by the availability of rather general classes of coherent – and, most importantly, computationally tractable – risk measures, such as the conditional value at risk (Rockafellar & Uryasev, 2013).

This paper focuses on distributionally robust MPC problems, formulated as ambiguity controllers. Here, the main idea is to propagate sets in the space of probability measures on the state space. The primary motivation for analyzing such a class of controllers is, however, not to develop yet another robust MPC formulation, but to develop a coherent stability theory for a very general class of distributionally robust MPC controllers, containing tube MPC as well as stochastic MPC as a special case.

Before we outline why such a general stability theory for ambiguity controllers is of fundamental interest in control theory – especially, in the emerging era of learning based MPC (Hewing, Wabersich, Menner, & Zeilinger, 2020; Zanon & Gros, 2021), where uncertain models are omni-present – one has to first mention the existence of a plethora of stability results for MPC. First of all, the stability of classical (certainty-equivalent) MPC has been thoroughly analyzed—be it for tracking or economic MPC, with or without terminal costs or regions (Chen & Allgöwer, 1998; Grüne, 2009; Rawlings et al., 2018). Similarly, the stability of variants of min–max MPC schemes have been analyzed exhaustively (Mayne et al., 2005; Villanueva, Quirynen, Diehl, Chachuat, & Houska,

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editors Gabriele Pannocchia, Daniele Casagrande under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail address: borish@shanghaitech.edu.cn (B. Houska).

https://doi.org/10.1016/j.automatica.2022.110648
0005-1098/© 2022 Elsevier Ltd. All rights reserved.

2017), although the development of a unified stability analysis for general set-based MPC controllers is a topic of ongoing research (Villanueva, De Lazzari, Müller, & Houska, 2020).

These stability results for certainty-equivalent and tube MPC controllers have in common that they rely on the construction of Lyapunov functions that descend along the closed-loop trajectories of the robust controller. This is in contrast to existing results on the stability of stochastic MPC, which are usually based on the theory of non-negative supermartingales (Doob, 1953; Feller, 1971). The mathematical foundation for such results has been developed by Bucy (1965) and Kushner (1965) in the context of general Markov processes—see also Kushner (2014) for a historical review. At the current status of research on stochastic MPC, martingale theory has been applied to special classes of linear MPC controllers with multiplicative uncertainty (Bernardini & Bemporad, 2012). Moreover, an impressive collection of articles by M. Cannon and B. Kouvaritakis has appeared during the last two decades, which has had significant impact on shaping the state-of-the-art of stochastic MPC. As we cannot possibly list all of their papers on this subject, we refer at this point to their textbook (Kouvaritakis & Cannon, 2015) for an overview of formulations and stability results for stochastic linear MPC. Additionally, the book chapter (Kouvaritakis & Cannon, 2016) comes along with an excellent overview of recursive feasibility results for stochastic MPC as well as a proof of stochastic stability with respect to ellipsoidal regions that are derived by using non-negative supermartingales, too.

Given the above list of articles it can certainly be stated that the question how to establish stability results for both robust mix-max and stochastic MPC has received significant attention. Nevertheless, looking back at the MPC literature from the last decade, it must also be stated that this question has raised a critical discussion. For example, the general critique of robust MPC in Mayne (2015) points out the lack of a satisfying treatment of stabilizing terminal conditions for stochastic MPC. Similarly, Chatterjee and Lygeros (2015) discusses various discrepancies between deterministic and stochastic MPC. From these articles it does get clear that one has to carefully distinguish between a rigorous stability analysis based on supermartigale theory – in the sense of Kushner and Bucy – and weaker properties of stochastic MPC from which stability may not be inferred. Among these are bounds on the asymptotic average performance of stochastic MPC (Kouvaritakis & Cannon, 2015, 2016), which do not necessarily imply stability. Moreover, in the past 5 years several articles have appeared exploiting input-to-state-stability assumptions for establishing convergence of stochastic MPC controllers to robust invariant sets (Lorenzen, Dabbene, Tempo, & Allgöwer, 2016; Sehr & Bitmead, 2018). One of the strongest results in this context appeared only a few weeks ago in Munoz-Carpintero and Cannon (2020), where an input-to-state-stability assumption in combination with the Borel–Cantelli lemma is used to establish conditions under which the state of a potentially nonlinear Markov process converges almost surely to a minimal robust invariant set. These conditions are applicable for establishing convergence of a variety of stochastic MPC formulations. Nevertheless, it has to be recalled here that, in general, neither stability implies convergence nor convergence implies stability. As such, none of the these contributions proposes a completely satisfying answer to the question how asymptotic stability conditions can be established for general stochastic, let alone distributionally robust, MPC.

Contribution. This paper is concerned with the mathematical formulation and stochastic stability analysis of distributionally robust MPC controllers for general, potentially nonlinear, but Lipschitz continuous stochastic discrete-time systems. Here, the focus is on ambiguity tube MPC controllers that are based on the propagation of sets in the space of state probability measures. The corresponding contributions of the current article can be outlined as follows.

(1) Section 2 develops a novel framework for formulating ambiguity tube MPC problems by exploiting measure-theoretic concepts from the field of modern optimal transport theory (Villani, 2005). In detail, we propose a Wasserstein metric based setting, which leads to a well-formulated class of ambiguity tube MPC controllers admitting a continuous value function; see Theorem 1.
(2) Section 3 presents a complete stability analysis for ambiguity tube MPC for Lipschitz continuous stochastic discrete-time systems under mild assumptions on the coherentness of the optimized performance and risk measures, as well as on the consistency of the terminal cost function of the MPC controller. In detail, Theorem 2 establishes conditions under which the cost function of the ambiguity tube MPC controller is a non-negative supermartingale along its closed-loop trajectories. This can be used to establish robust stability or, under a slightly stronger regularity assumption, robust asymptotic stability of the closed loop system in a stochastic sense, as summarized in Theorems 3 and 4. These results are more general than the existing stability and convergence statements about stochastic MPC in Chatterjee and Lygeros (2015), Kouvaritakis and Cannon (2016) and Munoz-Carpintero and Cannon (2020), as they apply to nonlinear systems and formulations based on ambiguity sets. Besides, Theorem 4 establishes conditions under which the closed-loop trajectories of ambiguity tube MPC controllers are asymptotically stable with respect to a minimal robust invariant set. This is in contrast to the results in Kouvaritakis and Cannon (2016), which only establish stability and convergence of linear stochastic MPC with respect to an ellipsoidal enclosure of the actual (in general, non-ellipsoidal) limit set of the stochastic ancillary closed-loop system.
(3) Section 4 discusses the practical implementation of ambiguity tube MPC. Here, our focus is on linear systems, although remarks on how this can be implemented for nonlinear systems are provided, too. The purpose of this section is to illustrate how the technical assumptions from Section 3 can be satisfied in practice. In this context, a relevant technical contribution is summarized in Lemma 3, which explains how to construct stabilizing terminal cost functions for stochastic and ambiguity tube MPC.

Notice that as much this paper attempts to take a step forward towards a more coherent stability analysis and treatment of stabilizing terminal conditions for stochastic and distributionally robust MPC, it must be stated clearly that we do not claim to be anywhere close to addressing the long list of conceptual issues of robust MPC that D. Mayne summarized in his critique (Mayne, 2015). Nevertheless, in order to assess the role of this paper in the context of recent developments in robust MPC, Section 5 comments on the long list of open problems that research on robust MPC is currently facing.

Notation. If (R, r) is a metric space with a given distance function $r : R \times R \to \mathbb{R}_+$, we use the notation $\mathbb{K}(R)$ to denote the set of compact subsets of R—assuming that it is clear from the context what r is. Similarly, if (R_1, r_1) and (R_2, r_2) are two metric spaces, $\mathcal{L}(R_1, R_2)$ denotes the set of Lipschitz continuous functions from R_1 to R_2 with respect to the distance functions r_1 and r_2. Moreover, $\mathcal{L}_1(R_1, R_2)$ denotes the subset of $\mathcal{L}(R_1, R_2)$ of all functions from R_1 to R_2 whose Lipschitz constant is smaller than

or equal to 1. Finally, $R_1 \times R_2$ is again a metric space with distance function

$$r((a,b),(c,d)) \stackrel{\text{def}}{=} r_1(a,c) + r_2(b,d)$$

for all $a, c \in R_1$ and all $b, d \in R_2$. If nothing else is stated, we assume that the distance function in the new metric space $R_1 \times R_2$ is constructed as above—without always stating this explicitly. Moreover, we denote the distance of a point $a \in \mathbb{R}^n$ to a compact set $B \in \mathbb{K}(\mathbb{R}^n)$ by

$$\text{dist}_q(a, B) \stackrel{\text{def}}{=} \min_{b \in B} \|a - b\|_q ,$$

with $\|\cdot\|_q$ being the Hölder q-norm. For $R_1 \in \mathbb{K}(\mathbb{R}^n)$ and $R_2 \in \mathbb{K}(\mathbb{R}^m)$ the map $d_{\mathcal{L}} : \mathcal{L}(R_1, R_2) \times \mathcal{L}(R_1, R_2) \to \mathbb{R}_+$ denotes the Hilbert–Sobolev distance,

$$d_{\mathcal{L}}(\mu, \nu) \stackrel{\text{def}}{=} \sqrt{\int \|\mu - \nu\|_2^2 + \|\nabla\mu - \nabla\nu\|_2^2 \, \mathrm{d}x} ,$$

which is defined for all $\mu, \nu \in \mathcal{L}(R_1, R_2)$. Here, $\|\cdot\|_2$ denotes the Euclidean norm and ∇ the weak gradient operator recalling that Lipschitz continuous functions are differentiable almost everywhere. Notice that $(\mathcal{L}(R_1, R_2), d_{\mathcal{L}})$ is a metric space.

Let $X \in \mathbb{K}(\mathbb{R}^n)$ denote a given compact set in $\mathbb{R}^n$. We use the symbol $\mathcal{P}(X)$ to denote the set of Borel probability measures on X. With this definition $\mathcal{P}(X)$ is convex and $p(X) = 1$ for all $p \in \mathcal{P}(X)$. Moreover, $\mathcal{B}(X)$ denotes the associated Borel σ-algebra of X. Notice that $\mathcal{P}(X)$ turns out to be a metric space with respect to the Wasserstein distance function,[1] defined as follows.

Definition 1. The function $d_{\mathrm{W}} : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}_+$ denotes the Wasserstein distance,

$$\forall p, q \in \mathcal{P}(X), \quad d_{\mathrm{W}}(p, q) \stackrel{\text{def}}{=} \sup_{\varphi \in \mathcal{L}_1(X, \mathbb{R})} \left| \int \varphi \, \mathrm{d}(p - q) \right| .$$

Notice that d_{W} is well-defined in our context, as all Lipschitz continuous functions are $\mathcal{B}(X)$-measurable. Thus, the integrals in Definition 1 exist and are finite, since we assume that X is compact.

Throughout this paper, we use the notation $\delta_y \in \mathcal{P}(X)$ to denote the Dirac measure at a point $y \in X$, given by

$$\forall Y \in \mathcal{B}(X), \qquad \delta_y(Y) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } y \in Y \\ 0 & \text{otherwise} . \end{cases}$$

Since this paper uses the concept of ambiguity sets intensely, we additionally introduce the shorthand

$$\mathcal{K}(X) \stackrel{\text{def}}{=} \mathbb{K}(\mathcal{P}(X))$$

to denote the set of compact subsets of $\mathcal{P}(X)$—in the Wasserstein space $(\mathcal{P}(X), d_{\mathrm{W}})$. By construction, $\mathcal{K}(X)$ is a metric space with respect to the Hausdorff–Wasserstein distance,

$$d_{\mathrm{H}}(P, Q) \stackrel{\text{def}}{=} \max \left\{ \max_{p \in P} \min_{q \in Q} d_{\mathrm{W}}(p, q), \ \max_{q \in Q} \min_{p \in P} d_{\mathrm{W}}(p, q) \right\} ,$$

defined for all $P, Q \in \mathcal{K}(X)$.

Remark 1. Let w be a random variable with given Lebesgue integrable probability distribution $\rho : \mathbb{R}^n \to \mathbb{R}_+$. The probability of the event $w \in W$ for a Borel set $W \subseteq \mathbb{R}^n$ is denoted by

$$\Pr(w \in W) \stackrel{\text{def}}{=} \int_W 1 \, \mathrm{d}p \stackrel{\text{def}}{=} p(W) \stackrel{\text{def}}{=} \int_W \rho(w) \, \mathrm{d}w .$$

Here, $p \in \mathcal{P}(\mathbb{R}^n)$ is called the probability measure of w. Notice that all four notations are, by definition, equivalent. Many articles on stochastic control, for instance, Bucy (1965) and Kushner (1965, 2014) work with measures rather than probability distributions, due to their technical advantages (Villani, 2005). This means that we specify the probability measure p rather than the probability distribution ρ. The relation

$$\rho = \frac{\mathrm{d}p}{\mathrm{d}w}$$

holds, where the right-hand expression denotes the Radon-Nikodyn derivative of the measure p with respect to the traditional Lebesgue measure (Taylor, 1996).

2. Control of ambiguity tubes

This section introduces a class of uncertain control systems and their Markovian kernels, which can be used to propagate state probability measures. Section 2.3 exploits the properties of these kernels to introduce a topologically coherent framework for defining ambiguity tubes with compact cross-sections in $\mathcal{K}(X)$. Section 2.4 focuses on an axiomatic characterization of proper risk and performance measures. These are used in Section 2.5 to introduce a general class of ambiguity tube MPC controllers, completing the problem formulation.

2.1. Uncertain control systems

This paper concerns uncertain nonlinear discrete-time control systems of the form

$$\forall k \in \mathbb{N}_0, \qquad x_{k+1} = f(x_k, u_k, w_k). \tag{1}$$

Here, x_k denotes the state at time $k \in \mathbb{N}_0$, evolving in the domain $X \in \mathbb{K}(\mathbb{R}^{n_x})$, $u_k \in U$ the control input at time k with domain $U \in \mathbb{K}(\mathbb{R}^{n_u})$, and w_k an uncertain disturbance with support $W \in \mathbb{K}(\mathbb{R}^{n_w})$.

Assumption 1. The potentially nonlinear right-hand side function f satisfies

$$f \in \mathcal{L}(X \times U \times W, X) . \tag{2}$$

Assumption 1 requires f to be Lipschitz continuous and it requires its image set to be contained in X. Thus, it should be mentioned that X is here interpreted as a sufficiently large region of interest in which we analyze the system. If f is Lipschitz continuous on $\mathbb{R}^{n_x}$ but unbounded, we redefine $f \leftarrow \text{proj}_X \circ f$ with proj_X denoting a Lipschitz continuous projection onto X, such that (2) holds by construction. Notice that X should not be mixed up with the set $\mathbb{X} \subseteq X$ that could, for example, model state-constraints; that is, a region in which we would like to keep the state with high probability.

We additionally introduce a compact set $\mathcal{U} \in \mathbb{K}(\mathcal{L}(X, U))$ in order to denote a class of ancillary feedback laws in $(\mathcal{L}(X, U), d_{\mathcal{L}})$. In the context of this paper, $\mathcal{U}$ models a suitable class of computer representable feedback laws.

Example 1. An example for a class of computer representable feedback laws is given by the set

$$\mathcal{U} = \left\{ x \mapsto Kx + k \,\middle|\, \begin{array}{l} \exists k \in \mathbb{R}^{n_u}, \exists K \in \mathbb{R}^{n_u \times n_x} : \\ \|K\| \leq \overline{K} \end{array} \right\}$$

of affine control laws with bounded feedback gain, where $\overline{K} < \infty$ is a given bound on the norm of K, such that all functions in $\mathcal{U}$ are Lipschitz continuous. Specific feedback laws can in this case be represented by storing the finite dimensional matrix K and the offset vector k.

[1] A detailed review of the history and mathematical properties of Wasserstein distances can be found in Villani (2005, Chapter 6).

2.2. Models of stochastic uncertainties

In order to refine (1), we introduce the probability spaces $(W, \mathcal{B}(W), \omega_k)$. Here, $\omega_k \in \mathcal{P}(W)$ is the probability measure of the random variable $w_k : W \to \mathbb{R}$ such that

$$\forall W' \in \mathcal{B}(W), \quad \Pr(w_k \in W') = \omega_k(W') \,.$$

In the most general setting, we might not know ω_k up to a high precision. Instead, we work with the assumption that an ambiguity set $\Omega \in \mathcal{K}(W)$ is given. This means that, for any given $k \in \mathbb{N}_0$, the probability measure ω_k is merely known to satisfy $\omega_k \in \Omega$.

In order to proceed with this modeling assumption, we analyze the closed-loop system

$$\forall k \in \mathbb{N}_0, \qquad x_{k+1} = f(x_k, \mu(x_k), w_k) \,, \tag{3}$$

for a given feedback law $\mu \in \mathcal{U}$. Since the sequence $w_0, w_1, \ldots$ consists of independent random variables, the states x_k are random variables, too. Now, if $p_k \in \mathcal{P}(X)$ denotes a probability measure associated with x_k, then the probability measure $p_{k+1} \in \mathcal{P}(X)$ of x_{k+1} in (3) is a function of p_k, μ, and ω_k. Formally, this propagation of measures can be defined using a parametric Markovian kernel, $\mathcal{N}[x, \mu, \omega] : \mathcal{B}(X) \to \mathbb{R}$, given by

$$\mathcal{N}[x, \mu, \omega](X^+) \stackrel{\text{def}}{=} \omega\left(\left\{ w \in W \,\middle|\, f(x, \mu(x), w) \in X^+ \right\} \right)$$

for all Borel sets $X^+ \in \mathcal{B}(X)$. The transition map

$$\Phi(p, \mu, \omega) \stackrel{\text{def}}{=} \int_X \mathcal{N}[x, \mu, \omega]\, p(\mathrm{d}x) \tag{4}$$

is then well-defined for all $p \in \mathcal{P}(X)$, all $\mu \in \mathcal{U}$, and all probability measures $\omega \in \mathcal{P}(W)$, where $\mathrm{d}x$ denotes the traditional Lebesgue measure (Feller, 1971). This follows from the Lipschitz continuity of f, such that the Markovian kernel $\mathcal{N}[\cdot, \mu, \omega](X^+)$ is for any given Borel set X^+ a $\mathcal{B}(X)$-measurable function in x. In summary, the recursion for the sequence of measures $p_0, p_1, \ldots$ can be written as

$$\forall k \in \mathbb{N}_0, \qquad p_{k+1} = \Phi(p_k, \mu, \omega_k). \tag{5}$$

The next lemma establishes an important property of Φ.

Lemma 1. *If Assumption 1 holds, then we have*

$$\Phi \in \mathcal{L}(\, \mathcal{P}(X) \times \mathcal{U} \times \mathcal{P}(W),\ \mathcal{P}(X)\,) \,.$$

Here, we recall that $\mathcal{P}(X)$ and $\mathcal{P}(W)$ are metric spaces with respect to their Wasserstein distances d_{W}, while $\mathcal{U}$ is equipped with the Hilbert–Sobolev distance $d_{\mathcal{L}}$.

Proof. First, Φ is well defined by (4): Assumption 1 guarantees the Lipschitz continuity of f and that its image set is contained in X. Hence, the image set of Φ is contained in $\mathcal{P}(X)$. Let $p, q \in \mathcal{P}(X)$ and $\omega, \xi \in \mathcal{P}(W)$ be given measures and $\mu, \nu \in \mathcal{U}$ given feedback laws. We set $p^+ = \Phi(p, \mu, \omega)$ and $q^+ = \Phi(q, \nu, \xi)$. The definition of the Wasserstein metric implies that

$$\begin{aligned} d_{\mathrm{W}}(p^+, q^+) &= \sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \left| \int \varphi \,\mathrm{d}p^+ - \int \varphi \,\mathrm{d}q^+ \right| \\ &= \sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \left| \iint \varphi \circ f_\mu \,\mathrm{d}p \,\mathrm{d}\omega - \iint \varphi \circ f_\nu \,\mathrm{d}q \,\mathrm{d}\xi \right| . \end{aligned}$$

Here, $\circ$ denotes the composition operator. Additionally, we have introduced the shorthand notation

$$\forall \mu \in \mathcal{U}, \qquad f_\mu(x, w) \stackrel{\text{def}}{=} f(x, \mu(x), w) \,.$$

Since we assume $\mathcal{U} \in \mathbb{K}(\mathcal{L}(X, U))$, and because the particular definition of $d_{\mathcal{L}}$ implies that all functions in $\mathcal{U}$ are uniformly Lipschitz continuous, each f_μ is – by construction – uniformly Lipschitz continuous on $\mathcal{U}$. Let $\gamma_1 < \infty$ denote its the uniform Lipschitz constant. Since φ is 1-Lipschitz continuous, $\varphi \circ f_\mu$ is also Lipschitz continuous with uniform Lipschitz constant $1 * \gamma_1 = \gamma_1$. Thus, the triangle inequality yields the estimate

$$\begin{aligned} &\left| \iint \varphi \circ f_\mu \,\mathrm{d}p \,\mathrm{d}\omega - \iint \varphi \circ f_\nu \,\mathrm{d}q \,\mathrm{d}\xi \right| \\ &\le \left| \iint \varphi \circ f_\mu \,\mathrm{d}p \,\mathrm{d}\omega - \iint \varphi \circ f_\mu \,\mathrm{d}q \,\mathrm{d}\omega \right| \\ &\quad + \left| \iint \varphi \circ f_\mu \,\mathrm{d}q \,\mathrm{d}\omega - \iint \varphi \circ f_\mu \,\mathrm{d}q \,\mathrm{d}\xi \right| \\ &\quad + \left| \iint \left[\varphi \circ f_\mu - \varphi \circ f_\nu \right] \mathrm{d}q \,\mathrm{d}\xi \right| \\ &\le \gamma_1 \cdot d_{\mathrm{W}}(p, q) + \gamma_1 \cdot d_{\mathrm{W}}(\omega, \xi) \\ &\quad + \left| \iint \left[\varphi \circ f_\mu - \varphi \circ f_\nu \right] \mathrm{d}q \,\mathrm{d}\xi \right| , \end{aligned} \tag{6}$$

which holds uniformly for all $\varphi \in \mathcal{L}_1(X, \mathbb{R})$ and all functions $\mu, \nu \in \mathcal{U}$. Additionally, since q and ξ are probability measures, the last integral term can be bounded as

$$\left| \iint \left[\varphi \circ f_\mu - \varphi \circ f_\nu \right] \mathrm{d}q \,\mathrm{d}\xi \right| \le \gamma_2 \cdot |X| \cdot |W| \cdot d_{\mathcal{L}}(\mu, \nu) \,,$$

where γ_2 denotes the Lipschitz constant of f with respect to its second argument, $|X|$ the diameter of the compact set X and $|W|$ the diameter of the set W. Finally, by substituting all the above inequalities we find that

$$d_{\mathrm{W}}(p^+, q^+) \le \gamma \left(d_{\mathrm{W}}(p, q) + d_{\mathrm{W}}(\omega, \xi) + d_{\mathcal{L}}(\mu, \nu) \right)$$

with $\gamma = \max\{\gamma_1, \gamma_2 \cdot |X| \cdot |W|\}$. But this inequality implies that Φ is indeed Lipschitz continuous. □

Remark 2. The proof of Lemma 1 relies on the properties of Wasserstein (Kantorovich–Rubinstein) distances, which have originally been introduced independently by several authors including Kantorovich (2006) and Vasershtein (1969), see also Villani (2005). To explain why this metric is remarkably powerful in the context of control system analysis, let us briefly discuss what would have happened if we had defined the metric space $\mathcal{P}(X)$, with respect to another metric, for example, a total variation distance,

$$d_{\mathrm{TV}}(p, q) \stackrel{\text{def}}{=} \sup_{A \in \mathcal{B}(X)} |p(A) - q(A)| \,.$$

Let us consider a scalar system with $f(x, u, w) = u$ and parametric feedback laws $\mu_\kappa(x) = \kappa$ with compact domain $\mathcal{U} = \{\mu_\kappa \mid \kappa \in [-1, 1]\}$. In this example, we have

$$\begin{aligned} d_{\mathrm{TV}}(\Phi(p, \mu_\kappa, \omega), \Phi(p, \mu_{\kappa'}, \omega)) &= d_{\mathrm{TV}}(\delta_\kappa, \delta_{\kappa'}) \\ &= \begin{cases} 0 & \text{if } \kappa = \kappa' \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

implying that Φ is not Lipschitz continuous with respect μ_κ on $\mathcal{U}$. In other words, the statement of Lemma 1 is wrong in general, if we replace the Wasserstein distance with other distances—such as the total variation.

2.3. Ambiguity tubes

In this section, we generalize the considerations from the previous section by introducing ambiguity tubes. The motivation for considering such a general setting is twofold: firstly, in practice, one might not know the exact probability measure ω_k of the process noise w_k, but only have a set $\Omega \in \mathcal{K}(W)$ of possible probability measures. For instance, one might know a couple of

lower order moments of w_k, such as the expected value and variance, while higher order moments are unknown. And secondly, in the contexts of nonlinear systems and high dimensional state spaces, propagating the exact state distribution can be difficult or impossible. In such cases, it may be easier to bound the true probability measure of the state by a so-called enclosure; that is, a set of probability measures that – in a suitable, yet to be defined sense – overestimates the actual probability measure of the state.

Let $F : \mathcal{K}(X) \times \mathcal{U} \to \mathcal{K}(X)$ denote the ambiguity transition map that is defined as

$$F(P, \mu) \stackrel{\text{def}}{=} \left\{ \Phi(p, \mu, \omega) \,\middle|\, p \in P,\ \omega \in \Omega \right\} \tag{7}$$

for all $P \in \mathcal{K}(X)$ and all $\mu \in \mathcal{U}$. The ambiguity set $\Omega \in \mathcal{K}(W)$ of possible disturbance probability measures is assumed to be given and constant.

Corollary 1. *Let Assumption 1 hold. The function F is Lipschitz continuous,*

$$F \in \mathcal{L}(\ \mathcal{K}(X) \times \mathcal{U},\ \mathcal{K}(X)\ )\ ,$$

recalling that $\mathcal{K}(X)$ is equipped with the Hausdorff–Wasserstein metric d_{H}.

Proof. Compactness of the image sets of F follows from (7) and the Lipschitz continuity of Φ. Moreover, F directly inherits the Lipschitz continuity of Φ (see Lemma 1), since d_{H} is the Hausdorff metric of d_{W}. □

In order to formalize the concept of ambiguity enclosures, the following definition is introduced.

Definition 2. Let $P, Q \in \mathcal{K}(X)$ be ambiguity sets. The set Q is called an enclosure of P, denoted by $P \preceq Q$, if

$$\sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \left[\max_{p \in P} \min_{q \in Q} \int \varphi \, \mathrm{d}(p - q) \right] \leq 0\ .$$

The ambiguity sets P and Q are equivalent, denoted by $P \simeq Q$, if both $P \preceq Q$ and $Q \preceq P$.

The above definition of the relation "$\preceq$" should not be mixed up with the set inclusion relation "$\subseteq$", as used for the definition of set enclosures in the field of set-theoretic tube MPC and global optimization. The conceptual difference is illustrated by the following example.

Example 2. Let us consider the ambiguity sets

$$P = \{\delta_0, \delta_1\} \quad \text{and} \quad Q = \{\delta_0, \delta_1, 0.5\delta_0 + 0.5\delta_1\}$$

of the compact set $X = [0, 1] \subseteq \mathbb{R}$ recalling that δ_0 and δ_1 denote the Dirac measures at 0 and 1, respectively. Here, the upper bounds of the integrals,

$$\max_{p \in P} \int \varphi \ \mathrm{dp} = \max\{\ \varphi(0),\ \varphi(1)\ \}$$

and $$\max_{q \in Q} \int \varphi \ \mathrm{dq} = \max\{\ \varphi(0),\ \varphi(1)\ \}\ ,$$

coincide for any Lipschitz continuous function φ. Similarly, the associated lower bounds

$$\min_{p \in P} \int \varphi \ \mathrm{dp} = \min\{\ \varphi(0),\ \varphi(1)\ \}$$

and $$\min_{q \in Q} \int \varphi \ \mathrm{dq} = \min\{\ \varphi(0),\ \varphi(1)\ \}\ ,$$

coincide, too. Consequently, in the sense of Definition 2, the ambiguity sets P and Q are equivalent, $Q \simeq P$. In particular, we have $Q \preceq P$ but we do not have $Q \subseteq P$. Thus, the relations $\preceq$ and $\subseteq$ are not the same.

The following proposition establishes that $\preceq$ defines a partial order on $\mathcal{K}(X)$ with respect to $\simeq$. Moreover, topological compatibility with respect to our Hausdorff–Wasserstein metric setting is established.

Proposition 1. *Let the enclosure relation $\preceq$ be defined as in Definition 2. Then, the following properties are satisfied for any $P, Q, T \in \mathcal{K}(X)$.*

(1) Reflexivity: we have $P \preceq P$.
(2) Anti-Symmetry: if $P \preceq Q$ and $Q \preceq P$, then $P \simeq Q$.
(3) Transitivity: if $P \preceq Q$ and $Q \preceq T$, then $P \preceq T$.
(4) Compactness: The set

$$\mathcal{S} = \{(P, Q) \in \mathcal{K}(X) \times \mathcal{K}(X) \mid P \preceq Q\}$$

is compact; that is, $\mathcal{S} \in \mathbb{K}(\mathcal{K}(X) \times \mathcal{K}(X))$.

Proof. Reflexivity, anti-symmetry with respect to the equivalence relation $\simeq$, and transitivity follow directly from Definition 2. Our focus is on the last statement, which claims to establish compatibility of Definition 2 and the proposed Wasserstein–Hausdorff metric setting. Let $P_0, P_1, P_2, \ldots \in \mathcal{K}(X)$ and $Q_0, Q_1, Q_2, \ldots \in \mathcal{K}(X)$ be two convergent sequences with

$$P^\star \stackrel{\text{def}}{=} \lim_{k \to \infty} P_k \qquad \text{and} \qquad Q^\star \stackrel{\text{def}}{=} \lim_{k \to \infty} Q_k$$

and such that $P_k \preceq Q_k$ for all $k \in \mathbb{N}$. Because all sets are compact, the maximizers

$$p^\star_{k,\varphi} \stackrel{\text{def}}{=} \operatorname*{argmax}_{p \in P_k} \int \varphi \, \mathrm{d}p \quad \text{and} \quad q^\star_{k,\varphi} \stackrel{\text{def}}{=} \operatorname*{argmax}_{q \in Q_k} \int \varphi \, \mathrm{d}q$$

exist for all $\varphi \in \mathcal{L}_1(X, \mathbb{R})$. Next, since $\mathcal{K}(X)$ is a compact set of compact sets, we have not only $P^\star, Q^\star \in \mathcal{K}(X)$, but the triangle inequality for the Hausdorff–Wasserstein metric additionally yields that

$$\widetilde{p}_\varphi \stackrel{\text{def}}{=} \lim_{k \to \infty} p^\star_{k,\varphi} = \operatorname*{argmax}_{p \in P^\star} \int \varphi \, \mathrm{d}p$$

and $$\widetilde{q}_\varphi \stackrel{\text{def}}{=} \lim_{k \to \infty} q^\star_{k,\varphi} = \operatorname*{argmax}_{q \in Q^\star} \int \varphi \, \mathrm{d}q\ .$$

A direct consequence of these equations is that we have

$$\begin{aligned} &\sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \left[\max_{p \in P^\star} \min_{q \in Q^\star} \int \varphi \, \mathrm{d}(p - q) \right] \\ &\quad = \sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \{\widetilde{p}_\varphi - \widetilde{q}_\varphi\} \\ &\quad = \sup_{\varphi \in \mathcal{L}_1(X,\mathbb{R})} \lim_{k \to \infty} \{\underbrace{p^\star_{\varphi,k} - q^\star_{\varphi,k}}_{\leq 0}\} \leq 0\ , \end{aligned}$$

which shows that $P^\star \preceq Q^\star$. Notice that this means that if $(P_k, Q_k) \in S$ is a Cauchy sequence, then the limit point satisfies $(P^\star, Q^\star) \in S$; that is, S is closed. Because S is bounded by construction, this also implies that S is compact, $S \in \mathbb{K}(\mathcal{K}(X) \times \mathcal{K}(X))$. □

After this technical preparation, the following definition of ambiguity tubes is possible.

Definition 3. The sequence $(P_0, P_1, \ldots, P_N) \in \mathcal{K}(X)^{N+1}$ is called an ambiguity tube of (1) on the discrete-time horizon $\{0, 1, \ldots, N\}$, if there exists a sequence of ancillary feedback controllers $\mu_0, \mu_1, \ldots, \mu_{N-1} \in \mathcal{U}$ such that

$$\forall k \in \{0, 1, \ldots, N-1\}, \qquad F(P_k, \mu_k) \preceq P_{k+1}\ .$$

Definition 3 is inspired by similar definitions from the field of set-theoretic methods in control (Houska & Villanueva, 2019;

Langson et al., 2004; Mayne et al., 2005). In detail, the step from set-valued robust forward invariant tubes to ambiguity tubes is, however, not straightforward. For instance, the inclusion relation "$\subseteq$" would be too strong for a practical definition of ambiguity tubes and is therefore replaced by the relation $\preceq$. This adaption of concepts to our measure based setting is needed, as the purpose of constructing tubes for standard set propagation and ambiguity set propagation is different. As discussed in the following sections, ambiguity tubes can be used to assess, analyze, and trade-off the risk of constraint violations with other performance measures rather than enforcing worst-case constraints used in traditional tube MPC.

Remark 3. An equivalent characterization of the relations in Definition 2 can be obtained by borrowing notation from the field of convex optimization that is related to the concept of duality and support functions (Boyd & Vandenberghe, 2004; Rockafellar & Wets, 2005). In order to explain this, we denote with $d_P : \mathcal{L}_1(X, \mathbb{R}) \to \mathbb{R}$ the support function of the ambiguity set P,

$$\forall \varphi \in \mathcal{L}_1(X, \mathbb{R}), \qquad d_P(\varphi) \stackrel{\text{def}}{=} \max_{p \in P} \int \varphi \, \mathrm{d}p \, .$$

This notation is such that we have $d_P = d_Q$ if and only if $P \simeq Q$. Similarly, we have $d_P \leq d_Q$ if and only if $P \preceq Q$.

2.4. Proper ambiguity measures

The goal of this section is to formalize certain concepts of modeling performance and risk in the space of ambiguity sets. For this aim, we introduce maps of the form

$$\ell : \mathcal{K}(X) \to \mathbb{R} \, ,$$

which assign real values to ambiguity sets. The following definition proposes a regularity condition under which an ambiguity measure is considered "proper".[2]

Definition 4. The ambiguity measure $\ell : \mathcal{K}(X) \to \mathbb{R}$ is proper, if it is Lipschitz continuous, $\ell \in \mathcal{L}(\mathcal{K}(X), \mathbb{R})$, linear with respect to weighted Minkowski sums,

$$\forall \theta \in [0, 1], \quad \ell(\theta P \oplus (1-\theta) Q) = \theta \ell(P) + (1-\theta)\ell(Q) \, ,$$

and monotonous; that is, $P \preceq Q$ implies $\ell(P) \leq \ell(Q)$ for $P, Q \in \mathcal{K}(X)$.

Notice that the above definition ensures that any proper ambiguity measure ℓ satisfies $\ell(P) = \ell(Q)$ for any $P, Q \in \mathcal{K}(X)$ with $P \simeq Q$. Together with the monotonicity, it implies that proper ambiguity measures are compatible with our definition of the relations "$\preceq$" and "$\simeq$" from Definition 2. In order to understand why practical performance and risk measures can be assumed to be proper without adding much of a restriction, it is helpful to have the following examples in mind.

Example 3. If $l \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ denotes a cost function, for example, the stage cost of a nominal MPC controller, the associated worst-case average performance

$$\ell(P) \stackrel{\text{def}}{=} \max_{p \in P} \int l \, \mathrm{d}p$$

is well defined, where the maximizer exists for compact ambiguity sets $P \in \mathcal{K}(X)$. It is easy to check that ℓ is a proper ambiguity measure in the sense of Definition 4.

Example 4. If $\mathbb{X} \in \mathbb{K}(X)$ denotes a state constraint, its maximum expected constraint violation at risk is given by

$$\mathcal{R}(P) \stackrel{\text{def}}{=} \max_{p \in P} \int \mathrm{dist}_1(x, \mathbb{X}) \, p(\mathrm{d}x)$$

recalling that $\mathrm{dist}(x, \mathbb{X}) = \min_{z \in \mathbb{X}} \|x - z\|_1$ denotes the distance function with respect to the 1-norm. Similar to the previous example, $\mathcal{R}$ is a proper ambiguity measure in the sense of Definition 4, which can here be interpreted as a risk measure. In fact, it is closely related to the so-called worst-case conditional value at risk, as introduced by Rockafellar and Uryasev (2013), which is accepted as one of the most practical and computationally tractable risk measures in engineering and management sciences.

2.5. Ambiguity tube MPC

This section focuses on the formulation of ambiguity tube MPC problems of the form

$$\mathcal{V}(y) \stackrel{\text{def}}{=} \min_{P, \mu} \sum_{k=0}^{N-1} L(P_k, \mu_k) + M(P_N)$$
$$\text{s.t.} \quad \begin{cases} \forall k \in \{0, 1, \ldots, N-1\}, \\ F(P_k, \mu_k) \preceq P_{k+1} \\ P_k \in \mathcal{K}(X), \;\; \mu_k \in \mathcal{U} \\ \delta_y \in P_0 \, . \end{cases} \tag{8}$$

Here, the sequence of ambiguity sets $P = (P_0, P_1, \ldots, P_N)$ and ancillary feedback laws $\mu = (\mu_0, \mu_1, \ldots, \mu_N)$ are optimization variables. The current state measurement is y – recalling that δ_y denotes its Dirac measure – while

$$L : \mathcal{K}(X) \times \mathcal{U} \to \mathbb{R} \quad \text{and} \quad M : \mathcal{K}(X) \to \mathbb{R}$$

denote the stage and end costs. If $\mu_0^\star[y] \in \mathcal{U}$ denotes the parametric minimizer of (8), the MPC feedback law is

$$\mu_{\text{MPC}}(y) \stackrel{\text{def}}{=} \mu_0^\star[y](y). \tag{9}$$

Notice that in this notation, the current time of the MPC controller is reset to 0 after every iteration. The next theorem introduces a minimum requirement under which one could call (8) well-formulated.

Theorem 1. *Let Assumption 1 hold and let L and M be continuous functions on the compact domains $\mathcal{K}(X) \times \mathcal{U}$ and $\mathcal{K}(X)$, respectively, then (8) admits a minimizer for any $y \in X$. Moreover, the function $\mathcal{V}$ is continuous on X.*

Proof. Since Assumption 1 holds, Corollary 1 can be combined with the fourth statement of Proposition 1 to conclude that the feasible set of (8) is non-empty and compact. Thus, if L and M are continuous, Weierstrass' theorem yields the first statement of this theorem. The second statement follows from a variant of Berge's theorem (Berge, 1963); see, also Rockafellar and Wets (2005, Thm. 1.17). □

The above theorem has been formulated under a rather weak requirement on the continuity of the functions L and M; that is, without necessarily requiring that these functions are proper ambiguity measures. However, as discussed in the following sections, stronger assumptions on L and M are needed, if one is interested in analyzing the stability properties of the MPC controller (8).

[2] The conditions in Definition 4 are of an axiomatic nature, inspired by similar axioms for coherent risk measures, as introduced in Rockafellar and Uryasev (2013). One difference is, however, that we work with ambiguity sets rather than probability measures. Moreover, Definition 4 is tailored to our Wasserstein–Hausdorff metric setting in which Lipschitz continuity (not only closedness of image sets as required for regular risk measures Rockafellar & Uryasev, 2013) is needed for ensuring topological compatibility.

Remark 4. The ambiguity tube MPC formulation (8) includes traditional tube MPC as well as stochastic MPC formulations as special cases. In the first case, Ω denotes the set of all probability distributions with support set W while, in the second case, $\Omega = \{\omega\}$ is a singleton. Here, (8) is formulated under the convention that state-constraints are taken into account by adding suitable risk measures to the objective, as explained by Example 4. This is, from the perspective of stochastic MPC, rather natural. In such a setting one would usually be interested in an objective that allows one to tradeoff between the risk of violating a constraint and control performance. Nevertheless, for the sake of generality of the following analysis, it should be mentioned that if one is interested in enforcing explicit chance constraints, the corresponding MPC controllers can only be reformulated as a problem of the form (8), if additional assumptions on the regularity[3] and recursive feasibility of these constraints are made—such that they can be added to the stage cost in the form of L_1-penalties without altering the problem formulation. Such conditions have been discussed in all detail in Rawlings et al. (2018) for min–max MPC and in Kouvaritakis and Cannon (2016) for stochastic MPC.

3. Stability analysis

As mentioned before, the basic concepts for analyzing stability of Markovian systems using martingale theory can be found in Bucy (1965) and Kushner (1965). The goal of this section is to lay the foundation for applying these concepts to analyze the stochastic closed-loop stability properties of the ambiguity tube MPC controller (8) in the presence of uncertainties. For this aim, this section is divided into three parts: Section 3.1 concisely presents all assumptions that will be needed for this stability analysis, Section 3.2 establishes an important technical result regarding the concavity of MPC cost functions with respect to Minkowski addition of ambiguity sets, and Section 3.3 uses this concavity property to construct a non-negative supermartingale, which finally leads to the stability results for ambiguity tube MPC that are summarized in Theorems 3 and 4.

3.1. Conditions on the stage and terminal cost function

Throughout the following stability analysis, two main assumptions on the stage and end cost function of the MPC controller (8) are needed, as introduced below.

Assumption 2. The functions $L(\cdot, \mu)$ and M are for any given $\mu \in \mathcal{U}$ proper ambiguity measures. Moreover, we assume that L is continuous on $\mathcal{K}(X) \times \mathcal{U}$.

Notice that Examples 3 and 4 discuss the formulation of practical risk and performance measures in such a way that Assumption 2 holds. A separate assumption on the continuity of M (as in Theorem 1) is not needed anymore, since proper ambiguity measures are Lipschitz continuous functions and, as such, continuous. Assumption 2 does, however, add a continuity requirement for L, as this function depends in general on the feedback law μ. In this way, we guarantee that the conditions of Theorem 1 are satisfied whenever Assumptions 1 and 2 are satisfied. The next assumption introduces an additional condition on the terminal cost function M.

Assumption 3. The functions L and M are non-negative and satisfy the terminal descent condition

$$\forall P \in \mathcal{K}(X),\ \exists \mu \in \mathcal{U}:\quad L(P, \mu) + M(F(P, \mu)) \leq M(P)\,.$$

[3] Using proper ambiguity measures in order to formulate constraints, is clearly sufficient to ensure regularity.

Assumption 3 can be interpreted as a Lyapunov decent condition. It is similar to the terminal descent conditions that are typically introduced in the context of certainty-equivalent and min–max MPC (Chen & Allgöwer, 1998; Grüne, 2009; Rawlings et al., 2018; Villanueva et al., 2020). The construction of functions L and M that satisfy Assumptions 2 and 3 simultaneously will be discussed in Section 4.2.

Remark 5. Assumptions 2 and 3 together imply that M must be a Lipschitz continuous Control Lyapunov Function (CLF). In the general context of nonlinear system analysis, conditions under which such Lipschitz continuous CLFs exist have been analyzed by various authors (Clarke, Ledyaev, & Stern, 1998; Ledyaev & Sontag, 1999). However, if one considers nonlinear MPC problems with explicit state constraints (see also Remark 4), it is possible to construct systems – for example, based on Artstein's circles – that are asymptotically stabilizable yet fail to admit a continuous CLF (Grimm, Messina, Tuna, & Teel, 2004). It is, however, also pointed out in Grimm et al. (2004) that systems admitting only discontinuous CLFs often lead to non-robust MPC controllers. Therefore, in the context of robust MPC design, the motivation behind Assumptions 2 and 3 is to exclude such pathological non-robustly stabilizable systems. Notice that more general regularity assumptions, under which a robust control design is possible, are beyond the scope of this paper.

3.2. On concave cost functions

After summarizing all main assumptions, we can now focus on the properties of certain cost functions that will later be used to construct a supermartingale for the ambiguity tube MPC controller (8). For this aim, we first introduce the auxiliary function

$$\mathcal{J}_\mu(Q) \stackrel{\text{def}}{=} \min_{P} \sum_{k=0}^{N-1} L(P_k, \mu_k) + M(P_N) \quad \text{s.t.} \begin{cases} \forall k \in \{0, 1, \ldots, N-1\}, \\ F(P_k, \mu_k) \preceq P_{k+1} \\ P_k \in \mathcal{K}(X) \\ Q \preceq P_0,\ P_N = P^\star\,, \end{cases} \tag{10}$$

defined for all $Q \in \mathcal{K}(X)$ and all feedback laws $\mu \in \mathcal{U}^N$.

Lemma 2. *If Assumptions 1 and 2 are satisfied, then $\mathcal{J}_\mu$ is for any given $\mu \in \mathcal{U}^N$ a proper ambiguity measure.*

Proof. In the following, we may assume that $\mu \in \mathcal{U}^N$ is constant and given. Our proof is divided into two parts. The first part focuses on establishing a linearity property of F. The second part of the proof builds upon the first part in to further analyze the properties of $\mathcal{J}_\mu$.

PART I: Recall that Φ, as defined in (4), is, by construction, linear in its first argument. Consequently, we have

$$\theta\Phi(p, \nu, \omega) + (1-\theta)\Phi(q, \nu, \omega) = \Phi(\theta p + (1-\theta)q, \nu, \omega)$$

for all $p, q \in \mathcal{P}(X)$ and all $\theta \in [0, 1]$, for any given $\omega \in \mathcal{P}(W)$ and $\nu \in \mathcal{U}$. This implies that F satisfies

$$\theta F(P, \nu) \oplus (1-\theta)F(Q, \nu) = F(\theta P \oplus (1-\theta)Q, \nu)\,,$$

for all $P, Q \in \mathcal{K}(X)$, all $\nu \in \mathcal{U}$, and all $\theta \in [0, 1]$, which follows from the definition of F in (7).

PART II: Let $\mathfrak{P}_0[Q], \ldots, \mathfrak{P}_N[Q] \in \mathcal{K}(X)$ be the solution of the recursion

$$\mathfrak{P}_0[Q] = Q\,,$$

$$\mathfrak{P}_{k+1}[Q] = F(\mathfrak{P}_k[Q], \mu_k)$$

for $k \in \{0, 1, \ldots, N-1\}$ recalling that μ is given. Corollary 1 ensures that the transition map F is Lipschitz continuous such that the above recursion generates compact ambiguity sets for any compact input set $Q \in \mathcal{K}(X)$, such that the sequence $\mathfrak{P}_0, \ldots, \mathfrak{P}_N$ is well-defined. Due to the linearity of F with respect to its first argument (see Part I), it follows by induction over k that

$$\mathfrak{P}_k[\theta Q \oplus (1-\theta)Q'] = \theta \mathfrak{P}_k[Q] \oplus (1-\theta)\mathfrak{P}_k[Q'] \tag{11}$$

for all $Q, Q' \in \mathcal{K}(X)$ and all $\theta \in [0, 1]$. Next, since both ambiguity measures L and M are, by Assumption 2, monotonous and Lipschitz continuous, we have

$$\mathcal{J}_\mu(Q) = \sum_{k=0}^{N-1} L(\mathfrak{P}_k[Q], \mu_k) + M(\mathfrak{P}_N[Q]) \tag{12}$$

for any $Q \in \mathcal{K}(X)$.

Finally, combining (11) and (12) with the assumption that L and M are proper ambiguity measures, imply that $\mathcal{J}_\mu$ is a proper ambiguity measure, too. □

Notice that $\mathcal{J}_\mu$ depends on the feedback law μ, which is optimized in the context of MPC. Consequently, we are in the following not directly interested in this auxiliary function, but rather in the actual cost-to-go function

$$J(Q) \stackrel{\text{def}}{=} \min_{\mu \in \mathcal{U}^N} \mathcal{J}_\mu(Q)\,, \tag{13}$$

which is defined for all $Q \in \mathcal{K}(X)$, too. Clearly, the function J is closely related to the value function $\mathcal{V}$ of the ambiguity tube MPC controller (8), as we have

$$\forall y \in X, \qquad \mathcal{V}(y) = J(\{\delta_y\}). \tag{14}$$

This follows directly by comparing the definition of $\mathcal{V}$ in (8) with (10) and (13). The following corollary summarizes an important consequence of Lemma 2.

Corollary 2. *Let Assumptions 1 and 2 be satisfied. The cost-to-go function J is concave with respect to weighted Minkowski addition; that is, we have*

$$J(\theta Q \oplus (1-\theta)Q') \;\geq\; \theta J(Q) + (1-\theta)J(Q') \tag{15}$$

for all $Q, Q' \in \mathcal{K}(X)$ and all $\theta \in [0, 1]$. Moreover, J is monotonous; that is, $Q \preceq Q'$ implies $J(Q) \leq J(Q')$.

Proof. The key idea for establishing the first statement is to use the linearity of $\mathcal{J}_\mu$ (Lemma 2). This yields

$$\begin{aligned}
&J(\theta Q \oplus (1-\theta)Q') \\
&= \min_{\mu \in \mathcal{U}^N} \mathcal{J}_\mu(\theta Q \oplus (1-\theta)Q') \\
&= \min_{\mu \in \mathcal{U}^N} \left\{\theta \mathcal{J}_\mu(Q) + (1-\theta)\mathcal{J}_\mu(Q')\right\} \\
&\geq \theta \left(\min_{\mu \in \mathcal{U}^N} \mathcal{J}_\mu(Q)\right) + (1-\theta)\left(\min_{\mu' \in \mathcal{U}^N} \mathcal{J}_{\mu'}(Q')\right) \\
&= \theta J(Q) + (1-\theta)J(Q')
\end{aligned}$$

for all $Q, Q' \in \mathcal{K}(X)$ and all $\theta \in [0, 1]$. This corresponds to the first statement of the lemma.

The second statement follows from the fact that any minimizer of

$$\begin{aligned}
J(Q') \;=\; \min_{P,\mu} &\sum_{k=0}^{N-1} L(P_k, \mu_k) + M(P_N) \\
\text{s.t.} &\begin{cases} \forall k \in \{0, 1, \ldots, N-1\}, \\ F(P_k, \mu_k) \preceq P_{k+1} \\ P_k \in \mathcal{K}(X)\,,\;\; \mu_k \in \mathcal{U} \\ Q' \preceq P_0\,, \end{cases}
\end{aligned} \tag{16}$$

is a feasible point of the optimization problem obtained when we replace the ambiguity set $Q' \in \mathcal{K}(X)$ with an ambiguity set Q that satisfies $Q \preceq Q'$, which then implies monotonicity; that is, $J(Q) \leq J(Q')$. □

Corollary 2 is central to establishing stability properties of ambiguity tube MPC schemes. The following example helps to understand the role of this concavity statement in the ongoing developments.

Example 5. Let us consider the example that $Q = \{\delta_a\}$ and $Q' = \{\delta_b\}$ for two given points $a, b \in X$. In this case, $J(Q)$ is the cost that is associated with knowing that we are currently at the point a. Similarly, $J(Q')$ can be interpreted as the cost that is associated with knowing that we are currently at the point b. Now, if we set $\theta = \frac{1}{2}$, the corresponding ambiguity set

$$\theta Q \oplus (1-\theta)Q' = \left\{\frac{1}{2}\delta_a + \frac{1}{2}\delta_b\right\}$$

can be associated with the situation that we do not know whether we are at a or b, as both events could happen with probability $\frac{1}{2}$. Thus, in this example, the first statement of Corollary 2 is saying that the cost that is associated with not knowing our state is larger or equal than the expected cost obtained when planning to first measure whether we are at a or b and then evaluating the cost function.

The following section exploits the conceptual idea from the above example and Corollary 2 for constructing a non-negative supermartingale for ambiguity tube MPC.

3.3. Supermartingales for ambiguity tube MPC

The goal of this section is to establish conditions under which the value function $\mathcal{V}$ is a supermartingale along the trajectories of the closed system associated with the ambiguity tube MPC feedback μ_{MPC}, as defined in (9). Here, we recall that the closed-loop stochastic process associated with (8) is denoted by

$$\forall k \in \mathbb{N}, \qquad y_{k+1} = f(y_k, \mu_{\text{MPC}}(y_k), w_k)\,, \tag{17}$$

with $w_0, w_1, \ldots : W \to W$, denoting independent $\mathcal{B}(W)$-measurable random variables in the probability spaces $(W, \mathcal{B}(W), \omega_k)$ that depend on the sequence of measures $\omega_0, \omega_1, \ldots \in \Omega$. This implies that $y_0, y_1, \ldots$ are random variables, too. In the following, we use the notation $\mathfrak{S}_k = \sigma(y_0, y_1, \ldots, y_k)$ to denote the minimal σ-field of the sequence $y_0, y_1, \ldots, y_k$, such that

$$\forall k \in \mathbb{N}, \qquad \mathfrak{S}_k \subseteq \mathfrak{S}_{k+1}$$

is a filtration of σ-algebras. Moreover, the conditional expectation of $\mathcal{V}(y_{k+1})$ given $\mathfrak{S}_k$ (Taylor, 1996) is denoted by

$$\mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} \stackrel{\text{def}}{=} \int \mathcal{V}(f(y_k, \mu_{\text{MPC}}(y_k), \cdot))\,\mathrm{d}\omega_k\,.$$

This definition depends on the sequence $\omega_0, \omega_1, \ldots \in \Omega$. If Assumptions 1 and 2 hold, Theorem 1 ensures that $\mathcal{V}$ is continuous. Since Assumption 1 also ensures that f is Lipschitz continuous, the integrand in the above expression is $\mathcal{B}(W)$-measurable and, consequently, the conditional expectation of $\mathcal{V}(y_{k+1})$ given $\mathfrak{S}_k$ is well-defined.

Theorem 2. *Let Assumptions 1, 2, and 3 be satisfied. Then the function $\mathcal{V}$ is a supermartingale along the trajectories of the closed-loop system (17); that is, we have*

$$\forall k \in \mathbb{N}, \qquad \mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} \leq \mathcal{V}(y_k)$$

independent of the choice of $\omega_0, \omega_1 \ldots \in \Omega$.

Proof. Let $P_0^\star(y), P_1^\star(y), \ldots \in \mathcal{K}(X)$ denote the parametric minimizers of (8) such that

$$\mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} \quad \leq \quad \max_{p \in P_1^\star(y_n)} \int \mathcal{V} \, \mathrm{d}p \tag{18}$$

holds – by definition of $P_1^\star(y_n)$ – for any choice of the sequence $\omega_0, \omega_1 \ldots \in \Omega$. Since Assumptions 1 and 2 are satisfied, the first statement of Corollary 2 implies that

$$\begin{aligned} \int \mathcal{V} \, \mathrm{d}p &\overset{(14)}{=} \int J\left(\{\delta_y\}\right) \, p(\mathrm{d}y) \\ &\overset{(15)}{\leq} J\left(\left\{\int \delta_y \, p(\mathrm{d}y)\right\}\right) = J(\{p\}) \end{aligned} \tag{19}$$

holds for all probability measure $p \in \mathcal{P}(X)$. Moreover, we know from the second statement of Corollary 2 that J is monotonous implying that

$$p \in P \quad \Longrightarrow \quad \{p\} \preceq P \quad \Longrightarrow \quad J(\{p\}) \leq J(P) \tag{20}$$

for all $P \in \mathcal{K}(X)$. In order to briefly summarize our intermediate results so far, we combine (18), (19), and (20), which yields the inequality

$$\mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} \quad \leq \quad J(P_1^\star(y_k)). \tag{21}$$

Next, in analogy to the construction of Lyapunov functions for traditional MPC controllers (Rawlings et al., 2018), we use that Assumption 3 implies that the cost-to-go function J descends along the iterates $P_i^\star(y_k)$, which means that

$$J(P_1^\star(y_k)) \quad \leq \quad J(P_0^\star(y_k)) \quad = \quad \mathcal{V}(y_k). \tag{22}$$

But then (21) and (22) imply that we also have

$$\mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} \quad \leq \quad \mathcal{V}(y_k) \, .$$

The latter inequality does not depend on the choice of the sequence $\omega_0, \omega_1, \ldots \in \Omega$ and, consequently, corresponds to the statement of the theorem. □

3.4. Stability of ambiguity tube MPC

As established in Theorem 1, Assumptions 1 and 2 are sufficient to ensure that $\mathcal{V}$ is a continuous function. Thus,

$$\mathcal{V}^\star \overset{\text{def}}{=} \min_{y \in X} \mathcal{V}(y) \quad \text{and} \quad Y^\star \overset{\text{def}}{=} \operatorname*{argmin}_{y \in X} \mathcal{V}(y) \tag{23}$$

are well-defined and the set $Y^\star$ is compact, $Y^\star \in \mathbb{K}(X)$, and non-empty (see Theorem 1). We may assume, without loss of generality, that $\mathcal{V}^\star = 0$, as adding constant offsets to $\mathcal{V}$ does not affect the result of Theorem 2. Let

$$\mathcal{N}_\varepsilon(Y^\star) \overset{\text{def}}{=} \left\{ x \in X \;\middle|\; \mathrm{dist}_2(x, Y^\star) < \varepsilon \right\}$$

denote an ε-neighborhood of $Y^\star$. The definition below introduces a (standard) notion of stability of the closed-loop system (17). It is helpful to recall that the probability measures p_k of the sequence y_k are given by the recursion

$$p_{k+1} = \Phi(p_k, \mu_{\mathrm{MPC}}, \omega_k) \, ,$$

which depends on the sequence $\omega_0, \omega_1, \ldots \in \Omega$ and on the given initial state measurement y_0, as we set $p_0 = \delta_{y_0}$.

Definition 5. The closed-loop system (17) is called robustly stable with respect to the set $Y^\star$, if there exists for every $\varepsilon > 0$ a $\delta > 0$ such that for any $y_0 \in \mathcal{N}_\delta(Y^\star)$ we have

$$\forall k \in \mathbb{N}, \quad p_k\left(\mathcal{N}_\varepsilon(Y^\star)\right) \; > \; 1 - \varepsilon \, ,$$

independent of the choice of the probability measures $\omega_0, \omega_1, \ldots \in \Omega$. If we additionally have that

$$\lim_{k \to \infty} p_k(Y^\star) = 1 \, ,$$

independent of the choice of $\omega_0, \omega_1, \ldots \in \Omega$, we say that the closed-loop system is robustly asymptotically stable with respect to $Y^\star$.

The main result of this section is summarized as follows.

Theorem 3. *If Assumptions 1, 2, and 3 hold, then (17) is robustly stable with respect to $Y^\star$.*

Proof. The assumptions of this theorem ensure that $\mathcal{V}$ is continuous (Theorem 1) and a supermartingale along the trajectories of (17), independent of the choice of the probability measures $\omega_0, \omega_1, \ldots \in \Omega$ (Theorem 2). Also, since we work with a compact support, all random variables are essentially bounded. Thus, we can apply Bucy's supermartingale stability theorem (Bucy, 1965, Thm. 1) to conclude that (17) is robustly stable with respect to $Y^\star$. □

Remark 6. The proof of Theorem 3 is based on the result of Bucy's original article on positive supermartingales, who formally only established stability of Markov processes with respect to an isolated equilibrium; that is, for the case that the set $Y^\star$ is a singleton. However, the proof of Theorem 1 in Bucy (1965) generalizes trivially to the version needed in our proof after replacing the distance between the states of the Markov system and the equilibrium point by the corresponding distance of these iterates to the set $Y^\star$. By now, this and other generalizations of the supermartingale based stability theorems by Bucy and Kushner for Markov processes are, of course, well-known and can in very similar versions be found in Feller (1971), Kushner (1965, 2014) and Taylor (1996).

For the case that we are not only interested in robust stability but also robust asymptotic stability, the above theorem can be extended after introducing a slightly stronger regularity requirement on the function L.

Definition 6. The function L is positive definite with respect to $Y^\star$ if $L(P, \mu) > 0$ for all $\mu \in \mathcal{U}$ and all $P \in \mathcal{K}(X)$ with

$$\min_{p \in P} p(Y^\star) < 1 \, .$$

A stronger version of Theorem 3 is formulated as follows.

Theorem 4. *Let Assumptions 1, 2, and 3 hold. If L is positive definite with respect to $Y^\star$, then (17) is robustly asymptotically stable with respect to $Y^\star$.*

Proof. The statement of this theorem is similar to the statement of Theorem 3, but we need to work with a slightly tighter version of the supermartingale inequality from Theorem 2. For this aim, we use that Assumption 3 implies that

$$L(P_0^\star(y_k), \mu^\star[y_k]) + J(P_1^\star(y_k)) \quad \leq \quad J(P_0^\star(y_k)). \tag{24}$$

Consequently, the inequality (22) can be replaced by its tighter version,

$$\min_{p \in P_0^\star(y_k)} p(Y^\star) < 1 \quad \Longrightarrow \quad J(P_1^\star(y_k)) < \mathcal{V}(y_k) \, .$$

Thus, the argument in the proof of Theorem 2 can be modified finding that we also have

$$\min_{p \in P_0^\star(y_k)} p(Y^\star) < 1 \quad \Longrightarrow \quad \mathbb{E}\{\mathcal{V}(y_{k+1}) \mid \mathfrak{S}_k\} < \mathcal{V}(y_k)$$

for all $k \in \mathbb{N}$, independent of the choice of the sequence $\omega_0, \omega_1 \ldots \in \Omega$. But this means that $\mathcal{V}$ is a strict non-negative supermartingale and we can apply the standard result from Bucy (1965, Thm. 2) (of course, again after replacing Bucy's outdated definition of equilibrium points with our definition of the set $Y^\star$—see Remark 6) to establish the statement of this theorem. □

4. Practical implementation of ambiguity tube MPC for linear systems

In order to illustrate and discuss the above theoretical results, this section develops a practical framework for reformulating a class of ambiguity tube MPC controllers for linear systems as convex optimization problems that can then be solved with existing MPC software. In particular, Section 4.2 focuses on the question how to construct stabilizing terminal costs for stochastic and ambiguity tube MPC.

4.1. Linear stochastic systems

Let us consider linear stochastic discrete-time systems with (projected) linear right-hand function

$$f = \text{proj}_X \circ \hat{f} \qquad \text{with} \qquad \hat{f}(x, u, w) = Ax + Bu + w \ ,$$

with given $A \in \mathbb{R}^{n_x \times n_x}$ and $B \in \mathbb{R}^{n_x \times n_u}$. We assume that an asymptotically stabilizing linear feedback gain $K \in \mathbb{R}^{n_u \times n_x}$ for (A, B) exists; that is, such that all eigenvalues of $A + BK$ are in the open unit disc. For simplicity of presentation, we focus on parametric ancillary feedback controllers of the form

$$\mu[u_c](x) = u_c + Kx \ , \quad \text{where} \quad \mathcal{U} = \{ \ \mu[u_c] \ | \ u_c \in U_c \ \}$$

denotes the compact set of representable ancillary control laws for the pre-computed control gain K. In this context, $U_c \in \mathbb{K}(\mathbb{R}^{n_u})$ denotes a control constraint that is associated with the central control offset u_c. Similarly, we introduce the notation

$$\omega[w_c](X') = \overline{\omega}(\{w_c\} \oplus X') \ \text{ and } \ \Omega = \{\omega[w_c] \mid w_c \in W_c\}$$

for all $X' \in \mathcal{B}(X)$, assuming that $\overline{\omega} \in \mathcal{P}(\overline{W})$, the central set $W_c \in \mathbb{K}(\mathbb{R}^{n_x})$, and the domain $\overline{W} \in \mathbb{K}(\mathbb{R}^{n_x})$ are given. Finally, the corresponding domain of the uncertainty sequence is denoted by

$$W \stackrel{\text{def}}{=} W_c \oplus \overline{W} \ .$$

At this point, there are two remarks in order.

Remark 7. Because K is a pre-stabilizing feedback, one can construct the compact domain $X \in \mathbb{K}(\mathbb{R}^{n_x})$ such that

$$X \supseteq (A + BK)X \oplus [BU_c \oplus W] \ .$$

This construction is such that the closed-loop uncertainty measure propagation operator Φ is unaffected by the projection onto X—it would have been the same, if we would have set $f = \hat{f}$. This illustrates how Assumption 1 can – at least for asymptotically stabilizable linear systems – formally be satisfied by a simple projection onto an invariant set, without altering the original physical problem formulation.

Remark 8. Notice that the ambiguity set Ω in the above system model can be used to overestimate nonlinear terms. For example, if we have a system of the form

$$x^+ = Ax + Bu + g(x) + \overline{w} \ ,$$

where $g(x)$ is bounded on X, such that $g(x) \in W_c$ while $\overline{w}$ is a random variable with probability measure $\overline{\omega}$, then the ambiguity set Ω is such that the probability measure ω_g of the random variable $w = g(x) + \overline{w}$ satisfies $\{\omega_g\} \preceq \Omega$. This example can be used as a starting point to develop computationally tractable ambiguity tube MPC formulations for nonlinear systems, although a discussion of less conservative nonlinearity bounders, as developed for set propagation in Villanueva et al. (2017), are beyond the scope of this paper.

4.2. Construction of stage and terminal costs

In order to discuss how to design stage and terminal costs, L and M, which satisfy the requirements from Assumptions 2 and 3, this section focuses on the case that the stage cost has the form

$$L(P, \mu) = \max_{p \in P} \int \Theta(x, u_c)\, p(\mathrm{d}x) \ , \tag{25}$$

where $\Theta \in \mathcal{L}(X \times U_c, \mathbb{R}_+)$ is a non-negative and Lipschitz continuous control performance function. In the following, we assume that we have $0 \in U_c$ as well $\Theta(y^\star, 0) = 0$ for all $y^\star \in Y^\star$, where $Y^\star$ denotes the limit set of the considered linear stochastic system that is obtained for the offset-free ancillary control law; that is,

$$Y^\star \ = \ \bigoplus_{k=0}^{\infty} (A + BK)^k \, W \ .$$

Notice that this set can be computed by using standard methods from the field of set based computing (Blanchini & Miani, 2015; Houska & Villanueva, 2019).

Example 6. Let us assume that $\mathbb{X} \in \mathbb{K}(\mathbb{R}^{n_x})$ is a given state constraint with $Y^\star \subseteq \mathbb{X}$. In this case, the risk and performance measure

$$\Theta(x, u) = \|u\|_2^2 + \text{dist}_2(x, Y^\star)^2 + \tau \cdot \text{dist}_1(x, \mathbb{X})$$

is Lipschitz continuous. It can be used to model a trade-off between the least-squares control performance term

$$\|u\|_2^2 + \text{dist}_2(x, Y^\star)^2$$

that penalizes control inputs and the distance of the state to the target region $Y^\star$, and the constraint violation penalty $\tau \cdot \text{dist}_1(x, \mathbb{X})$. Here, $\tau > 0$ is a tuning parameter that can be used to adjust how risk-averse the controller is; see also Examples 3 and 4.

The key idea for constructing the terminal cost M is to first construct a non-negative function $\Pi \in \mathcal{L}(X, \mathbb{R}_+)$ that satisfies the ancillary Lyapunov descent condition

$$\forall x \in X, \quad \Theta(x, 0) + \max_{w \in W} \Pi(f(x, Kx, w)) \ \leq \ \Pi(x). \tag{26}$$

Notice that such a function Π exists, as we assume that the closed-loop system matrix $(A+BK)$ is asymptotically stabilizing. It can be constructed as follows. Let Σ denote the positive definite solution of the algebraic Lyapunov equation

$$(A + BK)^\intercal \Sigma (A + BK) + I = \Sigma,$$

let $|\lambda_{\max}(A + BK)| \leq \overline{\lambda} < 1$ be an upper bound on the spectral radius of the matrix $A + BK$, and let $\Lambda > 0$ be the Lipschitz constant of Θ with respect to the weighted Euclidean norm, $\|x\|_\Sigma \stackrel{\text{def}}{=} \sqrt{x^\intercal \Sigma x}$, such that

$$\forall x \in X, \qquad \Theta(x, 0) \ \leq \ \Lambda \min_{x' \in Y^\star} \|x - x'\|_\Sigma \ .$$

Next, we claim that the function

$$\forall x \in X, \qquad \Pi(x) \ \stackrel{\text{def}}{=} \ \frac{\Lambda}{1 - \overline{\lambda}} \left[\min_{x' \in Y^\star} \|x - x'\|_\Sigma \right]$$

satisfies (26). In order to prove this, notice that the inequality

$$\forall x \in X, \qquad \max_{w \in W} \Pi(f(x, Kx, w)) \ \leq \ \overline{\lambda} \cdot \Pi(x)$$

holds by construction of Σ, Π and $Y^\star$. Thus, we have

$$\begin{aligned}
&\Theta(x,0) + \max_{w\in W} \Pi(f(x,Kx,w)) \\
&\quad \leq \Lambda \min_{x'\in Y^\star} \|x-x'\|_\Sigma + \bar{\lambda}\cdot \Pi(x) \\
&\quad = \left[\Lambda + \frac{\Lambda\bar{\lambda}}{1-\bar{\lambda}}\right] \min_{x'\in Y^\star} \|x-x'\|_\Sigma = \Pi(x)
\end{aligned} \tag{27}$$

for all $x \in X$; that is, Π satisfies (26). Next, the associated ambiguity measure

$$M(P) = \max_{p\in P} \int \Pi \, \mathrm{d}p \tag{28}$$

can be used as an associated terminal cost that satisfies Assumption 3. As this result is of high practical relevance, we summarize it in the form of the following lemma.

Lemma 3. *Let L and M be defined as in (25) and (28). If the functions Θ and Π are non-negative and Lipschitz continuous such that (26) is satisfies and if $0 \in U_c$, then L and M satisfy all requirements from Assumptions 2 and 3.*

Proof. Because Θ and Π are Lipschitz continuous and non-negative, L and M are, by construction, proper ambiguity measures and non-negative, too. Thus, Assumption 2 is satisfied. Moreover, (28) implies that

$$\begin{aligned}
M(F(P,\mu)) &\overset{(28)}{=} \max_{p\in F(P,\mu)} \int \Pi \, \mathrm{d}p \\
&= \max_{p\in P, \omega\in\Omega} \iint \Pi(f(x,Kx,w))\, p(\mathrm{d}x)\, \omega(\mathrm{d}w) \\
&\leq \max_{p\in P} \int \max_{w\in W} \Pi(f(x,Kx,w))\, p(\mathrm{d}x) \,,
\end{aligned} \tag{29}$$

where the second equation holds for the offset-free ancillary feedback law $\mu(x) = Kx$. Furthermore, according to (26), we have

$$\max_{w\in W} \Pi(f(x,Kx,w)) \;\leq\; \Pi(x) - \Theta(x,0) \tag{30}$$

for all $x \in X$. Consequently, we can substitute this inequality in (29) finding that

$$\begin{aligned}
M(F(P,\mu)) &\overset{(29),(30)}{\leq} \max_{p\in P} \int [\Pi - \Theta(\cdot,0)] \, \mathrm{d}p \\
&\overset{(28),(25)}{=} M(x) - L(P,\mu) \,.
\end{aligned}$$

In other words, because we assume that $0 \in U_c$, there exists for every $P \in \mathcal{K}(X)$ a $\mu \in \mathcal{U}$ for which

$$L(P,\mu) + M(F(P,\mu)) \;\leq\; M(P)$$

and the conditions from Assumption 3 are satisfied. This corresponds to the statement of the lemma. □

Remark 9. Notice that many articles on stochastic MPC, for example Chatterjee and Lygeros (2015), Kouvaritakis and Cannon (2016) and Mayne (2015), start their construction of the stage cost by assuming that a nominal (non-negative) cost function $l : X \times U \to \mathbb{R}_+$ is given. For example, in the easiest case, one could consider the least-squares cost

$$l(x,u) = x^2 + u^2 \,.$$

In the above context, however, we cannot simply set $\Theta = l$, as Condition (26) can only be satisfied if we have $\Theta(y^\star, 0) = 0$ for all $y^\star \in Y^\star$—but $Y^\star$ is usually not a singleton. However, one can find a Lipschitz continuous function Θ that approximates the function

$$\Theta(x,u) \approx \begin{cases} l(x,u) & \text{if } x \notin Y^\star \\ 0 & \text{if } x \in Y^\star \end{cases}$$

up to any approximation accuracy such that Θ coincides with l on the domain $X \setminus Y^\star$ with high precision. In fact, the approximation is in this context only needed for technical reasons, such that Θ is Lipschitz continuous. Notice that this construction satisfies the requirements of Lemma 3 and is, as such, fully compatible with our stability analysis framework. Next, we construct the hybrid feedback law

$$\tilde{\mu}(y) \overset{\text{def}}{=} \begin{cases} \mu_{\text{MPC}}(y) & \text{if } y \notin Y^\star \\ Ky & \text{if } y \in Y^\star \end{cases},$$

which simply switches to the ancillary control law $x \to Kx$ whenever the current state is already inside the target region $Y^\star$. This construction is compatible with the stability statements from Theorems 3 and 4, as we modify the closed-loop system only inside the robust control invariant target region. Of course, this is in the understanding that the control gain K is optimized beforehand and that this linear controller leads to a close-to-optimal control performance (with respect to worst-case expected value of the given cost function l) inside the region $Y^\star$—if not, one needs to work with more sophisticated ancillary controllers and redefine $\mathcal{U}$. The robust MPC controller is in this case only taking care of the case that the current state is in $X \setminus Y^\star$—but in this region Θ coincides with the given cost function l as desired.

4.3. Implementation details

Ambiguity tube MPC can be implemented by pre-computing the stage and terminal cost offline. This has the advantage that, in the online phase, a simple convex optimization problem is solved. For this aim, we pre-compute the central sets

$$Z_{k+1} \overset{\text{def}}{=} \bigoplus_{i=0}^{k} (A+BK)^i W_c \tag{31}$$

for all $k \in \{0,1,\ldots,N-1\}$ by using standard set computation techniques (Blanchini & Miani, 2015). Similarly, by introducing the Markovian kernel

$$\forall X' \in \mathcal{B}(X), \quad \Delta[x](X') \overset{\text{def}}{=} \omega(X' \oplus \{-(A+BK)x\}) \,,$$

we can pre-compute offset-free measures q_k via the Markovian recursion

$$\begin{aligned}
\forall k \in \mathbb{N}, \quad q_{k+1} &= \int \Delta[x]\, q_k(\mathrm{d}x) \\
q_0 &= \delta_0 \,.
\end{aligned} \tag{32}$$

For example, if ω denotes a uniform probability measure with compact zonotopic support, the measures q_k can be computed with high precision by using a generalized Lyapunov recursion in combination with a Gram–Charlier expansion (Villanueva & Houska, 2020). After this preparation, we can pre-compute Chebyshev representations of the functions

$$S_k(z,u) = \max_{z_c\in Z_k} \int \Theta(z+z_c+\bar{z},u)\, q_k(\mathrm{d}\bar{z})$$

and $$S_N(z) = \max_{z_c\in Z_N} \int \Pi(z+z_c+\bar{z})\, q_N(\mathrm{d}\bar{z})$$

with high accuracy, as discussed in Villanueva and Houska (2020), too. Here, the function Θ and Π are constructed as in the previous section. In particular, if Θ is convex, as in Example 6, S_k is convex. Similarly, if Π is convex, S_N is convex. Finally, the associated online optimization problem,

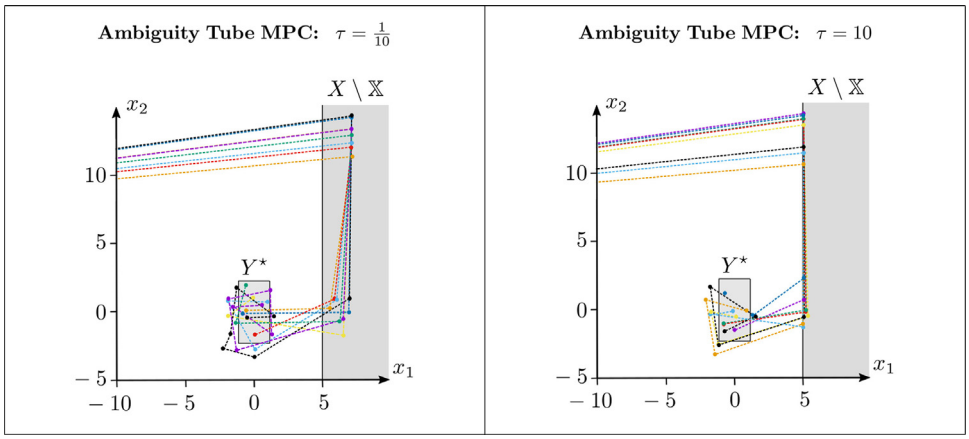


Fig. 1. Randomly generated closed-loop scenarios of the ambiguity tube MPC controller (33) with (LEFT) $\tau = 10^{-1}$ and (RIGHT) $\tau = 10$. The state constraint, $x_1 \leq 5$, is visualized in the form of the dark gray-shaded infeasible region $x_1 > 5$. For the very small penalty parameter $\tau = \frac{1}{10}$ the controller risks marginal constraint violation for the sake of better nominal control performance. All trajectories converge to the light gray-shaded target region $Y^\star$. Notice that the initial state, $y_0 = [-60, 5]^\intercal$, would be far off to the left of the figure and it is therefore not visualized—the colored dots correspond to the discrete-time states after the first MPC iteration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\mathcal{V}(y) = \min_{v,z} \sum_{k=0}^{N-1} S_k(z_k, v_k) + S_N(z_N)$$
$$\text{s.t.} \begin{cases} \forall k \in \{0, 1, \ldots, N-1\}, \\ z_{k+1} = Az_k + Bv_k \\ z_0 = y \,, \end{cases} \tag{33}$$

can be solved with existing MPC software. The associated ambiguity tube MPC feedback law has then, by construction, the form

$$\mu_{\text{MPC}}(y) = v_0^\star(y) \,,$$

where $v_0^\star(y)$ denotes the first element of an optimal control input sequence of (33) as a function of y. This construction can be further refined by implementing the hybrid control law from Remark 9.

4.4. Numerical illustration

This section illustrates the performance of the above ambiguity tube MPC controller for the nonlinear system

$$x_1^+ = -\frac{x_1}{8} + x_2 + u_1 + w_1,$$
$$x_2^+ = -\frac{x_1}{2} + \frac{x_2}{4} + u_2 + \cos(x_1)\sin(5x_2)^3 + w_2 \,.$$

In order to write this system in the above form, we introduce the notation

$$A \stackrel{\text{def}}{=} \begin{pmatrix} -\frac{1}{8} & 1 \\ -\frac{1}{2} & \frac{1}{4} \end{pmatrix}, \quad B \stackrel{\text{def}}{=} I \quad \text{and} \quad K \stackrel{\text{def}}{=} \begin{pmatrix} \frac{1}{8} & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} \end{pmatrix}$$

to denote the system matrices and a suitable ancillary control gain. Notice that the influence of the nonlinear term, $\cos(x_1)\sin(5x_2)^3$, can be over-estimated by introducing the central set $W_c \stackrel{\text{def}}{=} \{0\} \times [-1, 1]$ (see Remark 8). Additionally, we set $U_c \stackrel{\text{def}}{=} [-10, 10]^2$. Moreover, the uncertain input is modeled by the distribution measure

$$\omega(A) \stackrel{\text{def}}{=} \int_A \rho(w) \mathrm{d}w$$

with Radon–Nikodyn derivative (density function)

$$\rho(w) \stackrel{\text{def}}{=} \begin{Bmatrix} \mathfrak{d}(w_1) & \text{if } w_2 \in [-1, 1] \\ 0 & \text{otherwise} \end{Bmatrix},$$

where $\mathfrak{d} \stackrel{\text{def}}{=} \partial\delta/\partial w$ denotes the standard Dirac distribution. The objective is constructed as in Example 6,

$$\Theta(x, u) \stackrel{\text{def}}{=} \|u\|_2^2 + \text{dist}_2(x, Y^\star)^2 + \tau \cdot \text{dist}_1(x, \mathbb{X}) \,,$$

where $\tau > 0$ is a risk-parameter that is associated with the given state constraint set

$$\mathbb{X} \stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^2 \,\middle|\, x_1 \leq 5 \right\} .$$

Here, it is not difficult to check that the function

$$\Pi(x) \stackrel{\text{def}}{=} \left[2 + \frac{7}{27} \cdot \tau \right] \text{dist}_2(x, Y^\star)^2$$

satisfies the requirements from Lemma 3. Consequently, the associated ambiguity measures L and M satisfy all technical requirements of Theorem 4. Notice that the functions S_k in (33) are in our implementation pre-computed with high precision such that the convex optimization problem (33) can be solved in much less than 1 ms by using ACADO Toolkit (Houska, Ferreau, & Diehl, 2011). The prediction horizon of the MPC controller is set to $N = 5$.

Fig. 1 shows two ambiguity tube MPC closed-loop simulations that are both started at the initial point $y_0 = [-60, 5]^\intercal$. In the left figure, we have set $\tau = \frac{1}{10}$, which means that the constraint violation penalty is small compared to the nominal control performance objective. Consequently, during the randomly generated closed-loop scenarios marginal constraint violation can be observed. This is in contrast to the right part of Fig. 1, which shows randomly generated closed-loop scenarios for the case $\tau = 10$, leading to much smaller expected constraint violations at risk. In all cases, that is, independent of how the penalty parameter $\tau > 0$ is chosen and independent of the particular uncertainty scenario, the closed-loop trajectories converge to the terminal region $Y^\star$ after a short transition period. In this particular example, we observe that this happens typically after 3 to 8 discrete-time steps, which confirms the robust asymptotic convergence statement of Theorem 4.

5. Conclusion

This paper has presented a coherent measure-theoretic framework for analyzing the stability of a rather general class of ambiguity tube MPC controllers. In detail, we have proposed a Wasserstein–Hausdorff metric leading to our first main result in

Theorem 1, where conditions for the existence of a continuous value function of ambiguity tube MPC controllers have been established. Moreover, Theorem 2 has built upon this topological framework to establish conditions under which the stage and terminal cost are proper ambiguity measures, such that the cost function of the MPC controller can be turned into a non-negative supermartingale along the trajectories of the stochastic closed-loop system. Related stochastic stability and convergence results for ambiguity tube MPC have been summarized in Theorems 3 and 4.

In the sense that Lemma 3 proposes a practical strategy for constructing stabilizing terminal costs for stochastic and ambiguity tube MPC, the current article has outlined a path towards a more consistent stability theory that goes much beyond the existing convergence results from Chatterjee and Lygeros (2015), Kouvaritakis and Cannon (2016) and Munoz-Carpintero and Cannon (2020). At the same time, however, it should also be pointed out that these results are based on a slightly different strategy of modeling the stage cost of the MPC controller, as discussed in Remark 9, where it is also explained why it may be advisable to use a hybrid feedback control law that switches to a pre-optimized ancillary controller whenever the state is inside its associated target region $Y^\star$.

Last but not least, as much as this article has been attempting to make a step forward, towards a more consistent stability theory and practical formulation of robust MPC, it should also be stated that many open problems and conceptual challenges remain. In the line of this paper, for instance, a discussion of more advanced representations of ambiguity set representations and handling of nonlinearities, a more in-depth analysis of the interplay of the choice of $\mathcal{U}$, the performance of the ambiguity tube controller, and its computational tractability, bounds on the concentrations of the state distributions, economic objectives for ambiguity tube MPC, as well as a deeper analysis of risk measures and related problems regarding recursive feasibility are only a small and incomplete selection of open problems in the field of distributionally robust MPC.

References

Berge, C. (1963). *Topological spaces*. Oliver and Boyd.

Bernardini, D., & Bemporad, A. (2012). Stabilizing model predictive control of stochastic constrained linear systems. *IEEE Transactions on Automatic Control*, *57*(6), 1468–1480.

Blanchini, F., & Miani, S. (2015). *Set-theoretic methods in control*. Birkhäuser.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Bucy, R. S. (1965). Stability and positive supermartingales. *Journal of Differential Equations*, *1*, 151–155.

Chatterjee, D., & Lygeros, J. (2015). On stability and performance of stochastic predictive control techniques. *IEEE Transactions on Automatic Control*, *60*, 509–514.

Chen, H., & Allgöwer, F. (1998). A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica*, *34*(10), 1205–1217.

Clarke, F., Ledyaev, Y., & Stern, R. (1998). Asymptotic stability and smooth Lyapunov functions. *Journal of Differential Equations*, *149*(1), 69–114.

Doob, J. L. (1953). *Stochastic processes*. Wiley.

Feller, W. (1971). *Introduction to probability theory and its applications*. Wiley.

Grimm, M. J., Messina, G., Tuna, S. E., & Teel, A. R. (2004). Examples when nonlinear model predictive control is nonrobust. *Automatica*, *40*, 1729–1738.

Grüne, L. (2009). Analysis and design of unconstrained nonlinear MPC schemes for finite and infinite dimensional systems. *SIAM Journal on Control and Optimization*, *48*(2), 1206–1228.

Hewing, L., Wabersich, K. P., Menner, M., & Zeilinger, M. N. (2020). Learning-based model predictive control: Toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, *3*, 269–296.

Houska, B., Ferreau, H. J., & Diehl, M. (2011). An auto-generated real-time iteration algorithm for nonlinear MPC in the microsecond range. *Automatica*, *47*, 2279–2285.

Houska, B., & Villanueva, M. E. (2019). Robust optimzation for MPC. In S. Raković, & W. Levine (Eds.), *Handbook of model predictive control* (pp. 413–443). Birkäuser.

Kantorovich, L. V. (2006). On the translocation of masses. *Journal of Mathematical Sciences*, *133*(4), 1381–1382.

Kouvaritakis, B., & Cannon, M. (2015). *Model predictive control: Classical, robust and stochastic*. Springer.

Kouvaritakis, B., & Cannon, M. (2016). Feasibility, stability, convergence and Markov chains. In *Advanced textbooks in control and signal processing, Model predictive control* (pp. 271–301). Springer.

Kushner, H. J. (1965). On the stability of stochastic dynamical systems. *Proceedings of the National Academy of Sciences*, *53*(1), 8–12.

Kushner, H. J. (2014). A partial history of the early development of continuous-time nonlinear stochastic systems theory. *Automatica*, *50*(2), 303–334.

Langson, W., Chryssochoos, I., Raković, S. V., & Mayne, D. Q. (2004). Robust model predictive control using tubes. *Automatica*, *40*(1), 125–133.

Ledyaev, Y., & Sontag, E. (1999). A Lyapunov characterization of robust stabilization. *Nonlinear Analysis*, *37*, 813–840.

Lorenzen, M., Dabbene, F., Tempo, R., & Allgöwer, F. (2016). Constraint-tightening and stability in stochastic model predictive control. *IEEE Transactions on Automatic Control*.

Mayne, D. Q. (2015). Robust and stochastic MPC: are we going in the right direction? *IFAC-PapersOnLine*, *48*(23), 1–8.

Mayne, D. Q., Seron, M. M., & Raković, S. (2005). Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, *41*(2), 219–224.

Mesbah, A. (2016). Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control Systems Magazine*, *36*(6), 30–44.

Munoz-Carpintero, D., & Cannon, M. (2020). Convergence of stochastic nonlinear systems and implications for stochastic model predictive control. *IEEE Transactions on Automatic Control*, 1–8, (early access).

Raković, S., Kouvaritakis, B., Findeisen, R., & Cannon, M. (2012). Homothetic tube model predictive control. *Automatica*, *48*(8), 1631–1638.

Rawlings, J. B., Mayne, D. Q., & Diehl, M. M. (2018). *Model predictive control: Theory and design*. Madison, WI: Nob Hill Publishing.

Rockafellar, R. T., & Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, *18*, 33–53.

Rockafellar, R. T., & Wets, R. J. (2005). *Variational analysis*. Springer.

Sehr, M. A., & Bitmead, R. R. (2018). Stochastic output-feedback model predictive control. *Automatica*, *94*, 315–323.

Sopasakis, P., Herceg, D., Bemporad, A., & Patrinos, P. (2019). Risk-averse model predictive control. *Automatica*, *100*, 281–288.

Taylor, J. C. (1996). *An introduction to measure and probability*. Springer.

Van Parys, B., Kuhn, D., Goulart, P., & Morari, M. (2016). Distributionally robust con- trol of constrained stochastic systems. *IEEE Transactions on Automatic Control*, *61*(2), 430–442.

Vasershtein, L. N. (1969). Markov processes over denumerable products of spaces describing large system of automata. *Problemy Peredači Informacii*, *5*(3), 64–72.

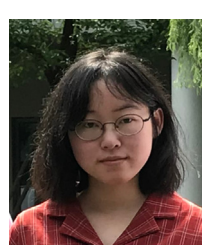
Villani, C. (2005). *Optimal transport, old and new*. Springer.

Villanueva, M. E., De Lazzari, E., Müller, M. A., & Houska, B. (2020). A set-theoretic generalization of dissipativity with applications in Tube MPC. *Automatica*, *122*(109179).

Villanueva, M. E., & Houska, B. (2020). On stochastic linear systems with zonotopic support sets. *Automatica*, *111*(108652).

Villanueva, M. E., Quirynen, R., Diehl, M., Chachuat, B., & Houska, B. (2017). Robust MPC via min–max differential inequalities. *Automatica*, *77*, 311–321.

Zanon, M., & Gros, S. (2021). Safe reinforcement learning using robust MPC. *IEEE Transactions on Automatic Control*, *66*(8), 3638–3652.



Fan Wu is a graduate student at the School of Information Science and Technology at ShanghaiTech University. She has received bachelor's degrees in mathematics and law from East China University of Science And Technology in 2019. Her research interests include model based learning and control algorithms.

Mario Eduardo Villanueva is an assistant professor at the IMT School for Advanced Studies Lucca. He received a master and Ph.D. in chemical engineering from Imperial College London in 2011 and 2016, respectively. He was a postdoctoral researcher at Texas A&M University in 2016 and at the School of Information Science and Technology at ShanghaiTech University from 2017 to 2022. Mario E. Villanueva is the recipient of the 2016 Dudley Newitt Price and the 2018 as well as 2019 SIST Excellent Postdoc Award. His research interests include set based computing, robust and stochastic control, global optimization and system theoretic tools for data analysis.

Boris Houska is an associate professor at the School of Information Science and Technology at ShanghaiTech University. He received a diploma in mathematics and physics from the University of Heidelberg in 2007, and a Ph.D. in Electrical Engineering from KU Leuven in 2011. From 2012 to 2013 he was a postdoctoral researcher at the Centre for Process Systems Engineering at Imperial College London. His research interests include numerical optimization and optimal control, robust and global optimization, as well as fast model predictive control algorithms.