

On The Inherent Distributional Robustness of Stochastic and Nominal Model Predictive Control

Robert D. McAllister and James B. Rawlings

Abstract—We define a notion of distributional robustness, via the Wasserstein metric, for closed-loop systems subject to errors in the disturbance distribution used to construct the controller. We then establish sufficient conditions for stochastic model predictive control (SMPC) to satisfy this definition of distributional robustness and establish a similar notion of distributional robustness for economic applications of SMPC. These results address incorrectly or unmodeled disturbances, demonstrate the efficacy of scenario optimization as a means to approximate and solve the SMPC problem, and unify the descriptions of robustness for stochastic and nominal model predictive control. This definition of distributional robustness for closed-loop systems is general and can be applied to other stochastic optimal control algorithms and, potentially, the developing field of distributionally robust control.

Index Terms—Model predictive control, stochastic optimal control, distributional robustness, stochastic systems

I. INTRODUCTION

In practice, a control algorithm must ensure some degree of robustness to disturbances and modeling errors for successful industrial implementation. By virtue of feedback, nominal model predictive control (MPC) is known to be inherently robust to sufficiently small disturbances [1, 11, 30]. We use the term *inherent* robustness to indicate that this robustness is achieved through feedback and without explicitly considering the disturbances in the problem formulation.

Stochastic MPC (SMPC) offers a means to improve on the inherent robustness of nominal MPC by including a stochastic description of the disturbance directly in the problem formulation. In general, the SMPC optimization problem minimizes the expected value of the cost function subject to deterministic and probabilistic constraints [8, 16, 20]. Using the same rolling horizon approach as MPC, this stochastic optimization problem is solved at each sample time, with an updated state estimate, to determine the control action at that time. We focus this article on the closed-loop properties of nonlinear SMPC.

We briefly review some contributions to the study of closed-loop properties of SMPC. Primbs and Sung [21] use a global stochastic Lyapunov function as the terminal cost to establish that the origin is asymptotically stable with probability one for linear systems with multiplicative disturbances that vanish as the origin. Cannon, Kouvaritakis, and co-authors use a terminal constraint and *local* Lyapunov function to ensure recursive feasibility and stability in expectation for linear

systems with bounded disturbances (that do not vanish at the origin) and input/state constraints [4–6, 14]. Lorenzen et al. [15] propose a less restrictive constraint tightening approach and establish that linear SMPC asymptotically stabilizes, with probability one, the minimal robust positive invariant set of the system. For *nonlinear* SMPC, Chatterjee and Lygeros [7] uses a global stochastic Lyapunov function to establish, for unconstrained systems, that the expected value of the optimal cost is bounded along the closed-loop trajectory. Mayne and Falugi [17] use a terminal constraint and a *local* Lyapunov function to extend the results in [7] to systems with bounded inputs and disturbances. In [19], the authors establish that SMPC renders the closed-loop system robustly asymptotic stable in expectation (RASiE).

Although these results are interesting and instructive, the key assumption made in SMPC closed-loop analysis, and indeed much of stochastic optimal control in general, is that the stochastic description of uncertainty used in the SMPC optimization problem is exact and comprehensive. While this assumption is a reasonable starting point, it does not hold for any practical implementation of SMPC. Much like the nominal model identified for nominal MPC, we cannot expect that a stochastic model, typically identified from data, is exact.

In this work, we remove this assumption of exact disturbance models and distributions. Specifically, we address an open question in the field of SMPC that is of significant practical concern: What, if any, robustness does SMPC confer for unmodeled or incorrectly modeled disturbances? This question is asking about the *distributional robustness* of SMPC, i.e., the robustness of SMPC to errors in the probability distribution used in the problem formulation. The unwritten hypothesis is that feedback provides some margin of inherent distributional robustness to SMPC and thereby addresses any small discrepancies in the disturbance model. This hypothesis, however, has never been established for SMPC.

To address this question of distributional robustness, we require a notion of distance between probability measures, i.e., a probability metric. For this task, we select the Wasserstein metric. While this metric initially gained popularity in the field of optimal transport [28], there are several recent applications of the Wasserstein metric in machine learning [2, 10], state estimation [26], and optimal control [29]. In contrast to these approaches, we do not use the Wasserstein metric in the formulation of the SMPC optimization problem. Instead, we use the Wasserstein metric only as a means to quantify the distance between the true disturbance distribution and the model disturbance distribution used in the SMPC optimization problem.

This work was supported by the National Science Foundation (NSF) under Grant 2027091. (Corresponding author: R. D. McAllister)

R. D. McAllister and J. B. Rawlings are with the Department of Chemical Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: rdmcallister@ucsb.edu, jbraw@ucsb.edu)

We summarize the paper as follows. In Section II, we define the system of interest and introduce the SMPC problem formulation and associated assumptions. In Section III, we introduce the Wasserstein metric and a few important results for this metric. In Section IV, we define distributionally robust asymptotic stability in expectation (DRASiE) and establish a sufficient condition for this form of robustness via a stochastic input-to-state stable (SISS) Lyapunov function. In Section V, we establish that SMPC is inherently distributionally robust in terms of DRASiE and establish an alternative result for economic applications of SMPC, i.e., problems in which the stage cost is not necessarily positive definite with respect to the origin. In Section VI, we discuss several insights for both stochastic and nominal MPC afforded by these results. We provide a numerical example to illustrate these results in Section VII.

Notation: Let $\mathbb{I}$ and $\mathbb{R}$ denote the integers and reals. Let superscripts and subscripts denote dimension and restrictions (e.g., $\mathbb{R}_{\geq 0}^n$ denotes nonnegative reals of dimension n). Let $|\cdot|$ denote Euclidean norm and $|x|_Y := \inf_{y \in Y} |x - y|$ denote Euclidean point-to-set distance. The function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is in class $\mathcal{K}$ if it is continuous, strictly increasing, and $\alpha(0) = 0$. The function $\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is in class $\mathcal{K}_\infty$ if $\alpha(\cdot) \in \mathcal{K}$ and unbounded, i.e., $\lim_{s \rightarrow \infty} \alpha(s) = \infty$. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{I}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is in class $\mathcal{KL}$ if for every $k \in \mathbb{I}_{\geq 0}$ the function $\beta(\cdot, k)$ is in class $\mathcal{K}$ and for fixed $s \in \mathbb{R}_{\geq 0}$ the function $\beta(s, \cdot)$ is nonincreasing and $\lim_{k \rightarrow \infty} \beta(s, k) = 0$. Let $\mathcal{B}(\Omega)$ denote the Borel algebra of some set Ω . Let $\text{tr}(A)$ denote the trace of a matrix A . Let $\delta_x(A)$ denote the Dirac measure defined for a given $x \in \mathbb{R}^n$ and set $A \subseteq \mathbb{R}^n$, i.e., $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise.

II. PROBLEM FORMULATION AND PRELIMINARIES

A. The stochastic system(s)

We consider the following discrete-time stochastic system

$$x^+ = f(x, u, w) \quad f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}^n \quad (1)$$

in which $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the controlled input, $w \in \mathbb{W} \subseteq \mathbb{R}^q$ is the stochastic disturbance, and x^+ is the successor state. Let $(\Omega, \mathcal{F}, P)$ denote the probability space for the sequence $\mathbf{w}_\infty : \Omega \rightarrow \mathbb{W}^\infty$. We denote a subsequence of $\mathbf{w}_\infty$ as $\mathbf{w}_i := (w(0), w(1), \dots, w(i-1))$ and define expected value of a Borel measurable function $g : \mathbb{W}^i \rightarrow \mathbb{R}$ as the Lebesgue integral

$$\mathbb{E}[g(\mathbf{w}_i)] := \int_{\Omega} g(\mathbf{w}_i(\omega)) dP(\omega) \quad (2)$$

We then consider the following standard assumption.

Assumption 1. The random variables $w(i) : \Omega \rightarrow \mathbb{W}$ are independent and identically distributed (i.i.d.) in time and $\mathbb{W}$ is compact.

Given Assumption 1, each $w(i)$ has an equivalent probability measure that we denote $\mu : \mathcal{B}(\mathbb{W}) \rightarrow [0, 1]$, defined such that $\mu(F) = P(\{\omega \in \Omega : w(i; \omega) \in F\})$ for all $F \in \mathcal{B}(\mathbb{W})$ and $i \in \mathbb{I}_{\geq 0}$. We use $\mathcal{M}(\mathbb{W})$ to denote the collection of all probability measures on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$. All moments of w are

finite for $\mu(\cdot) \in \mathcal{M}(\mathbb{W})$, i.e., $\int |w|^p d\mu(w) < \infty$ for all $p \in \mathbb{I}_{\geq 1}$, because $\mathbb{W}$ is compact. Note that $\mathbb{W}$ is larger than, but not necessarily equal to, the support of μ . Therefore, $\mu(\mathbb{W}) = 1$ but there may exist a nonempty set $S \subseteq \mathbb{W}$ such that $\mu(S) = 0$.

Since we intend to consider problems with bounded inputs, we require that $\mathbb{W}$ is bounded (compact). Otherwise, the control problem is not well posed, i.e., we are attempting to reject an unbounded disturbance with a bounded input.

In this work, we do not assume that we know the set $\mathbb{W}$ or measure μ for the random variable w . Instead, we have access to only a *model* of the set $\mathbb{W}$ and probability measure μ , that we denote $\hat{\mathbb{W}}$ and $\hat{\mu}$, respectively. Note that we may assume $\hat{\mathbb{W}} \subseteq \mathbb{W}$ without loss of generality because we can always increase the size of $\mathbb{W}$ to fit $\hat{\mathbb{W}}$ and assign these additional values measure zero with μ . We may also, without loss of generality, define $\hat{\mu}$ on the domain $\mathcal{B}(\mathbb{W})$ and assign zero measure to all the points in $\mathbb{W}$ that are not in $\hat{\mathbb{W}}$, i.e., $\hat{\mu} : \mathcal{B}(\mathbb{W}) \rightarrow [0, 1]$ such that $\hat{\mu}(\mathbb{W} \setminus \hat{\mathbb{W}}) = 0$. Specifically, we have that

$$\int_{\mathbb{W}} g(\hat{w}) d\hat{\mu}(\hat{w}) = \int_{\hat{\mathbb{W}}} g(\hat{w}) d\hat{\mu}(\hat{w})$$

for all measurable functions $g(\cdot)$ because $\hat{\mu}(\mathbb{W} \setminus \hat{\mathbb{W}}) = 0$. Thus, $\hat{\mathbb{W}}$ includes the support of $\hat{\mu}$. Note that we allow for the possibility that $\hat{\mathbb{W}}$ is *larger* than the support of $\hat{\mu}$. We define $\hat{\mu}$ on the larger set $\mathbb{W}$ only to facilitate the comparison of μ and $\hat{\mu}$. In the SMPC optimization problem, the stochastic system evolves according to the following stochastic model and without knowledge of $\mathbb{W}$ or μ .

$$x^+ = f(x, u, \hat{w}) \quad \hat{w} \in \hat{\mathbb{W}} \quad (3)$$

in which $\hat{w}$ is distributed according to the measure $\hat{\mu}$.

We formalize these requirements of the disturbance model used in the SMPC formulation through the following assumption.

Assumption 2. The random variables $\hat{w}(i)$ are i.i.d. in time, with a probability measure $\hat{\mu} : \mathcal{B}(\mathbb{W}) \rightarrow [0, 1]$. The set $\hat{\mathbb{W}}$ is compact and contains the origin. The probability distribution satisfies $\hat{\mu}(\hat{\mathbb{W}}) = 1$.

We use $\hat{\mathcal{M}}(\mathbb{W})$ to denote the collection of all probability measures on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$ that satisfy Assumption 2, i.e., $\hat{\mu}(\hat{\mathbb{W}}) = 1$ for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. Note that $\hat{\mathcal{M}}(\mathbb{W}) \subseteq \mathcal{M}(\mathbb{W})$. We provide an illustration of these sets in Figure 1. We emphasize that the framework we have introduced is capable of representing incorrectly modeled ($\hat{\mu} \neq \mu$), unmodeled ($\hat{\mathbb{W}} \neq \mathbb{W}$), or out-of-sample ($\hat{\mathbb{W}}$ is finite) disturbances. We discuss each of these cases further in Section VI.

For the i.i.d. random variables $\hat{\mathbf{w}} = (\hat{w}(i), \hat{w}(i+1), \dots, \hat{w}(i+N-1))$ and $N \in \mathbb{I}_{\geq 1}$, their joint distribution measure $\hat{\mu}^N : \mathcal{B}(\mathbb{W}^N) \rightarrow [0, 1]$ is defined as $\hat{\mu}^N(F) = \hat{\mu}(F_i) \hat{\mu}(F_{i+1}) \dots \hat{\mu}(F_{i+N-1})$ for all $F = (F_i, F_{i+1}, \dots, F_{i+N-1}) \in \mathcal{B}(\mathbb{W}^N)$. For any Borel measurable function $g : \mathbb{W}^N \rightarrow \mathbb{R}$, we define expected value with respect to $\hat{\mu}$ as

$$\hat{\mathbb{E}}[g(\hat{\mathbf{w}})] = \int_{\mathbb{W}^N} g(\hat{\mathbf{w}}) d\hat{\mu}^N(\hat{\mathbf{w}}) = \int_{\hat{\mathbb{W}}^N} g(\hat{\mathbf{w}}) d\hat{\mu}^N(\hat{\mathbf{w}})$$

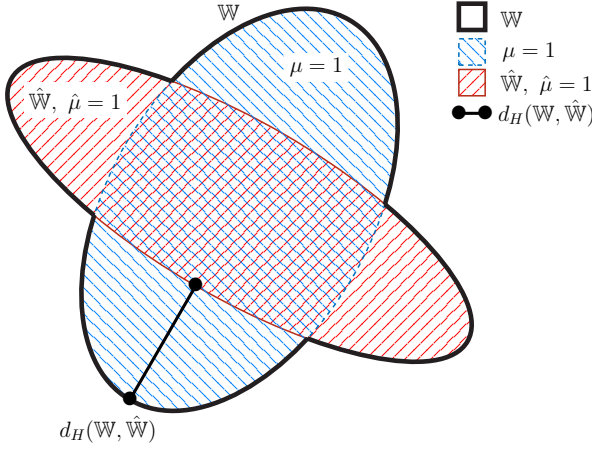


Fig. 1. Illustration of the sets $\mathbb{W}$, $\hat{\mathbb{W}}$ and the probability measures μ , $\hat{\mu}$. We also show the Hausdorff distance $d_H(\cdot)$ between these two sets.

in which the second equality holds because $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ and therefore $\hat{\mu}(\mathbb{W} \setminus \hat{\mathbb{W}}) = 0$. Note that we use $\hat{\mathbb{E}}[\cdot]$ to indicate expected value with respect to $\hat{\mu}$ instead of μ . We frequently use the expected value of $|\hat{w}|$ as a bounding quantity in subsequent analysis and note the following inequality

$$\hat{\mathbb{E}}[|\hat{w}|] \leq \sqrt{\text{tr}(\hat{\Sigma}) + |\hat{\mathbb{E}}[\hat{w}]|^2}$$

in which $\hat{\Sigma} \succeq 0$ is the covariance matrix of $\hat{w}$ and the upper bound can be constructed via Jensen's inequality.

Remark 1. Sampling-based approximation of μ may not produce empirical distributions with zero mean. We therefore do not restrict $\hat{\mu}$ to only measures of zero mean and instead leave $\hat{\mathbb{E}}[\hat{w}] = 0$ as a special case of the subsequent results.

Remark 2. We can extend these results to systems with time-varying probability measures (i.e., $w(i)$ and $\hat{w}(i)$ are independent but not identically distributed). This extension, however, requires that we consider a time-varying problem and adds additional notation to the following analysis. Given the introductory nature of this paper, we address only i.i.d. disturbances.

B. SMPC problem formulation

We now introduce the SMPC problem formulation. We assume that there is a fixed set $\hat{\mathbb{W}}$ used in the SMPC algorithm, as is typical in SMPC analysis. This requirement is important because the subsequent assumptions and sets for SMPC are based on a single set $\hat{\mathbb{W}}$ (e.g., Assumption 5). We do, however, allow for different $\hat{\mu}$ and derive bounds that apply for any such $\hat{\mu}$. Thus, the optimization problem and control law is subsequently defined as a function of $\hat{\mu}$.

We define the parameterized control policy $\pi : \mathbb{R}^n \times \mathbb{V} \rightarrow \mathbb{R}^m$ in which $x \in \mathbb{R}^n$ is the current state of the system and $v \in \mathbb{V} \subseteq \mathbb{R}^l$ are the parameters in the control policy, e.g., $\pi(x, v) = Kx + v$. Thus, the resulting system of interest is defined as

$$x^+ = f(x, \pi(x, v), \hat{w}) \quad \hat{w} \in \hat{\mathbb{W}} \quad (4)$$

in which $\hat{w}$ is distributed according to $\hat{\mu}$. We use $\hat{\phi}(k; x, \mathbf{v}, \hat{\mathbf{w}})$ to denote the solution of (4) at time k , given the initial condition $x \in \mathbb{R}^n$, the trajectory of control policy parameters $\mathbf{v} = (v(0), v(1), \dots, v(N-1)) \in \mathbb{V}^N$, and disturbance trajectory $\hat{\mathbf{w}} \in \hat{\mathbb{W}}^N$.

We consider the case of hard input constraints, i.e., $u \in \mathbb{U} \subseteq \mathbb{R}^m$. We do not, however, consider hard or probabilistic constraints on the state since we do not assume that the disturbance model is exact. Thus, a disturbance not included in $\hat{\mathbb{W}}$ may cause the closed-loop system to violate these state constraints. This fact is also true for nominal MPC [1, 30]. Instead, we assume that all state constraints are converted to exact penalty functions in the stage cost [13, 25, 31]. We do, however, include a terminal state constraint $\mathbb{X}_f \subseteq \mathbb{R}^n$ in the following SMPC problem that must satisfy specific requirements detailed in subsequent assumptions.

For a horizon of $N \in \mathbb{I}_{\geq 1}$, we denote the set of admissible control law parameter trajectories given $x \in \mathbb{R}^n$ as

$$\begin{aligned} \mathcal{V}(x) := \{ \mathbf{v} \in \mathbb{V}^N : \\ \pi(\hat{\phi}(k; x, \mathbf{v}, \hat{\mathbf{w}}), v(k)) \in \mathbb{U} \quad \forall \hat{\mathbf{w}} \in \hat{\mathbb{W}}^N, k \in \mathbb{I}_{[0, N-1]} \\ \hat{\phi}(N; x, \mathbf{v}, \hat{\mathbf{w}}) \in \mathbb{X}_f \quad \forall \hat{\mathbf{w}} \in \hat{\mathbb{W}}^N \} \end{aligned}$$

and the set of all feasible initial states is denoted

$$\mathcal{X} := \{x \in \mathbb{R}^n : \mathcal{V}(x) \neq \emptyset\}$$

Note that $\mathcal{V}(x)$ and $\mathcal{X}$ depend on $\hat{\mathbb{W}}$, but not $\hat{\mu}$. We define the stage cost $\ell : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, terminal cost $V_f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, and the function

$$J(x, \mathbf{v}, \hat{\mathbf{w}}) := \sum_{k=0}^{N-1} \ell(x(k), \pi(x(k), v(k))) + V_f(x(N)) \quad (5)$$

in which $x(k) = \hat{\phi}(k; x, \mathbf{v}, \hat{\mathbf{w}})$. We define the SMPC cost function based on the expected value of $J(\cdot)$ given the disturbance model available, i.e.,

$$\begin{aligned} V_{\hat{\mu}}(x, \mathbf{v}) &:= \hat{\mathbb{E}}[J(x, \mathbf{v}, \hat{\mathbf{w}})] \\ &= \int_{\hat{\mathbb{W}}^N} J(x, \mathbf{v}, \hat{\mathbf{w}}) d\hat{\mu}^N(\hat{\mathbf{w}}) \end{aligned}$$

The optimization problem for any $x \in \mathcal{X}$ is defined as

$$\mathbb{P}_{\hat{\mu}}^0(x) : V_{\hat{\mu}}^0(x) = \min_{\mathbf{v} \in \mathcal{V}(x)} V_{\hat{\mu}}(x, \mathbf{v}) \quad (6)$$

and the optimal solutions for a given distribution $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ are defined by the set-valued mapping $\mathbf{v}_{\hat{\mu}}^0 : \mathcal{X} \rightrightarrows \mathbb{V}^N$ such that

$$\mathbf{v}_{\hat{\mu}}^0(x) := \arg \min_{\mathbf{v} \in \mathcal{V}(x)} V_{\hat{\mu}}(x, \mathbf{v})$$

Note that $\mathbf{v}_{\hat{\mu}}^0(x)$ is a set-valued mapping because there may be multiple solutions to $\mathbb{P}_{\hat{\mu}}^0(x)$.

To streamline the following presentation, we assume that there exists some Borel measurable selection rule that defines a single-valued control law $\kappa_{\hat{\mu}} : \mathcal{X} \rightarrow \mathbb{U}$ such that $\kappa_{\hat{\mu}}(x) \in \{\pi(x, v(0)) : \mathbf{v} \in \mathbf{v}_{\hat{\mu}}^0(x)\}$ for all $x \in \mathcal{X}$, in which $v(0)$ is the first parameter vector in the sequence $\mathbf{v}$.

The resulting closed-loop system is then

$$x^+ = f(x, \kappa_{\hat{\mu}}(x), w) \quad w \in \mathbb{W} \quad (7)$$

in which w is distributed according to μ . We use $\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ to denote the solution to (7) at time $k \in \mathbb{I}_{\geq 0}$ given the initial condition x and disturbance sequence $\mathbf{w}_k = (w(0), \dots, w(k-1)) \in \mathbb{W}^k$. In the subsequent analysis, the control law $\kappa_{\hat{\mu}}(\cdot)$ and therefore the deterministic value of the closed-loop trajectory $\phi_{\hat{\mu}}(\cdot)$ depend on the disturbance model $\hat{\mu}$. The disturbance, however, takes values $w \in \mathbb{W}$ and the expected value of the closed-loop state trajectory is evaluated based on the disturbance distribution $\mu \in \mathcal{M}(\mathbb{W})$. Thus, we discuss quantities such as

$$\begin{aligned} \mathbb{E}[\|\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)\|] &= \int_{\Omega} |\phi_{\hat{\mu}}(k; x, \mathbf{w}_k(\omega))| dP(\omega) \\ &= \int_{\mathbb{W}^k} |\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)| d\mu(w(0)) d\mu(w(1)) \dots d\mu(w(k-1)) \end{aligned}$$

C. Assumptions for SMPC

We require the following assumptions for SMPC.

Assumption 3 (Continuity of system and cost). The model $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}^n$, control parameterization $\pi : \mathbb{R}^n \times \mathbb{V} \rightarrow \mathbb{R}^m$, stage cost $\ell : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, and terminal cost $V_f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ are locally Lipschitz continuous. Furthermore, we have, without loss of generality, that $f(0, 0, 0) = 0$, $\ell(0, 0) = 0$, and $V_f(0) = 0$.

Note that we have strengthened the usual assumption of continuity to local Lipschitz continuity. However, local Lipschitz continuity is typically required already (and satisfied for many problems of interest) if we intend to use standard nonlinear optimization methods to solve the SMPC optimization problem. This assumption of local Lipschitz continuity, however, does not imply that the optimal cost function $V_{\hat{\mu}}^0(\cdot)$ or control law $\kappa_{\hat{\mu}}(\cdot)$ are continuous functions of $x \in \mathcal{X}$.

Assumption 4 (Properties of the constraint sets). The sets $\mathbb{U}$ and $\mathbb{V}$ are compact and contain the origin. The set $\mathbb{X}_f$ is defined by $\mathbb{X}_f := \{x \in \mathbb{R}^n : V_f(x) \leq \tau\}$ for some $\tau > 0$. The set $\mathcal{X}$ is bounded. The control law parameterization satisfies $\pi(x, v) \in \mathbb{U}$ for all $x \in \mathbb{R}^n$ and $v \in \mathbb{V}$.

Note that the final requirement of Assumption 4 means that $\pi(x, v) = Kx + v$ may not be a valid control law parameterization. Instead, we can define $\pi(x, v) = \text{sat}_{\mathbb{U}}(Kx + v)$ in which $u = \text{sat}_{\mathbb{U}}(s)$ maps s to the closest value that satisfies $u \in \mathbb{U}$, i.e., $\text{sat}_{\mathbb{U}}(s) = \arg \min_{u \in \mathbb{U}} |u - s|$.

Assumption 5 (Terminal control law). There exists a locally Lipschitz continuous terminal control law $\kappa_f : \mathbb{X}_f \rightarrow \mathbb{U}$ and $\tilde{\tau} < \tau$ such that for all $x \in \mathbb{X}_f$,

$$f(x, \kappa_f(x), \hat{w}) \in \{x : V_f(x) \leq \tilde{\tau}\} \subset \mathbb{X}_f, \quad \forall \hat{w} \in \hat{\mathbb{W}} \quad (8)$$

$$V_f(f(x, \kappa_f(x), 0)) \leq V_f(x) - \ell(x, \kappa_f(x)) \quad (9)$$

Furthermore, $\pi(x, 0) = \kappa_f(x)$ for all $x \in \mathbb{X}_f$.

Thus, the terminal control law must drive any $x \in \mathbb{X}_f$ to the interior of $\mathbb{X}_f$ for all $\hat{w} \in \hat{\mathbb{W}}$. This assumption is stronger than the assumption of robust positive invariance for the terminal set typically used in SMPC. Assumption 5 ensures that $\mathcal{X}$ is nonempty because $\mathbb{X}_f$ is nonempty and $\mathbb{X}_f \subseteq \mathcal{X}$.

We also require the usual lower bound on the stage cost.

Assumption 6 (Stage cost bound). There exists a function $\alpha_{\ell}(\cdot) \in \mathcal{K}_{\infty}$ such that for all $(x, u) \in \mathbb{R}^n \times \mathbb{U}$, we have that

$$\alpha_{\ell}(|x|) \leq \ell(x, u)$$

Note that all of these assumptions address the construction of the SMPC optimization problem and do not specify any requirements concerning the true disturbance set and distribution ($\mathbb{W}$ and μ). Instead, these assumptions consider only the set $\hat{\mathbb{W}}$. We can also choose any $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ for the set $\hat{\mathbb{W}}$. Given Assumptions 2-5, we can establish that all stochastic quantities considered in the subsequent results are well defined [19, Prop. 5].

III. THE WASSERSTEIN METRIC

The goal of this work is to show that if the difference between μ and $\hat{\mu}$ is small, the degradation in performance of the closed-loop system relative to the idealized SMPC result is also small. Moreover, we wish to show that as $\hat{\mu} \rightarrow \mu$ (e.g., a sampling based approximation), we recover the idealized SMPC guarantee found in [19]. Thus, we require a concept of distance between probability measures, i.e., a probability metric, and a notion of convergence for probability measures.

We use the Wasserstein metric for this task for three main reasons. First, the Wasserstein distance between continuous and discrete distributions is well defined (in contrast to Kullback-Leibler divergence). Second, the Kantorovich-Rubinstein dual representation of the Wasserstein distance provides a useful upper bound for the difference between expected values of a nonlinear, Lipschitz continuous function for two different probability distributions. Third, the Wasserstein distance metrizes weak convergence on $\mathcal{M}(\mathbb{W})$ for compact $\mathbb{W} \subseteq \mathbb{R}^q$ (in contrast to total variation distance which does not converge to zero for sampling based approximations of continuous distributions).

We consider the type-1 version of this metric, sometimes known as the Kantorovich-Rubinstein metric, defined as follows.

Definition 3 (Wasserstein metric). The (type-1) Wasserstein metric $W : \mathcal{M}(\mathbb{W}) \times \mathcal{M}(\mathbb{W}) \rightarrow \mathbb{R}_{\geq 0}$ is defined as

$$W(\mu_1, \mu_2) := \inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{W} \times \mathbb{W}} |w_1 - w_2| d\gamma(w_1, w_2)$$

for all $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{W})$, in which $\Gamma(\mu_1, \mu_2)$ denotes the collection of all measures on $\mathbb{W} \times \mathbb{W}$ with marginals μ_1 and μ_2 , i.e., $\gamma \in \Gamma(\mu_1, \mu_2)$ must satisfy

$$\mu_1(\cdot) = \int_{\mathbb{W}} \gamma(\cdot, w_2) dw_2 \quad \mu_2(\cdot) = \int_{\mathbb{W}} \gamma(w_1, \cdot) dw_1$$

The measure $\gamma(\cdot)$ can be viewed as a *transport plan* for moving a distribution described by μ_1 to another one described by μ_2 . Thus, determining the Wasserstein distance amounts to solving for the optimal transport plan in which the cost is given by the Euclidean distance $|\cdot|$. Note that the Wasserstein metric satisfies all the axioms of a distance on $\mathcal{M}(\mathbb{W})$ for compact $\mathbb{W} \subset \mathbb{R}^n$. Specifically, the metric is finite, symmetric, satisfies the triangle inequality, and $W(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$.

A particularly useful result for the Wasserstein metric (type-1) is the following dual representation.

Theorem 4 (Kantorovich-Rubinstein). *For any probability measures $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{W})$ we have*

$$W(\mu_1, \mu_2) = \sup_{g \in \mathcal{L}} \left\{ \int_{\mathbb{W}} g(w) d\mu_1(w) - \int_{\mathbb{W}} g(w) d\mu_2(w) \right\}$$

in which $\mathcal{L}$ denotes the space of all Lipschitz continuous functions with $|g(w_1) - g(w_2)| \leq |w_1 - w_2|$ for all $w_1, w_2 \in \mathbb{W}$.

See Villani [28, Remark 6.5] for a discussion of this result. Thus, for any Lipschitz continuous function $g(\cdot)$ with Lipschitz constant $L \geq 0$ on $\mathbb{W}$, we have that

$$\int_{\mathbb{W}} g(w) d\mu_1(w) - \int_{\mathbb{W}} g(w) d\mu_2(w) \leq LW(\mu_1, \mu_2)$$

for all $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{W})$.

We now introduce a notion of convergence for probability measures in the space $\mathcal{M}(\mathbb{W})$. For a sequence of probability measures $\{\mu_i\}_{i=1}^{\infty}$ in $\mathcal{M}(\mathbb{W})$, we say that μ_i converges weakly to $\mu \in \mathcal{M}(\mathbb{W})$ if

$$\int_{\mathbb{W}} g(w) d\mu_i(w) \rightarrow \int_{\mathbb{W}} g(w) d\mu(w)$$

for all continuous and bounded functions $g : \mathbb{W} \rightarrow \mathbb{R}$. We use the notation $\mu_i \rightarrow \mu$ to denote weak convergence. For compact $\mathbb{W}$, convergence in the Wasserstein metric is equivalent to weak convergence, i.e., for the sequence $\{\mu_i\}_{i=1}^{\infty}$, $\mu_i \rightarrow \mu$ if and only if $W(\mu_i, \mu) \rightarrow 0$ [28, Theorem 6.9].

The notion of convergence for probability measures is particularly important for sampling-based empirical approximations of a probability measure. For example, consider a probability measures $\mu \in \mathcal{M}(\mathbb{W})$. We draw s random samples from μ that we denote $\{\hat{\omega}_i\}_{i=1}^s$ and define the empirical probability measure as $\hat{\mu}_s := \frac{1}{s} \sum_{i=1}^s \delta_{\hat{\omega}_i}$ in which $\delta_{\hat{\omega}_i}$ is the Dirac measure at $\hat{\omega}_i$. One can show that $\hat{\mu}_s \rightarrow \mu$ as $s \rightarrow \infty$ via the strong law of large numbers [27].

IV. DISTRIBUTIONAL ROBUSTNESS

We first define robust positive invariance (RPI) as follows.

Definition 5 (Robust positive invariance). The set $\mathcal{X}$ is robustly positive invariant (RPI) for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ if $x \in \mathcal{X}$ implies that $x^+ \in \mathcal{X}$ for all $w \in \mathbb{W}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$.

We define distributional robustness for a closed-loop nonlinear systems as follows.

Definition 6 (Distributionally Robust Asymptotic Stability in Expectation). The origin of the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ is distributionally robustly asymptotically stable in expectation (DRASiE) in the RPI set $\mathcal{X}$ if there exist $\beta(\cdot) \in \mathcal{KL}$ and $\gamma_1(\cdot), \gamma_2(\cdot) \in \mathcal{K}$ such that

$$\begin{aligned} \mathbb{E}[\|\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)\|] \\ \leq \beta(|x|, k) + \gamma_1(\hat{\mathbb{E}}[\|\hat{w}\|]) + \gamma_2(W(\mu, \hat{\mu})) \end{aligned} \quad (10)$$

for all $x \in \mathcal{X}$, $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, $\mu \in \mathcal{M}(\mathbb{W})$, and $k \in \mathbb{I}_{\geq 0}$.

The first part of the upper bound in (10) is a $\mathcal{KL}$ function that ensures the effect of the initial condition $x \in \mathcal{X}$ (asymptotically) vanishes as $k \rightarrow \infty$. The second function $\gamma_1(\hat{\mathbb{E}}[\|\hat{w}\|])$ accounts for the persistent effect of the modeled disturbance ($\hat{w}$) in the control law design and the ideal system with $\mu = \hat{\mu}$. The third function $\gamma_2(W(\mu, \hat{\mu}))$ accounts for the discrepancy between the disturbance distribution model $\hat{\mu}$, used in the SMPC optimization problem, and the true disturbance distribution μ . If $\mu = \hat{\mu}$, then $\gamma_2(W(\mu, \hat{\mu})) = 0$ and we recover the usual bound for idealized SMPC analysis. The most significant consequence of this result is that the effect of arbitrarily small errors between $\hat{\mu}$ and μ produce similarly small deviations from the closed-loop bound derived for idealized SMPC analysis. We further discuss the implications of this property for SMPC in Section VI.

Next, we define an SISS Lyapunov function to serve as a sufficient condition for this definition of distributional robustness.

Definition 7 (SISS Lyapunov function). The measurable function $V_{\hat{\mu}} : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ is an SISS Lyapunov function, defined for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ in the RPI set $\mathcal{X}$ if there exist $\alpha_1(\cdot), \alpha_2(\cdot), \alpha_3(\cdot) \in \mathcal{K}_{\infty}$ and $\sigma_2(\cdot), \sigma_3(\cdot), \sigma_4(\cdot) \in \mathcal{K}$ such that

$$\alpha_1(|x|) \leq V_{\hat{\mu}}(x) \leq \alpha_2(|x|) + \sigma_2(\hat{\mathbb{E}}[\|\hat{w}\|]) \quad (11)$$

$$\begin{aligned} \int_{\mathbb{W}} V_{\hat{\mu}}(f(x, \kappa_{\hat{\mu}}(x), w)) d\mu(w) \\ \leq V_{\hat{\mu}}(x) - \alpha_3(|x|) + \sigma_3(\hat{\mathbb{E}}[\|\hat{w}\|]) + \sigma_4(W(\mu, \hat{\mu})) \end{aligned} \quad (12)$$

for all $x \in \mathcal{X}$, $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, $\mu \in \mathcal{M}(\mathbb{W})$.

Note that we allow the upper bound for $V_{\hat{\mu}}(x)$ to depend on $\hat{\mathbb{E}}[\|\hat{w}\|]$ and therefore $\hat{\mu}$. Since we intend to use the optimal cost of SMPC as the SISS Lyapunov function, this generalization is necessary since $V_{\hat{\mu}}^0(x)$ is usually nonzero for $x = 0$ and grows with increasing $\hat{\mathbb{E}}[\|\hat{w}\|]$.

We can now use this SISS Lyapunov function as a sufficient condition for DRASiE.

Proposition 8. *If a system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ admits an SISS Lyapunov function in the RPI bounded set $\mathcal{X}$, then the origin is DRASiE.*

As the proof of this result is similar to the proof in [19, Prop. 13], we defer this proof to the Appendix.

V. INHERENT DISTRIBUTIONAL ROBUSTNESS OF SMPC

In addition to characterizing the distance between μ and $\hat{\mu}$, we must also characterize the distance between the sets $\mathbb{W}$ and $\hat{\mathbb{W}}$. A natural metric for the distance between sets is the Hausdorff distance defined for two sets $X, Y \subseteq \mathbb{R}^n$ as

$$d_H(X, Y) := \max \left\{ \sup_{x \in X} \inf_{y \in Y} \|x - y\|, \sup_{y \in Y} \inf_{x \in X} \|y - x\| \right\}$$

Note that since $\hat{\mathbb{W}} \subseteq \mathbb{W}$ and both sets are compact, we have that $d_H(\mathbb{W}, \hat{\mathbb{W}}) = \max_{w \in \mathbb{W}} \|w\|_{\hat{\mathbb{W}}}$. We show an example of

this distance in Figure 1. Note that $W(\mu, \hat{\mu}) \rightarrow 0$ does *not* imply that $d_H(\mathbb{W}, \hat{\mathbb{W}}) \rightarrow 0$.¹

The main result of this paper is now stated.

Theorem 9. *Let Assumptions 1-6 hold. Then there exists $\delta > 0$ such that for any set $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$, the feasible set $\mathcal{X}$ is RPI for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ and the origin of this system is DRASiE in $\mathcal{X}$.*

Thus, for a sufficiently small difference between $\mathbb{W}$ and $\hat{\mathbb{W}}$, the SMPC problem remains robustly recursively feasible and is distributionally robust. Note that if $d_H(\mathbb{W}, \hat{\mathbb{W}})$ is too large, the feasible set $\mathcal{X}$ may not be RPI even if $W(\mu, \hat{\mu}) \rightarrow 0$. In other words, disturbances $w \notin \hat{\mathbb{W}}$ that occur with arbitrarily small (but nonzero) probability can still affect the feasibility of the SMPC optimization problem.

To establish this result, we require the following intermediate lemmata. Their corresponding proofs can be found in the Appendix.

Lemma 10. *Consider a Lipschitz continuous function $g : X \times S \rightarrow \mathbb{R}$ with $X \subseteq \mathbb{R}^n$ and $S \subseteq \mathbb{R}^m$. Then $G(x) := \int_S g(x, s) d\mu(s)$ is also a Lipschitz continuous function with the same Lipschitz constant for all $\mu \in \mathcal{M}(S)$.*

Lemma 11. *Let Assumptions 2-5 hold. Then there exists $L_f \geq 0$ such that*

$$\begin{aligned} \int_{\hat{\mathbb{W}}} V_f(f(x, \kappa_f(x), \hat{w})) d\hat{\mu}(\hat{w}) \\ \leq V_f(x) - \ell(x, \kappa_f(x)) + L_f \hat{\mathbb{E}}[|\hat{w}|] \end{aligned} \quad (13)$$

for all $x \in \mathbb{X}_f$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$.

Lemma 12. *Let Assumptions 2-5 hold. Then there exist $\alpha_2(\cdot) \in \mathcal{K}_\infty$ and $\sigma_2(\cdot) \in \mathcal{K}$ such that $V_{\hat{\mu}}^0(x) \leq \alpha_2(|x|) + \sigma_2(\hat{\mathbb{E}}[|\hat{w}|])$ for all $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$.*

We now proceed to the proof of Theorem 9.

Proof of Theorem 9. We proceed by first establishing the there exists $\delta > 0$ such that $\mathcal{X}$ is RPI. Since $f(\cdot)$ and $\pi(\cdot)$ are locally Lipschitz continuous and $\mathcal{X}$ is bounded, there exists $L_x > 0$ such that

$$|f(x, \pi(x, v), w) - f(x, \pi(x, v), \hat{w})| \leq L_x |w - \hat{w}|$$

for all $x \in \mathcal{X}$, $v \in \mathbb{V}$, $w \in \mathbb{W}$, and $\hat{w} \in \hat{\mathbb{W}}$. For $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, choose $\mathbf{v}^0 \in \mathbf{v}_{\hat{\mu}}^0(x)$ such that $\kappa_{\hat{\mu}}(x) = \pi(x, v^0(0))$, any $\hat{\mathbf{w}} \in \hat{\mathbb{W}}^N$, and define

$$\hat{\mathbf{v}}^+ := (v^0(1), v^0(2), \dots, v^0(N-1), 0)$$

and

$$\hat{\mathbf{w}}^+ := (\hat{w}(1), \hat{w}(2), \dots, \hat{w}(N-1), \hat{w}(N))$$

for some $\hat{w}(N) \in \hat{\mathbb{W}}$. We denote $x^+(w) = f(x, \kappa_{\hat{\mu}}(x), w)$, $x(N) = \hat{\phi}(N; x, \mathbf{v}^0, \hat{\mathbf{w}})$, and $x^+(N; w) = \hat{\phi}(N; x^+(w), \hat{\mathbf{v}}^+, \hat{\mathbf{w}}^+)$. Note that $x \in \mathcal{X}$, $w \in \mathbb{W}$, $\hat{\mathbf{v}}^+ \in \mathbb{V}^N$, and $\hat{\mathbf{w}}^+ \in \hat{\mathbb{W}}^N$ are all bounded. The function $\hat{\phi}(N; \cdot)$ is locally Lipschitz continuous since it is a

composition of a finite number of locally Lipschitz continuous functions, i.e., the composition of $f(\cdot)$ with itself N times. Therefore, $V_f(\hat{\phi}(N; \cdot))$ is also locally Lipschitz continuous and there exists $\tilde{L}_f > 0$ such that

$$\begin{aligned} V_f(x^+(N; w)) - V_f(x^+(N; \hat{w})) \\ \leq |V_f(x^+(N; w)) - V_f(x^+(N; \hat{w}))| \\ \leq \tilde{L}_f |x^+(w) - x^+(\hat{w})| \\ \leq \tilde{L}_f L_x |w - \hat{w}| \end{aligned}$$

for all $w \in \mathbb{W}$ and $\hat{w} \in \hat{\mathbb{W}}$. Since $x(N) \in \mathbb{X}_f$, we have from Assumption 5 that $V_f(x^+(N; \hat{w})) \leq \tilde{\tau}$ for all $\hat{w} \in \hat{\mathbb{W}}$ and therefore,

$$V_f(x^+(N; w)) \leq \tilde{\tau} + \tilde{L}_f L_x |w - \hat{w}|$$

for any $w \in \mathbb{W}$ and $\hat{w} \in \hat{\mathbb{W}}$. Thus, for any $w \in \mathbb{W}$, we can choose $\hat{w} \in \hat{\mathbb{W}}$ to minimize the value of $|w - \hat{w}|$, i.e., the value of $\hat{w}$ closest to w , and we have that

$$V_f(x^+(N; w)) \leq \tilde{\tau} + \tilde{L}_f L_x |w|_{\hat{\mathbb{W}}}$$

We define $\delta := (\tau - \tilde{\tau})/(\tilde{L}_f L_x) > 0$. Thus, for all sets $\mathbb{W}$ such that $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$, we have that $|w|_{\hat{\mathbb{W}}} \leq \delta$ and therefore $V_f(x^+(N; w)) \leq \tau$ and $x^+(N; w) \in \mathbb{X}_f$ for all $w \in \mathbb{W}$. Since $\pi(x, v) \in \mathbb{U}$ for any $v \in \mathbb{V}$ and $x \in \mathbb{R}^n$, we have that $\mathbf{v}^+ \in \mathcal{V}(x^+(w))$ and therefore $x^+(w) \in \mathcal{X}$ for all $w \in \mathbb{W}$. Since the choice of $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ was arbitrary, we have that $\mathcal{X}$ is RPI for the system $f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ and any set $\mathbb{W} \subseteq \mathbb{R}^q$ such that $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$.

We now establish an expected cost decrease condition for the probability measure $\hat{\mu}$ similar to [19, Prop. 11]. Using the definition of $J(\cdot)$, we obtain

$$\begin{aligned} J(x^+(\hat{w}), \hat{\mathbf{v}}^+, \hat{\mathbf{w}}^+) \\ = J(x, \mathbf{v}^0, \hat{\mathbf{w}}) - \ell(x, \kappa_{\hat{\mu}}(x)) + \eta(x(N), \hat{w}(N)) \end{aligned} \quad (14)$$

in which

$$\eta(x, w) := -V_f(x) + \ell(x, \kappa_f(x)) + V_f(f(x, \kappa_f(x), w))$$

From Lemma 11 and the fact that $x(N) \in \mathbb{X}_f$, there exists $L_f > 0$ such that

$$\int_{\hat{\mathbb{W}}^{N+1}} \eta(x(N), \hat{w}(N)) d\hat{\mu}^N(\hat{\mathbf{w}}) d\hat{\mu}(\hat{w}(N)) \leq L_f \hat{\mathbb{E}}[|\hat{w}|] \quad (15)$$

We also have the equality

$$V_{\hat{\mu}}^0(x) = \int_{\hat{\mathbb{W}}^{N+1}} J(x, \mathbf{v}^0, \hat{\mathbf{w}}) d\hat{\mu}^N(\hat{\mathbf{w}}) d\hat{\mu}(\hat{w}(N)) \quad (16)$$

We integrate both sides of (14) with respect to $\hat{\mu}^{N+1}$ and apply (15), (16), and the definition of $V_{\hat{\mu}}(\cdot)$ to give

$$\begin{aligned} \int_{\hat{\mathbb{W}}} V_{\hat{\mu}}(x^+(\hat{w}), \hat{\mathbf{v}}^+) d\hat{\mu}(\hat{w}) \\ \leq V_{\hat{\mu}}^0(x) - \ell(x, \kappa_{\hat{\mu}}(x)) + L_f \hat{\mathbb{E}}[|\hat{w}|] \end{aligned} \quad (17)$$

in which we can exchange $\hat{\mathbb{W}}$ with $\mathbb{W}$ for the domain of integration since Assumption 2 ensures that $\hat{\mu}(\mathbb{W} \setminus \hat{\mathbb{W}}) = 0$.

Now, we use Theorem 4 to exchange $\hat{\mu}$ with μ . The function $J(x, \mathbf{v}, \mathbf{w})$ is a composition of a finite number of

¹Consider $\mu_\varepsilon = (1 - \varepsilon)\delta_{\{0\}} + \varepsilon\delta_{\{1\}}$ and $\hat{\mu} = \delta_{\{0\}}$ with $\mathbb{W} = \{0, 1\}$ and $\hat{\mathbb{W}} = \{0\}$. As $\varepsilon \rightarrow 0$, $W(\mu_\varepsilon, \hat{\mu}) \rightarrow 0$, but $d_H(\mathbb{W}, \hat{\mathbb{W}}) = 1$.

locally Lipschitz continuous functions and is therefore locally Lipschitz continuous. Thus, $J(x, \mathbf{v}, \mathbf{w})$ is Lipschitz continuous on the compact set $\mathcal{X} \times \mathbb{V}^N \times \mathbb{W}^N$. From Lemma 10, we have that $V_{\hat{\mu}}(x, \mathbf{v})$ is also Lipschitz continuous with the same Lipschitz constant for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. Thus, there exists $L_J > 0$ such that

$$|V_{\hat{\mu}}(f(x, u, w), \mathbf{v}^+) - V_{\hat{\mu}}(f(x, u, \hat{w}), \mathbf{v}^+)| \leq L_J |w - \hat{w}|$$

for all $w, \hat{w} \in \mathbb{W}$, $x \in \mathcal{X}$, $u \in \mathbb{U}$, $\mathbf{v}^+ \in \mathbb{V}^N$, and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. We choose arbitrary $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ and use Theorem 4 to give,

$$\begin{aligned} & \int_{\mathbb{W}} V_{\hat{\mu}}(f(x, u, w), \mathbf{v}^+) d\mu(w) \\ & \leq \int_{\mathbb{W}} V_{\hat{\mu}}(f(x, u, \hat{w}), \mathbf{v}^+) d\hat{\mu}(\hat{w}) + L_J W(\mu, \hat{\mu}) \end{aligned} \quad (18)$$

for all $x \in \mathcal{X}$, $u \in \mathbb{U}$, $\mathbf{v}^+ \in \mathbb{V}^N$, $\mu \in \mathcal{M}(\mathbb{W})$. Note that the choice of $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ was arbitrary and therefore (18) holds for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ with the same value of $L_J > 0$. We combine (17), (18) and by optimality we have

$$\begin{aligned} & \int_{\mathbb{W}} V_{\hat{\mu}}^0(f(x, \kappa_{\hat{\mu}}(x), w)) d\mu(w) \\ & \leq V_{\hat{\mu}}^0(x) - \ell(x, \kappa_{\hat{\mu}}(x)) + L_f \hat{\mathbb{E}}[|\hat{w}|] + L_J W(\mu, \hat{\mu}) \end{aligned} \quad (19)$$

for all $x \in \mathcal{X}$, $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, and $\mu \in \mathcal{M}(\mathbb{W})$.

We now establish that $V_{\hat{\mu}}^0(\cdot)$ is an SISS Lyapunov function. From Assumption 6, there exists $\alpha_{\ell}(\cdot) \in \mathcal{K}_{\infty}$ such that $-\ell(x, \kappa_{\hat{\mu}}(x)) \leq -\alpha_{\ell}(|x|)$ in (19). Therefore, (12) holds with $V_{\hat{\mu}}(\cdot) := V_{\hat{\mu}}^0(\cdot)$, $\alpha_3(\cdot) := \alpha_{\ell}(\cdot) \in \mathcal{K}_{\infty}$, $\sigma_3(s) := L_f s \in \mathcal{K}$, and $\sigma_4(s) := L_J s \in \mathcal{K}$. We also use Assumption 6 to show that $\alpha_1(|x|) := \alpha_{\ell}(|x|) \leq \ell(x, \kappa_{\hat{\mu}}(x)) \leq V_{\hat{\mu}}(x)$ for all $x \in \mathcal{X}$. We then use Lemma 12 to construct the upper bound for $V_{\hat{\mu}}^0(\cdot)$. Thus, $V_{\hat{\mu}}^0(\cdot)$ satisfies all the requirements in Definition 7 for an SISS Lyapunov function. By Proposition 8, the origin is DRASiE. $\square$

An important class of applications for SMPC are economic problems in which the stage cost is defined to directly represent a performance metric for the process (e.g., economic cost, carbon production). If this cost is positive definite with respect to the origin (the specified steady state), then the results of Theorem 9 also hold. But this requirement restricts the space of economic cost functions that we may consider with SMPC and can exclude many relevant problems. Thus, in economic applications of MPC, i.e., economic MPC, the key generalization is that we do not require Assumption 6 to hold. By dropping this assumption, we obtain a weaker, but still instructive result, for economic applications of SMPC.²

Theorem 13. *Let Assumptions 1-5 hold. Then there exists $\delta > 0$ such that for any set $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$, the feasible set $\mathcal{X}$ is RPI for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$,*

²If we assume some form of “stochastic dissipativity” for the stage cost and system, we may be able to retain the results in Theorem 9, but the form of this condition is another open research question for even idealized SMPC.

$w \in \mathbb{W}$. Furthermore, there exist $L_f, L_J > 0$ such that the closed-loop trajectory satisfies

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\ell(x(k), \kappa_{\hat{\mu}}(x(k)))] \\ & \leq L_f \hat{\mathbb{E}}[|\hat{w}|] + L_J W(\mu, \hat{\mu}) \end{aligned} \quad (20)$$

in which $x(k) = \phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ for all $x \in \mathcal{X}$, $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, $\mu \in \mathcal{M}(\mathbb{W})$.

Proof. In the proof of Theorem 9, we established that $\mathcal{X}$ is RPI and that the bound in (19) holds for some $\delta > 0$ without using Assumption 6. We choose $x \in \mathcal{X}$ and denote the closed-loop trajectory $x(k) = \phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$. We then apply the law of total expectation to (19) and rearrange to give,

$$\begin{aligned} & \mathbb{E}[\ell(x(k), \kappa_{\hat{\mu}}(x(k)))] \\ & \leq \mathbb{E}[V_{\hat{\mu}}^0(x(k))] - \mathbb{E}[V_{\hat{\mu}}^0(x(k+1))] + L_f \hat{\mathbb{E}}[|\hat{w}|] + L_J W(\mu, \hat{\mu}) \end{aligned}$$

We sum both sides of this inequality from $k = 0$ to $T - 1$, cancel terms, and divide by T to give,

$$\begin{aligned} & \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\ell(x(k), \kappa_{\hat{\mu}}(x(k)))] \\ & \leq \frac{V_{\hat{\mu}}^0(x) - \mathbb{E}[V_{\hat{\mu}}^0(x(T))]}{T} + L_f \hat{\mathbb{E}}[|\hat{w}|] + L_J W(\mu, \hat{\mu}) \end{aligned}$$

Since $V_{\hat{\mu}}(x, \mathbf{v})$ is Lipschitz continuous on $\mathcal{X} \times \mathbb{V}^N$ for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, we have that $V_{\hat{\mu}}^0(x)$ is bounded uniformly for all $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. Therefore, $\mathbb{E}[V_{\hat{\mu}}^0(x(T))]$ is bounded since $x(T) \in \mathcal{X}$. We take the limit supremum as $T \rightarrow \infty$ and note that $V_{\hat{\mu}}^0(x)/T$ and $\mathbb{E}[V_{\hat{\mu}}^0(x(T))]/T$ vanish as $T \rightarrow \infty$ to give (20). $\square$

VI. DISCUSSION

We now discuss several insights derived from Theorems 9 and 13.

A. SMPC

Idealized SMPC: For idealized SMPC, we have that the disturbance model and distribution are exact and therefore $\mu = \hat{\mu}$ and $\mathbb{W} = \hat{\mathbb{W}}$. Under these conditions, $W(\mu, \hat{\mu}) = 0$, $d_H(\mathbb{W}, \hat{\mathbb{W}}) = 0$, and (10), (20) reduce to their idealized SMPC counterparts discussed in [19], i.e.,

$$\mathbb{E}[\|\phi_{\mu}(k; x, \mathbf{w}_k)\|] \leq \beta(|x|, k) + \gamma_1(\hat{\mathbb{E}}[|\hat{w}|]) \quad \forall k \in \mathbb{I}_{\geq 0} \quad (21)$$

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\ell(x(k), \kappa_{\mu}(x(k)))] \leq L_f \hat{\mathbb{E}}[|\hat{w}|] \quad (22)$$

for all $x \in \mathcal{X}$ and $\mu \in \mathcal{M}(\mathbb{W})$. The bound in (22) is a standard result for idealized SMPC performance that was first derived for nonlinear systems in Chatterjee and Lygeros [7]. Similar results are also available for (idealized) robust MPC [3].

Incorrectly modeled disturbances: We assume that $\mathbb{W} = \hat{\mathbb{W}}$, but the distribution is incorrect: $\mu \neq \hat{\mu}$. In this case, $d_H(\mathbb{W}, \hat{\mathbb{W}}) = 0$ and recursive feasibility is guaranteed ($\mathcal{X}$ is RPI). The performance bound for either $\mathbb{E}[\|x(k)\|]$ or $\mathbb{E}[\ell(x(k), u(k))]$ degrades relative the idealized case with respect to the distance between $\hat{\mu}$ and μ , i.e., $W(\mu, \hat{\mu})$. Thus, as $\hat{\mu} \rightarrow \mu$ we recover the idealized SMPC bound. Furthermore, arbitrarily small differences between $\hat{\mu}$ and μ , in terms of $W(\cdot)$, produce similarly small deviations in the performance bound. If we select $\hat{\mathbb{W}}$ larger than the support of μ , i.e., $\mu(\hat{\mathbb{W}}) = 1$, then we have that $\mathbb{W} = \hat{\mathbb{W}}$. Thus, designing for a larger disturbance than the system experiences does not affect feasibility. This additional conservatism, however, may reduce performance. Moreover, for a sufficiently large set $\hat{\mathbb{W}}$, there may not exist any $\kappa_f(\cdot)$, $\mathbb{X}_f$, and $V_f(\cdot)$ that satisfy Assumption 5.

Unmodeled disturbances: In this case, we have that $\mathbb{W} \neq \hat{\mathbb{W}}$ and $\mu \neq \hat{\mu}$. Note that this case represents disturbances that were “undermodeled” in which the true disturbance may be larger than the disturbances included in $\hat{\mathbb{W}}$, e.g., we define $\hat{\mathbb{W}} := \{w \in \mathbb{R}^q : |w| \leq 1\}$, but $\mathbb{W} := \{w \in \mathbb{R}^q : |w| \leq 2\}$ with $\mu(\hat{\mathbb{W}}) < 1$. This case also covers elements or directions of $w \in \mathbb{W}$ that are entirely absent in $\hat{\mathbb{W}}$, e.g., $\hat{\mathbb{W}} := \{w \in \mathbb{R}^q : |w| \leq 1 \text{ and } w_1 = 0\}$ or $\hat{\mathbb{W}} := \{0\}$. This representation can also be interpreted as an error in the dynamic model $f(\cdot)$. For example, we consider the dynamic model $\hat{f}(x, u, w_1)$, but the actual system evolves according to $f(x, u, w) = \hat{f}(x, u + w_2, 0) + w_3$ in which $w = [w_1, w_2, w_3]'$. Nonetheless, we may still define $\hat{\mathbb{W}} := \{w \in \mathbb{R}^3 : |w_1| \leq 1, w_2 = 0, w_3 = 0\}$

B. Scenario Optimization

Scenario optimization methods are often used to approximate and solve the stochastic optimization problem in (6), particularly for nonlinear systems. By selecting a finite set of possible scenarios from the underlying disturbance distribution and set, the stochastic objective can be approximated by the average cost of these scenarios and the constraints in the optimization problem are required to hold for all scenarios considered. The quality of and performance bounds for this approximate optimal solution/cost are topics that have generated much interest, with applications beyond SMPC. However, the quality of this approximation is irrelevant for SMPC if near exact approximations still produce poor controllers. The contribution in this subsection is novel because we are able to bound the performance of the closed-loop system generated by repeated solutions to this approximated optimization problem and can thereby directly address performance of the controller subject to a scenario-based approximation of the stochastic optimization problem.

We proceed by redefining the optimization problem in (6) based on an empirical distribution generated via a scenario-based approximation of the original stochastic optimal control problem. Thus, we can analyze the scenario approximation error as an additional error in representing the true disturbance distribution for w . Specifically, we construct this scenario optimization problem by drawing $s \in \mathbb{I}_{\geq 1}$ samples, denoted $\hat{\omega}_i$, from the model disturbance distribution $\hat{\mu}$ and set $\hat{\mathbb{W}}$.

We then define the set $\hat{\mathbb{W}}_s = \{\hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_s\}$ and empirical distribution

$$\hat{\mu}_s(\cdot) := \frac{1}{s} \sum_{i=1}^s \delta_{\hat{\omega}_i}(\cdot)$$

Note that $\hat{\mu}_s(\cdot)$ and $\hat{\mathbb{W}}_s$ satisfy the requirements in Assumption 2 for all $s \in \mathbb{I}_{\geq 1}$. Moreover, if Assumption 5 holds for $\hat{\mathbb{W}}$, Assumption 5 also holds for $\hat{\mathbb{W}}_s \subseteq \hat{\mathbb{W}}$. Thus, we may use $\hat{\mu}_s(\cdot)$, $\hat{\mathbb{W}}_s$ in place of $\hat{\mu}$, $\hat{\mathbb{W}}$ for all algorithms and results in this work including Theorems 9 and 13.

This fact allows us to draw several important conclusions for scenario-based approximations of the stochastic optimal control problem. First, if $\hat{\mathbb{W}}$ is sufficiently close to $\mathbb{W}$ ($d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta/2$), and the sampling of $\hat{\mathbb{W}}$ is sufficiently dense ($d_H(\hat{\mathbb{W}}, \hat{\mathbb{W}}_s) \leq \delta/2$), then $\mathcal{X}$ is RPI ($d_H(\mathbb{W}, \hat{\mathbb{W}}_s) \leq \delta$). This observation implies that scenario-based methods, if a sufficient number of samples are used, can ensure robust recursive feasibility for SMPC. In fact, deliberate construction of $\hat{\mathbb{W}}_s$ to ensure the approximation is sufficiently dense on $\hat{\mathbb{W}}$ may be preferable to constructing $\hat{\mathbb{W}}_s$ via random sampling, in which $d_H(\hat{\mathbb{W}}_s, \hat{\mathbb{W}})$ may take large values with some small probability.

Second, the performance is bounded by the distance between $\hat{\mu}_s$ and μ . By the triangle inequality, we have that

$$W(\mu, \hat{\mu}_s) \leq W(\mu, \hat{\mu}) + W(\hat{\mu}, \hat{\mu}_s)$$

By the weak convergence of sampling, we have

$$W(\mu, \hat{\mu}_s) \rightarrow W(\mu, \hat{\mu})$$

with probability one as $s \rightarrow \infty$. Thus, the performance bounds in (10) and (20) converge to their values for the original stochastic optimization problem as the number of samples increases. We can further quantify this convergence of the Wasserstein metric in terms of the number samples and dimension of the w via results that can be found in [9]. As the number of samples increases, we expect the closed-loop performance of this scenario-based approximation to improve at the expense of increased computational cost. The trade-off between the performance and computational cost is therefore important to understand and manage.

C. Constraint-tightened MPC

Although the conclusions of Theorema 9 and 13 are informative for SMPC, these results are particularly interesting given their ability to unify notions of stochastic robustness across different MPC formulations. Since different MPC formulations may be treated as special cases of SMPC, we can use this single theorem to draw conclusions about two additional MPC formulations: constraint-tightened MPC (discussed in this subsection) and nominal MPC (discussed in the next subsection).

For constraint-tightened MPC (CMPC), we optimize a nominal objective function subject to the tightened constraints used in the SMPC problem formulation, i.e.,

$$\min_{\mathbf{v} \in \mathcal{V}(x)} J(x, \mathbf{v}, \mathbf{0}) \quad (23)$$

For $\hat{\mu}(\{0\}) = 1$, the SMPC optimization problem in (6) reduces to (23). Furthermore, we have that $\mathbb{E}[\|\hat{w}\|] = 0$ and the Wasserstein metric reduces to

$$W(\mu, \hat{\mu}) = \int_{\mathbb{W}} |w| d\mu = \mathbb{E}[\|w\|]$$

Therefore, (10), (20) can be simplified as follows.

Corollary 14 (CMPC). *Let Assumptions 1-5 hold with $\hat{\mu}(\{0\}) = 1$. Then there exists $\delta > 0$ such that for any set $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$, the feasible set $\mathcal{X}$ is RPI for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$. Furthermore, there exists $L_J > 0$ such that the closed-loop trajectory satisfies*

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\ell(x(k), \kappa_{\hat{\mu}}(x(k)))] \leq L_J \mathbb{E}[\|w\|] \quad (24)$$

in which $x(k) = \phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ for all $x \in \mathcal{X}$ and $\mu \in \mathcal{M}(\mathbb{W})$. If Assumption 6 also holds, there exist $\beta(\cdot) \in \mathcal{KL}$ and $\gamma_2(\cdot) \in \mathcal{K}$ such that

$$\mathbb{E}[\|\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)\|] \leq \beta(|x|, k) + \gamma_2(\mathbb{E}[\|w\|]) \quad (25)$$

for all $x \in \mathcal{X}$, $\mu \in \mathcal{M}(\mathbb{W})$, and $k \in \mathbb{I}_{\geq 0}$.

By using these tightened constraints with a nominal objective function, we still guarantee that CMPC renders the feasible set $\mathcal{X}$ RPI for the closed-loop system and all $w \in \mathbb{W}$ with $\mathbb{W}$ sufficiently close to $\hat{\mathbb{W}}$. Note that $\gamma_2(\cdot)$ and L_J appear in (25) and (24), but $\gamma_1(\cdot)$ and L_f appear in the similar bounds for idealized SMPC in (21) and (22). This observation suggests, as one may expect, that the performance of CMPC and SMPC may differ quantitatively, but the qualitative behavior is likely similar for an otherwise equivalent problem. In general, however, these bounds are too conservative to provide useful quantitative information.

D. Nominal MPC

If we also choose $\hat{\mathbb{W}} = \{0\}$, the SMPC problem reduces to a nominal MPC problem in which we have embedded the feedback law $\pi(x, v)$ in the optimization problem. Note that this type of parameterization has been previously used in nominal MPC formulations to “pre-stabilize” the open-loop system and thereby ensure that the MPC optimization problem is well conditioned [12, 24]. If we choose $\pi(x, v) = v$ and $\mathbb{V} = \mathbb{U}$, the problems are equivalent since $\mathcal{V}(x)$ reduces to

$$\mathcal{V}(x) = \mathcal{U}(x) = \{\mathbf{u} \in \mathbb{U}^N : \hat{\phi}(N; x, \mathbf{u}, \mathbf{0}) \in \mathbb{X}_f\}$$

and the optimization problem becomes

$$\min_{\mathbf{v} \in \mathcal{V}(x)} J(x, \mathbf{v}, \mathbf{0}) = \min_{\mathbf{u} \in \mathcal{U}(x)} J(x, \mathbf{u}, \mathbf{0})$$

With this choice of $\pi(x, v)$, Assumptions 3, 4, and 6 are equivalent to the assumptions used in [1] to establish the inherent robustness of nominal MPC.³ In Assumption 5, however, the requirement in (8) reduces to

$$f(x, \kappa_f(x), 0) \in \{x \in \mathbb{R}^n : V_f(x) \leq \tilde{\tau}\} \quad (26)$$

³With the exception that Assumption 4 also requires bounded $\mathcal{X}$, which is a minor restriction.

i.e., the terminal control law must drive the subsequent state for the nominal system to the interior of $\mathbb{X}_f$ when $x \in \mathbb{X}_f$. If Assumption 6 also holds, the nominal cost decrease in (9) is sufficient to guarantee that (26) also holds. If Assumption 6 does not hold, however, the requirement in (26) is notably different than the requirements typically considered or used for economic MPC. As we discuss in the subsequent paragraphs, this requirement allows us to derive new results for nominal economic MPC.

For this problem, $d_H(\mathbb{W}, \{0\}) = \max_{w \in \mathbb{W}} |w|$ and the bounds in (10), (20) reduce to the equations in (25), (24). Thus, nominal MPC also confers some margin of inherent stochastic robustness to sufficiently small disturbances, as previously discussed in [18]. We summarize this result in the following corollary.

Corollary 15 (Nominal MPC). *Let Assumptions 1-5 hold with $\hat{\mu}(\{0\}) = 1$ and $\hat{\mathbb{W}} = \{0\}$. Then there exists $\delta > 0$ such that for any set $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $\max_{w \in \mathbb{W}} |w| \leq \delta$, the feasible set $\mathcal{X}$ is RPI for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$. Furthermore, there exists $L_J > 0$ such that the closed-loop trajectory satisfies*

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[\ell(x(k), \kappa_{\hat{\mu}}(x(k)))] \leq L_J \mathbb{E}[\|w\|] \quad (27)$$

in which $x(k) = \phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ for all $x \in \mathcal{X}$ and $\mu \in \mathcal{M}(\mathbb{W})$. If Assumption 6 also holds, there exist $\beta(\cdot) \in \mathcal{KL}$ and $\gamma_2(\cdot) \in \mathcal{K}$ such that

$$\mathbb{E}[\|\phi_{\hat{\mu}}(k; x, \mathbf{w}_k)\|] \leq \beta(|x|, k) + \gamma_2(\mathbb{E}[\|w\|]) \quad (28)$$

for all $x \in \mathcal{X}$, $\mu \in \mathcal{M}(\mathbb{W})$, and $k \in \mathbb{I}_{\geq 0}$.

The value of $\delta > 0$ in this result, however, is *not* necessarily the same as in Theorems 9, 13, or Corollary 14. For an otherwise equivalent problem (i.e., the same dynamic model, stage cost, constraints, terminal cost, etc.), the value of $\tilde{\tau} > 0$ in Assumption 5 may be significant smaller (but not larger) for $\hat{\mathbb{W}} = \{0\}$ than for sets $\hat{\mathbb{W}}$ that include more than the origin. Note that in the Proof of Theorem 9, the value of δ is given by

$$\delta := \frac{\tau - \tilde{\tau}}{\tilde{L}_f L_x}$$

However, the feasible set $\mathcal{X}$ may also be larger (but not smaller) for $\hat{\mathbb{W}} = \{0\}$ than for sets $\hat{\mathbb{W}}$ that include more than the origin. Thus, the Lipschitz constants $\tilde{L}_f, L_x > 0$ may also increase for nominal MPC relative to SMPC. Depending on the relative magnitude of these changes, the value of $\delta > 0$ may be larger or smaller for nominal MPC than SMPC, for an otherwise equivalent problem.

By explicitly requiring the terminal control law to satisfy (26), Corollary 15 does not require Assumption 6 to ensure that $\mathcal{X}$ is RPI for sufficiently small disturbances. This result is, to the best of our knowledge, new for nominal economic MPC, i.e., MPC without Assumption 6. By ensuring that $\mathcal{X}$ is RPI, we can guarantee recursive feasibility of nominal economic MPC and thereby derive the performance bound (27) that is analogous to the bounds derived for idealized versions of stochastic MPC in [7] and robust MPC in [3]. Thus, nominal

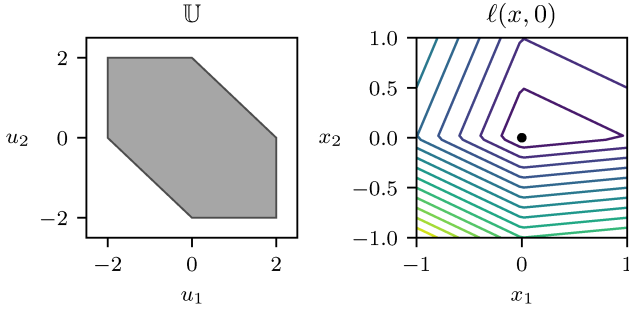


Fig. 2. Left: Input constraints $\mathbb{U}$. Right: Contour plot of $\ell(x, 0)$.

economic MPC is robust to sufficiently small disturbances in an economic context, i.e., (27). Terminal control laws that satisfy both (26) and (9), however, may prove difficult to construct for economic cost functions.

VII. NUMERICAL EXAMPLE

We consider a simple (linear) numerical example with two states (x_1, x_2) , two inputs (u_1, u_2) , and two disturbances (w_1, w_2) described by the following discrete-time dynamic equations.

$$\begin{aligned} x_1^+ &= x_1 + u_1 + w_1 \\ x_2^+ &= x_2 + u_2 + w_2 \end{aligned}$$

The dynamics of these two states are decoupled, but the input is subject to the constraints $u_1, u_2 \in [-2, 2]$ and $u_1 + u_2 \in [-2, 2]$. We plot $\mathbb{U}$ in Figure 2.

For the SMPC controller design, we define $(\hat{w}_1, \hat{w}_2) \in \hat{\mathbb{W}} := \{-0.5, 0, 0.5\} \times \{0\}$, i.e., $\hat{w}_1 \in \{-0.5, 0, 0.5\}$ and $\hat{w}_2 = 0$. We define $\hat{\mu}(\cdot)$ such that $\hat{\mu}(\{(-0.5, 0)\}) = \hat{\mu}(\{(0.5, 0)\}) = \hat{\varepsilon}_1/2$ and $\hat{\mu}(\{(0, 0)\}) = 1 - \hat{\varepsilon}_1$ for some $\hat{\varepsilon}_1 \in [0, 1]$. We use the stage cost

$$\begin{aligned} \ell(x, u) &= \max\{x_1, 0\} + 0.5 \max\{x_2, 0\} + 5 \max\{-x_1, 0\} \\ &\quad + 2.5 \max\{-x_2, 0\} + 0.5|u_1| + 0.5|u_2| \end{aligned}$$

and note that this cost is asymmetric about the origin. We plot the cost function for $u = 0$ in Figure 2. We choose the terminal cost $V_f(x) := 6|x_1| + 6|x_2|$ and define $\mathbb{X}_f := \{x : V_f(x) \leq 6\} = \{x : |x_1| + |x_2| \leq 1\}$. We define the terminal control law as the deadbeat controller $\kappa_f(x) := -x$. We use the parameterization $\pi(x, v) := \text{sat}_{\mathbb{U}}(-x + v)$ and let $\mathbb{V} := \{v : |v_1| \leq 10, |v_2| \leq 10\}$. We verify that this SMPC formulation satisfies Assumptions 3-6. Since $\hat{\mathbb{W}}$ is finite, we solve this SMPC problem exactly.

To illustrate changes in the control law with respect to changes in $\hat{\mu}$, we fix $x_2 = 0$ and calculate the values of $\kappa_{\hat{\mu}}(x)$ for $x_1 \in [-3, 3]$. In the top plot of Figure 3, we plot the first element of $\kappa_{\hat{\mu}}(x)$, denoted u_1^0 for multiple values of $\hat{\varepsilon}_1 \in [0, 1]$.⁴ We observe three distinct regions of $\hat{\varepsilon}_1$ that produce three different control laws. Moreover, we observe that $\kappa_{\hat{\mu}}(x)$ is *discontinuous* with respect to changes in $\hat{\varepsilon}_1$.

⁴The second element is zero for all $x_1 \in [-3, 3]$ and $x_2 = 0$.

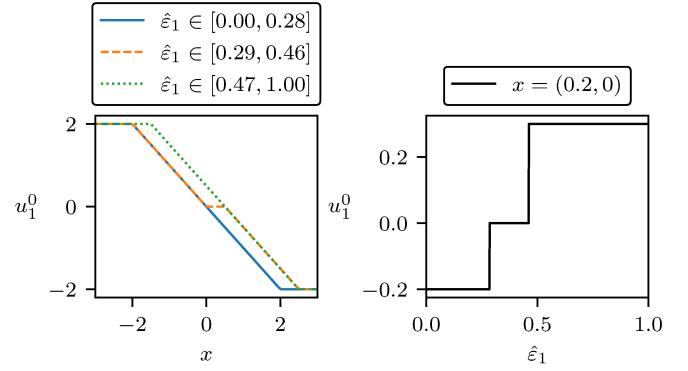


Fig. 3. Top: The first element of $\kappa_{\hat{\mu}}(x)$, denoted u_1^0 , for all $x_1 \in [-3, 3]$ and different values of $\hat{\varepsilon}_1$. Bottom: The first element of $\kappa_{\hat{\mu}}(x)$ for all $\hat{\varepsilon}_1 \in [0, 1]$ and $x = (0.2, 0)$.

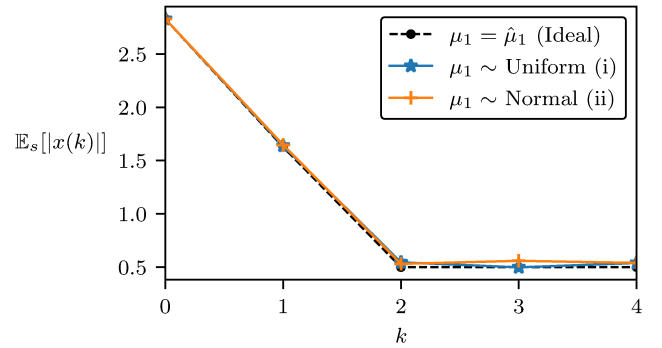


Fig. 4. Sample average (100 trajectories) of the norm of the closed-loop state, $\mathbb{E}_s[\|x(k)\|]$, for different distributions of w_1 with the same variance and $w_2 = 0$.

For $x = (0.2, 0)$, we plot the first element of $\kappa_{\hat{\mu}}(x)$ for $\hat{\varepsilon}_1 \in [0, 1]$ in the bottom plot of Figure 3 to better illustrate this discontinuity. Thus, arbitrarily small changes in $\hat{\mu}$ can produce significantly different controllers. For the exact two values of $\hat{\varepsilon}_1$ at which these discontinuities occur, we note that all values of $u_1 \in [-0.2, 0]$ and $u_1 \in [0, 0.25]$ are optimal for $x = (0.2, 0)$.

First, we demonstrate the out-of-sample performance of SMPC by simulating the closed-loop trajectory subject to a continuous distribution for w_1 . Let $\mathbb{W} := \mathbb{W}_1 \times \mathbb{W}_2$ and $\mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2)$ with $S_1 \subseteq \mathbb{W}_1$, $S_2 \subseteq \mathbb{W}_2$. Let $\mu_2(\{0\}) = 1$ and $\mathbb{W}_2 = \{0\}$, i.e., $\Pr(w_2 = 0) = 1$. We then consider two options for $\mu_1(\cdot)$: (i) a uniform distribution on $\mathbb{W}_1 := \{w_1 \in \mathbb{R} : |w_1| \leq 1/\sqrt{2}\}$ and (ii) a truncated normal distribution on $\mathbb{W}_1 := \{w_1 \in \mathbb{R} : |w_1| \leq 2\}$ with variance $1/6$. Note that the variance of both continuous distributions and the variance of $\hat{\mu}$ used the SMPC optimization problem are all equal to $1/6$. In Figure 4, we plot the sample average (100 trajectories) of the norm of the closed-loop state, $\mathbb{E}_s[\|x(k)\|]$, for these different distributions of w_1 . Note that the performance of SMPC is similar for both continuous distributions and the ideal (discrete) distribution $\mu = \hat{\mu}$.

We now introduce an unmodeled disturbance. Let $\mathbb{W} := \{-0.5, 0, 0.5\}^2$, i.e., $w_1, w_2 \in \{-0.5, 0, 0.5\}$, and $\mu(S_1 \times$

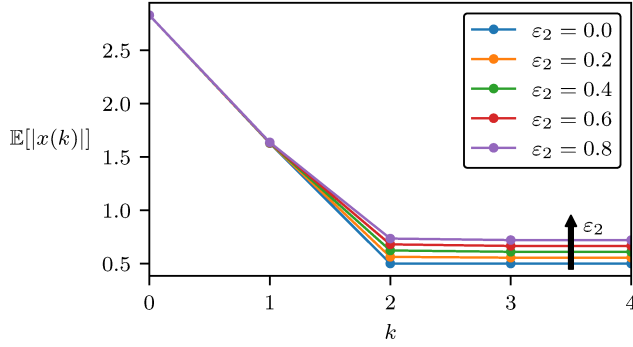


Fig. 5. Expected value of the norm of the closed-loop state, $\mathbb{E}[\|x(k)\|]$ for multiple values of ε_2 . The arrow indicates that the value of $\mathbb{E}[\|x(k)\|]$ increases with increasing ε_2 .

$S_2) = \mu_1(S_1)\mu_2(S_2)$ with $S_1, S_2 \subseteq \{-0.5, 0, 0.5\}$. For $\varepsilon_1, \varepsilon_2 \geq 0$, let $\mu_i(\{-0.5\}) = \mu_i(\{0.5\}) = \varepsilon_i/2$ and $\mu_i(\{0\}) = 1 - \varepsilon_i$. We assume that the distribution for w_1 is chosen correctly in the SMPC problem, i.e., $\varepsilon_1 = \hat{\varepsilon}_1 = 2/3$, but we now include the unmodeled disturbance w_2 via ε_2 . We note that $W(\mu, \hat{\mu}) = \varepsilon_2/2$. Starting from $x_0 = (2, 2)$, we simulate the closed-loop trajectories for SMPC subject to this unmodeled disturbance for multiple values of ε_2 . In Figure 5, we plot the expected norm of the closed-loop state trajectory ($\mathbb{E}[\|x(k)\|] = \mathbb{E}[\|\phi_{\hat{\mu}}(k; x_0, \mathbf{w}_k)\|]$) for multiple values of ε_2 . Note that this expected value is exact since $\mathbb{W}$ is finite. As ε_2 and therefore $W(\mu, \hat{\mu})$ increase, the value of $\mathbb{E}[\|x(k)\|]$ increases for each $k \geq 2$. This behavior is consistent with DRASiE and Theorem 9.

To demonstrate the robustness of SMPC to differences between $\mathbb{W}$ and $\hat{\mathbb{W}}$, we first compute $\mathcal{X}$ for the disturbance set $\hat{\mathbb{W}}$. We then sample this feasible region and simulate one step of the closed-loop trajectory from these points subject to $\hat{\mathbb{W}}$ to construct the region

$$\mathcal{X}^+ := \{f(x, \kappa_{\hat{\mu}}(x), \hat{w}) : x \in \mathcal{X}, \hat{w} \in \hat{\mathbb{W}}\}$$

We plot both of these sets in Figure 6. Note that $\mathcal{X}^+$ is a strict subset of $\mathcal{X}$. Thus, there exists some margin of additional disturbances that can be injected into the closed-loop system such that $\mathcal{X}$ remains RPI, i.e., SMPC maintains some nonzero margin of robustness.

We can in fact compute the largest $\delta > 0$ such that $\mathcal{X}$ is RPI for the closed-loop system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ and any $\mathbb{W} \in \mathbb{R}^2$ that satisfies $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$. We find a value of $\delta = (3/4)\sqrt{2}$ for this SMPC formulation as shown in Figure 6. For nominal MPC ($\hat{\mathbb{W}} = \{0\}$), the feasible set is larger than SMPC and we have $\delta_{MPC} = \sqrt{2}$. Note that this value is larger than the value of δ computed for SMPC, corroborating the previous discussion that followed Corollary 15.

VIII. CONCLUSIONS

Analogous to the inherent robustness of nominal MPC to sufficient small errors in the deterministic dynamic model, we established that SMPC is inherently robust to sufficient small errors in the stochastic dynamic model. In both cases, this

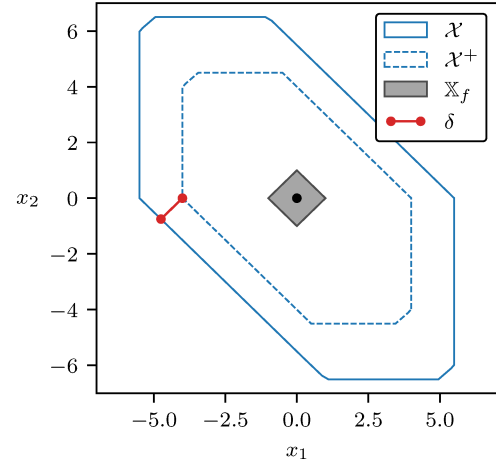


Fig. 6. The feasible set $\mathcal{X}$ for SMPC and the set of states after one step of the closed-loop trajectory subject to $\hat{w} \in \hat{\mathbb{W}}$, denoted $\mathcal{X}^+$.

inherent robustness is afforded by feedback. These errors may include disturbances that were entirely absent from the disturbance model and errors introduced by scenario-based approximations of the stochastic optimization problem. Moreover, this same result allows us to characterize the stochastic robustness of other MPC formulations, e.g., constraint-tightened and nominal MPC, as special cases of SMPC, thereby unifying the analysis of these three different problem formulations. In fact, this approach revealed a means to design the terminal region for nominal economic MPC such that the feasible set remains RPI for sufficiently small disturbances.

In addition, these results can likely be extend to suboptimal SMPC algorithms via the same approach used in [1], thereby alleviating some of the large computational burden imposed by stochastic optimization. We emphasize that this definition of distributional robustness for closed-loop systems is not restricted to SMPC and is applicable to the larger field of stochastic optimal control and, potentially, the developing field of distributionally robust control.

APPENDIX

Proof of Proposition 8. Define $\alpha_4(s) := \alpha_3(\alpha_2^{-1}(s/2))$ and note that $\alpha_4(\cdot) \in \mathcal{K}_\infty$ and $\alpha_4(s) \leq s$ for all $s \in \mathbb{R}_{\geq 0}$ because $\alpha_3(s) \leq \alpha_2(s)$ for all $s \in \mathbb{R}_{\geq 0}$.⁵ We have the following inequality.

$$\begin{aligned} \alpha_4(V_{\hat{\mu}}(x)) &\leq \alpha_4(\alpha_2(|x|) + \sigma_2(\hat{\mathbb{E}}[\|\hat{w}\|])) \\ &\leq \alpha_4(2\alpha_2(|x|)) + \alpha_4(2\sigma_2(\hat{\mathbb{E}}[\|\hat{w}\|])) \\ &= \alpha_3(|x|) + \alpha_4(2\sigma_2(\hat{\mathbb{E}}[\|\hat{w}\|])) \end{aligned}$$

By rearranging, we have $-\alpha_3(|x|) \leq -\alpha_4(V_{\hat{\mu}}(x)) + \alpha_4(2\sigma_2(\hat{\mathbb{E}}[\|\hat{w}\|]))$ and therefore

$$\begin{aligned} \int_{\mathbb{W}} V_{\hat{\mu}}(f(x, \kappa_{\hat{\mu}}(x), w)) d\mu(w) \\ \leq V_{\hat{\mu}}(x) - \alpha_4(V_{\hat{\mu}}(x)) + \tilde{\sigma}_3(\hat{\mathbb{E}}[\|\hat{w}\|]) + \sigma_4(W(\mu, \hat{\mu})) \end{aligned}$$

⁵If this inequality does not hold, we simply construct a new $\tilde{\alpha}_2(\cdot) \in \mathcal{K}_\infty$ such that $\alpha_2(s) \leq \tilde{\alpha}_2(s)$ and $\alpha_3(s) \leq \tilde{\alpha}_2(s)$.

in which $\tilde{\sigma}_3(s) := \sigma_3(s) + \alpha_4(2\sigma_2(s))$ and $\tilde{\sigma}_3(\cdot) \in \mathcal{K}$. Since $\mathcal{X}$ and $\mathbb{W}$ are bounded, there exists $b \geq 0$ such that $V_{\hat{\mu}}(x) \leq \alpha_2(|x|) + \sigma_2(\hat{\mathbb{E}}[|\hat{w}|]) \leq b$ for all $x \in \mathcal{X}$, $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ and corresponding $\hat{\mathbb{E}}[|\hat{w}|]$. Using Lemma 6 in [19], we can construct $\alpha_v(\cdot) \in \mathcal{K}_\infty$ such that $\alpha_v(\cdot)$ is convex and $\alpha_v(V_{\hat{\mu}}(x)) \leq \alpha_4(V_{\hat{\mu}}(x))$ for all $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. Therefore, we have

$$\begin{aligned} & \int_{\mathbb{W}} V_{\hat{\mu}}(f(x, \kappa_{\hat{\mu}}(x), w)) d\mu(w) \\ & \leq V_{\hat{\mu}}(x) - \alpha_v(V_{\hat{\mu}}(x)) + \tilde{\sigma}_3(\hat{\mathbb{E}}[|\hat{w}|]) + \sigma_4(W(\mu, \hat{\mu})) \end{aligned}$$

For arbitrary $x \in \mathcal{X}$, $\hat{\mu} \in \mathcal{M}(\mathbb{W})$, and $\mu \in \mathcal{M}(\mathbb{W})$, let $x(k) := \phi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ for all $k \in \mathbb{I}_{\geq 0}$. To streamline notation, we also define $c(\mu, \hat{\mu}) := \tilde{\sigma}_3(\hat{\mathbb{E}}[|\hat{w}|]) + \sigma_4(W(\mu, \hat{\mu}))$. By the law of total expectation and Jensen's inequality, we have

$$\begin{aligned} & \mathbb{E}[V_{\hat{\mu}}(x(k+1))] \\ & \leq \mathbb{E}[V_{\hat{\mu}}(x(k))] - \alpha_v(\mathbb{E}[V_{\hat{\mu}}(x(k))]) + c(\mu, \hat{\mu}) \end{aligned}$$

Note that $\mathbb{E}[\cdot]$ is defined with respect to μ and not $\hat{\mu}$.

Define $\tilde{\gamma}(s) := 2 \max\{\alpha_v^{-1}(s), s\}$ and note that $\tilde{\gamma}(\cdot) \in \mathcal{K}$. If $\mathbb{E}[V_{\hat{\mu}}(x(k))] \leq \tilde{\gamma}(c(\mu, \hat{\mu}))/2$, then

$$\begin{aligned} \mathbb{E}[V_{\hat{\mu}}(x(k+1))] & \leq \tilde{\gamma}(c(\mu, \hat{\mu}))/2 + c(\mu, \hat{\mu}) \\ & \leq \tilde{\gamma}(c(\mu, \hat{\mu}))/2 + \tilde{\gamma}(c(\mu, \hat{\mu}))/2 \\ & = \tilde{\gamma}(c(\mu, \hat{\mu})) \end{aligned}$$

If $\tilde{\gamma}(c(\mu, \hat{\mu}))/2 \leq \mathbb{E}[V_{\hat{\mu}}(x(k))]$, then

$$\begin{aligned} & \mathbb{E}[V_{\hat{\mu}}(x(k+1))] \\ & \leq \mathbb{E}[V_{\hat{\mu}}(x(k))] - \alpha_v(\tilde{\gamma}(c(\mu, \hat{\mu}))/2) + c(\mu, \hat{\mu}) \\ & \leq \mathbb{E}[V_{\hat{\mu}}(x(k))] \leq \tilde{\gamma}(c(\mu, \hat{\mu})) \end{aligned}$$

Thus, for $\mathbb{E}[V_{\hat{\mu}}(x(k))] \leq \tilde{\gamma}(c(\mu, \hat{\mu}))$, we know that

$$\mathbb{E}[V_{\hat{\mu}}(x(k+1))] \leq \tilde{\gamma}(c(\mu, \hat{\mu})) \quad (29)$$

If $\mathbb{E}[V_{\hat{\mu}}(x(k))] \geq \tilde{\gamma}(c(\mu, \hat{\mu}))$, we have

$$\mathbb{E}[V_{\hat{\mu}}(x(k+1))] \leq \lambda_1(\mathbb{E}[V_{\hat{\mu}}(x(k))])$$

in which $\lambda_1(s) := s - \alpha_v(s) + \alpha_v(s/2)$. We have that $\lambda_1(\cdot)$ is continuous, $\lambda_1(0) = 0$, and $\lambda_1(s) < s$ for all $s > 0$. By the same process used in [23, Theorem B.15], we construct $\lambda(\cdot) \in \mathcal{K}_\infty$ such that $\lambda_1(s) \leq \lambda(s) < s$ for $s > 0$. Thus, we have

$$\mathbb{E}[V_{\hat{\mu}}(x(k+1))] \leq \lambda(\mathbb{E}[V_{\hat{\mu}}(x(k))]) \quad (30)$$

Repeated application of (30) and the fact that $\mathbb{E}[V_{\hat{\mu}}(x(0))] = V_{\hat{\mu}}(x)$ gives

$$\mathbb{E}[V_{\hat{\mu}}(x(k))] \leq \tilde{\beta}(V_{\hat{\mu}}(x), k) := \lambda^k(V_{\hat{\mu}}(x)) \quad (31)$$

in which $\lambda^k(\cdot)$ is the composition of $\lambda(\cdot)$ with itself k times. Using the same approach as [23, Theorem B.15], we conclude that $\tilde{\beta}(\cdot) \in \mathcal{KL}$. We combine (29) and (31) to give

$$\mathbb{E}[V_{\hat{\mu}}(x(k))] \leq \max\{\tilde{\beta}(V_{\hat{\mu}}(x), k), \tilde{\gamma}(c(\mu, \hat{\mu}))\}$$

Using [19, Lemma 6] and the fact that $\mathcal{X}$ is bounded, we construct a convex function $\alpha_{1,v}(\cdot) \in \mathcal{K}_\infty$ such that

$\alpha_{1,v}(|x|) \leq \alpha_1(|x|) \leq V_{\hat{\mu}}(x)$ for all $x \in \mathcal{X}$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. Thus, we apply Jensen's inequality to give

$$\alpha_{1,v}(\mathbb{E}[|x|]) \leq \mathbb{E}[\alpha_{1,v}(|x|)] \leq \mathbb{E}[V_{\hat{\mu}}(x)]$$

and therefore

$$\begin{aligned} \mathbb{E}[|x(k)|] & \leq \max\left\{\alpha_{1,v}^{-1}\left(\tilde{\beta}(V_{\hat{\mu}}(x), k)\right), \alpha_{1,v}^{-1}(\tilde{\gamma}(c(\mu, \hat{\mu})))\right\} \\ & \leq \tilde{\beta}_1(V_{\hat{\mu}}(x), k) + \tilde{\gamma}_2(c(\mu, \hat{\mu})) \end{aligned}$$

in which $\tilde{\beta}_1(\cdot) := \alpha_{1,v}^{-1}(\tilde{\beta}(\cdot)) \in \mathcal{KL}$ and $\tilde{\gamma}_2(\cdot) := \alpha_{1,v}^{-1}(\tilde{\gamma}(\cdot)) \in \mathcal{K}$. We use the upper bound for $V_{\hat{\mu}}(x)$ to give

$$\begin{aligned} \mathbb{E}[|x(k)|] & \leq \tilde{\beta}_1(\alpha_2(|x|) + \sigma_2(\hat{\mathbb{E}}[|\hat{w}|]), k) + \tilde{\gamma}_1(c(\mu, \hat{\mu})) \\ & \leq \tilde{\beta}_1(2\alpha_2(|x|), k) + \tilde{\beta}_1(\sigma_2(\hat{\mathbb{E}}[|\hat{w}|]), k) \\ & \quad + \tilde{\gamma}_2(c(\mu, \hat{\mu})) \\ & \leq \beta(|x|, k) + \tilde{\gamma}_1(\hat{\mathbb{E}}[|\hat{w}|]) + \tilde{\gamma}_2(c(\mu, \hat{\mu})) \quad (32) \end{aligned}$$

in which $\beta(s, k) := \tilde{\beta}_1(2\alpha_2(s), k) \in \mathcal{KL}$ and $\tilde{\gamma}_1(s) := \tilde{\beta}_1(2\sigma_2(s), 0) \in \mathcal{K}$.

We now unpack the function $c(\mu, \hat{\mu})$ to give

$$\begin{aligned} & \tilde{\gamma}_1(\hat{\mathbb{E}}[|\hat{w}|]) + \tilde{\gamma}_2(c(\mu, \hat{\mu})) \\ & = \tilde{\gamma}_1(\hat{\mathbb{E}}[|\hat{w}|]) + \tilde{\gamma}_2(2\tilde{\sigma}_3(\hat{\mathbb{E}}[|\hat{w}|])) + \tilde{\gamma}_2(2\sigma_4(W(\mu, \hat{\mu}))) \\ & = \gamma_1(\hat{\mathbb{E}}[|\hat{w}|]) + \gamma_2(W(\mu, \hat{\mu})) \end{aligned}$$

in which $\gamma_1(s) := \tilde{\gamma}_1(s) + \tilde{\gamma}_2(2\tilde{\sigma}_3(s)) \in \mathcal{K}$ and $\gamma_2(s) := \tilde{\gamma}_2(2\sigma_4(s)) \in \mathcal{K}$. Substitute these definitions into (32) to complete the proof. $\square$

Proof of Lemma 10. Let $L \geq 0$ denote the Lipschitz constant for $g(\cdot)$. For $x_1, x_2 \in X$ we have

$$\begin{aligned} |G(x_1) - G(x_2)| & = \left| \int_S (g(x_1, w) - g(x_2, w)) d\mu(w) \right| \\ & \leq \int_S |g(x_1, w) - g(x_2, w)| d\mu(w) \\ & \leq \int_S L|x_1 - x_2| d\mu(w) \\ & = L|x_1 - x_2| \end{aligned}$$

in which the final equality holds because $\mu \in \mathcal{M}(S)$ is a probability measure, i.e., $\int_S d\mu(w) = 1$. Thus, $G(\cdot)$ is a Lipschitz continuous function with the same Lipschitz constant as $g(\cdot)$. $\square$

Proof of Lemma 11. Since $V_f(\cdot)$, $f(\cdot)$, $\kappa_f(\cdot)$ are locally Lipschitz continuous and $\mathbb{X}_f, \hat{\mathbb{W}}$ are bounded, there exists $L_f > 0$ such that

$$|V_f(f(x, \kappa_f(x), \hat{w})) - V_f(f(x, \kappa_f(x), 0))| \leq L_f|\hat{w}|$$

for all $x \in \mathbb{X}_f$ and $\hat{w} \in \hat{\mathbb{W}}$. Therefore,

$$\begin{aligned} & \int_{\hat{\mathbb{W}}} V_f(f(x, \kappa_f(x), \hat{w})) d\hat{\mu}(\hat{w}) \\ & \leq V_f(f(x, \kappa_f(x), 0)) + L_f \hat{\mathbb{E}}[|\hat{w}|] \end{aligned}$$

We use this bound with (9) to give (13). $\square$

Proof of Lemma 12. Choose $x \in \mathbb{X}_f$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ and consider a trajectory generated by repeated application of the

terminal control law, i.e., $x(k) = \hat{\phi}(k; x, \mathbf{0}, \hat{\mathbf{w}})$ since $\pi(x, 0) = \kappa_f(x)$. The set $\mathbb{X}_f$ is RPI for $\hat{\mathbb{W}}$ and this control law due to Assumption 5 and therefore $x(k) \in \mathbb{X}_f$ for all $k \in \mathbb{I}_{[0, N]}$ and $\hat{\mathbf{w}} \in \hat{\mathbb{W}}^N$. From Assumption 5 and Lemma 11, we have that

$$\begin{aligned} & \int_{\hat{\mathbb{W}}^N} (V_f(x(k+1)) - V_f(x(k))) d\hat{\mu}^N(\hat{\mathbf{w}}) \\ & \leq - \int_{\hat{\mathbb{W}}^N} \ell(x(k), \kappa_f(x(k))) d\hat{\mu}^N(\hat{\mathbf{w}}) + L_f \hat{\mathbb{E}}[|\hat{w}|] \end{aligned}$$

We sum both sides of the inequality from $k = 0$ to $k = N - 1$ to give

$$\begin{aligned} & \int_{\hat{\mathbb{W}}^N} (V_f(x(N)) - V_f(x(0))) d\hat{\mu}^N(\hat{\mathbf{w}}) \\ & \leq - \int_{\hat{\mathbb{W}}^N} \sum_{k=0}^{N-1} \ell(x(k), \kappa_f(x(k))) d\hat{\mu}^N(\hat{\mathbf{w}}) + N L_f \hat{\mathbb{E}}[|\hat{w}|] \end{aligned}$$

By rearranging and using the definition of $J(\cdot)$ and $x = x(0)$, we have

$$\int_{\hat{\mathbb{W}}^N} J(x, \mathbf{0}, \hat{\mathbf{w}}) d\hat{\mu}^N(\hat{\mathbf{w}}) \leq V_f(x) + N L_f \hat{\mathbb{E}}[|\hat{w}|]$$

By optimality, we have

$$V_{\hat{\mu}}^0(x) \leq V_f(x) + N L_f \hat{\mathbb{E}}[|\hat{w}|]$$

Since the choice of $x \in \mathbb{X}_f$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ was arbitrary, this inequality must hold for all $x \in \mathbb{X}_f$ and $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$.

We now define

$$H(x) := \max \left\{ \sup_{\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})} \left(V_{\hat{\mu}}^0(x) - N L_f \hat{\mathbb{E}}[|\hat{w}|] \right), 0 \right\}$$

and note that $0 \leq H(x) \leq V_f(x)$ for all $x \in \mathbb{X}_f$. Since $V_f(\cdot)$ is continuous, $H(0) = V_f(0) = 0$, and $\mathbb{X}_f$ contains the origin in its interior, we know that $H(x)$ is continuous at the origin. We also establish that $H(\cdot)$ is locally bounded on $\mathcal{X}$. Let X be a compact subset of $\mathcal{X}$. The function $J(\cdot)$ is a composition of a finite number of continuous functions and is therefore continuous. Thus, $J(\cdot)$ has an upper bound on the compact set $X \times \mathbb{V}^N \times \hat{\mathbb{W}}^N$. Since $\mathcal{V}(x) \subseteq \mathbb{V}^N$ for all $x \in \mathcal{X}$, $V_{\hat{\mu}}^0(\cdot)$ must satisfy the same upper bound for all $x \in X$, $\mu \in \hat{\mathcal{M}}(\mathbb{W})$. Thus, $H(x)$ satisfies this same upper bound because $H(x) \leq \sup_{\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})} V_{\hat{\mu}}^0(x)$. Since $0 \leq H(x)$ also and the choice of X was arbitrary, $H(x)$ is locally bounded on $\mathcal{X}$.

Since $H(x)$ is locally bounded, satisfies $H(0) = 0$, and is continuous at $x = 0$, we can apply [22, Prop. 14] to construct $\alpha_2(\cdot) \in \mathcal{K}_{\infty}$ such that $H(x) \leq \alpha_2(|x|)$ for all $x \in \mathcal{X}$. Furthermore, we have that

$$V_{\hat{\mu}}^0(x) - N L_f \hat{\mathbb{E}}[|\hat{w}|] \leq H(x) \leq \alpha_2(|x|)$$

for all $x \in \mathcal{X}$ and $\mu \in \hat{\mathcal{M}}(\mathbb{W})$. Define $\sigma_2(s) = N L_f s \in \mathcal{K}$ and rearrange to complete the proof. $\square$

REFERENCES

- [1] D. A. Allan, C. N. Bates, M. J. Risbeck, and J. B. Rawlings. On the inherent robustness of optimal and suboptimal nonlinear MPC. *Sys. Cont. Let.*, 106:68 – 78, 2017. ISSN 0167-6911. doi: 10.1016/j.sysconle.2017.03.005.
- [2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [3] F. A. Bayer, M. Lorenzen, M. A. Müller, and F. Allgöwer. Robust economic model predictive control using stochastic information. *Automatica*, 74:151–161, 2016.
- [4] M. Cannon, B. Kouvaritakis, and D. Ng. Probabilistic tubes in linear stochastic model predictive control. *Sys. Cont. Let.*, 58(10-11):747–753, 2009.
- [5] M. Cannon, B. Kouvaritakis, and X. Wu. Probabilistic constrained mpc for multiplicative and additive stochastic uncertainty. *IEEE Trans. Auto. Cont.*, 54(7):1626–1632, 2009.
- [6] M. Cannon, B. Kouvaritakis, S. V. Rakovic, and Q. Cheng. Stochastic tubes in model predictive control with probabilistic constraints. *IEEE Trans. Auto. Cont.*, 56(1):194–200, 2010.
- [7] D. Chatterjee and J. Lygeros. On stability and performance of stochastic predictive control techniques. *IEEE Trans. Auto. Cont.*, 60(2):509–514, 2014.
- [8] M. Farina, L. Giulioni, and R. Scattolini. Stochastic linear model predictive control with chance constraints—a review. *J. Proc. Cont.*, 44: 53–67, 2016.
- [9] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory and Relat. Fields*, 162(3):707–738, 2015.
- [10] C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio. Learning with a Wasserstein loss. *Adv. Neural Inf. Process. Syst.*, 28, 2015.
- [11] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel. Examples when nonlinear model predictive control is nonrobust. *Automatica*, 40:1729–1738, 2004.
- [12] J. L. Jerez, E. C. Kerrigan, and G. A. Constantinides. A condensed and sparse QP formulation for predictive control. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 5217–5222. IEEE, 2011.
- [13] E. C. Kerrigan and J. M. Maciejowski. Soft constraints and exact penalty functions in model predictive control. In *Control 2000 Conference, Cambridge*, pages 2319–2327, 2000.
- [14] B. Kouvaritakis, M. Cannon, S. V. Raković, and Q. Cheng. Explicit use of probabilistic distributions in linear predictive control. *Automatica*, 46 (10):1719 – 1724, 2010. ISSN 0005-1098. doi: <https://doi.org/10.1016/j.automatica.2010.06.034>.
- [15] M. Lorenzen, F. Dabbene, R. Tempo, and F. Allgöwer. Constraint-tightening and stability in stochastic model predictive control. *IEEE Trans. Auto. Cont.*, 62(7):3165–3177, 2016.
- [16] D. Q. Mayne. Robust and stochastic model predictive control: Are we going in the right direction? *Annual Rev. Control*, 41:184 – 192, 2016. ISSN 1367-5788. doi: 10.1016/j.arcontrol.2016.04.006.
- [17] D. Q. Mayne and P. Falugi. Stabilizing conditions for model predictive control. *Int. J. Robust and Nonlinear Control*, 29(4):894–903, 2019.
- [18] R. D. McAllister and J. B. Rawlings. The stochastic robustness of nominal and stochastic model predictive control. *IEEE Trans. Auto. Cont.*, 2022. doi: <https://doi.org/10.1109/TAC.2022.3226712>. Online Early Access.
- [19] R. D. McAllister and J. B. Rawlings. Nonlinear stochastic model predictive control: Existence, measurability, and stochastic asymptotic stability. *IEEE Trans. Auto. Cont.*, 68(3):1524–1536, March 2023. doi: <https://doi.org/10.1109/TAC.2022.3157131>.
- [20] A. Mesbah. Stochastic model predictive control. *IEEE Ctl. Sys. Mag.*, pages 30–44, Dec 2016.
- [21] J. A. Primbs and C. H. Sung. Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise. *IEEE Trans. Auto. Cont.*, 54(2):221–230, 2009.
- [22] J. B. Rawlings and M. J. Risbeck. On the equivalence between statements with epsilon-delta and K-functions. Technical Report 2015–01, TWCCC Technical Report, December 2015. URL <https://engineering.ucsb.edu/~jbraw/jbrweb-archives/tech-reports/twccc-2015-01.pdf>.
- [23] J. B. Rawlings, D. Q. Mayne, and M. M. Diehl. *Model Predictive Control: Theory, Design, and Computation*. Nob Hill Publishing, Santa Barbara, CA, 2nd, paperback edition, 2020. 770 pages, ISBN 978-0-9759377-5-4.
- [24] J. A. Rossiter, B. Kouvaritakis, and M. J. Rice. A numerically robust state-space approach to stable-predictive control strategies. *Automatica*, 34(1):65–73, 1998.
- [25] P. O. M. Scokaert and J. B. Rawlings. Feasibility issues in linear model predictive control. *AIChE J.*, 45(8):1649–1659, Aug 1999.
- [26] S. Shafieezadeh Abadeh, V. A. Nguyen, D. Kuhn, and P. M. Mohajerin Esfahani. Wasserstein distributionally robust Kalman filtering. *Adv. Neural Inf. Process. Syst.*, 31, 2018.

- [27] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā*, 19(1/2):23–26, 1958.
- [28] C. Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [29] I. Yang. Wasserstein distributionally robust stochastic control: A data-driven approach. *IEEE Trans. Auto. Cont.*, 66(8):3863–3870, 2020.
- [30] S. Yu, M. Reble, H. Chen, and F. Allgöwer. Inherent robustness properties of quasi-infinite horizon nonlinear model predictive control. *Automatica*, 50(9):2269 – 2280, 2014. ISSN 0005-1098. doi: 10.1016/j.automatica.2014.07.014.
- [31] A. Zheng and M. Morari. Control of linear unstable systems with constraints. In *American Control Conference, Seattle, Washington*, pages 3704–3708, 1995.



Robert D. McAllister Robert D. McAllister received the Bachelor of Chemical Engineering degree from the University of Delaware in 2017 and the Ph.D. degree in Chemical Engineering from the University of California, Santa Barbara in 2022. He was a postdoctoral researcher at TU Delft in the Delft Center for Systems and Control for six months. He is currently an assistant professor at TU Delft in the Delft Center for Systems and Controls. His research interests include model predictive control, closed-loop scheduling, stochastic and distributional

robustness of closed-loop systems, and data-driven control methods with a particular focus on agricultural applications of these methods.



James B. Rawlings James B. Rawlings received the BS from the University of Texas at Austin and the Ph.D. from the University of Wisconsin–Madison, both in Chemical Engineering. He spent one year at the University of Stuttgart as a NATO postdoctoral fellow and then joined the faculty at the University of Texas at Austin. He moved to the University of Wisconsin–Madison in 1995, and then to the University of California, Santa Barbara in 2018, where he is currently the Mellichamp Process Control Chair in the Department of Chemical Engineering, and the

codirector of the Texas-Wisconsin-California Control Consortium (TWCCC). He is a fellow of IFAC, IEEE, and AIChE.

Professor Rawlings' research interests are in the areas of chemical process modeling, monitoring and control, nonlinear model predictive control, moving horizon state estimation, and molecular-scale chemical reaction engineering. He has written numerous research articles and coauthored three textbooks: “Model Predictive Control: Theory Computation, and Design,” 2nd ed. (2020), with David Mayne and Moritz Diehl, “Modeling and Analysis Principles for Chemical and Biological Engineers” (2013), with Mike Graham, and “Chemical Reactor Analysis and Design Fundamentals,” 2nd ed. (2020), with John Ekerdt.