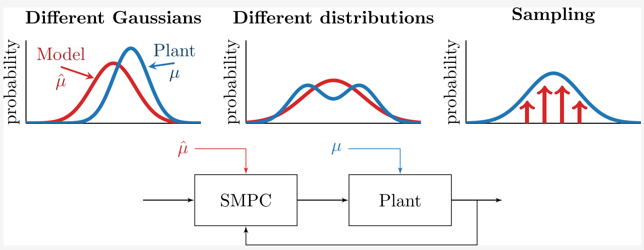
I&EC research
Industrial & Engineering Chemistry Research
pubs.acs.org/IECR
Article

On Using Feedback Control to Contend with Nature's Randomness

Robert D. McAllister* and James B. Rawlings

Cite This: *Ind. Eng. Chem. Res.* 2023, 62, 2175−2190
Read Online
ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Probability distributions are often used to characterize the randomness of nature. In stochastic model predictive control (SMPC), disturbances are described by a probability distribution that is used within a stochastic optimization problem to construct a feedback control law. While powerful, these probability distributions are themselves subject to their own type of uncertainty, often called distributional uncertainty. In this work, we establish that SMPC, under suitable assumptions, provides a nonzero margin of robustness to this distributional uncertainty. This inherent distributional robustness is afforded by feedback and careful algorithm design. Through a small example, we demonstrate the implications of this result for incorrectly modeled, out-of-sample, and even unmodeled disturbances. This result also covers scenario-based approximations of stochastic optimal control problems and unifies the description of robustness for nominal and stochastic model predictive control.



1. FOREWORD: SOME RECOLLECTIONS OF TUNDE OGUNNAIKE

During my (R.D.M.) time as an undergraduate student at the University of Delaware between 2013 and 2017, Professor Ogunnaike was the Dean of the College of Engineering. Despite what I assume was a demanding schedule as Dean, Prof. Ogunnaike made it a point to sneak into many chemical engineering courses as a guest lecturer. The first of these lectures I was fortunate enough to attend was in the Introduction to Engineering course required for all of the freshman engineering majors. Prof. Ogunnaike, of course, devoted the lecture to probability and statistics. Perhaps less important than the material covered was the statement made by this choice. The Dean of the College of Engineering, with a long and acclaimed career in industry and academia, thought that the most important thing he could tell the freshman engineering class was about probability and statistics. Clearly, that point stuck with me for the years to come.

Although not directly, Prof. Ogunnaike's influence was still felt in the chemical engineering curriculum at the University of Delaware. The process dynamics and control course was designed by Prof. Ogunnaike and taught from his book.[1] The chemical engineering department even offered a probability and statistic course based another one of Prof. Ogunnaike's books.[2] Prof. Ogunnaike's efforts undoubtedly shaped my education for the better with effects that still resonate in my doctoral research in stochastic dynamical systems and control. While I only got the opportunity to interact with Prof. Ogunnaike on a few occasions, his friendly and considerate nature was apparent. During my senior year, I decided to pursue a Ph.D. at the University of Wisconsin following my graduation. At the commencement ceremony, as I walked across the stage to shake his hand and accept my diploma, Prof. Ogunnaike had only two words for me: "Go Badgers!"

I (J.B.R.) first met Tunde Ogunnaike on a cold January day in 1980 in Madison, WI, when I arrived to enroll in the chemical engineering Ph.D. program at the University of Wisconsin. Tunde took me under his wing, showed me where the best apartments close to campus were located, and introduced me around to the other graduate students who would become my colleagues for the next five years. I later chose Professor Harmon Ray as my research advisor, and since Tunde was also one of Harmon's Ph.D. students, we became close colleagues during graduate school and lifelong friends. Besides his quiet and thoughtful demeanor and his kindness toward other people, one of the first professional attributes I noticed about Tunde was his enthusiasm for and deep understanding of probability and statistics. Where I might investigate the Physics department's graduate offerings for relevant courses in mechanics, Tunde would scour UW's outstanding Statistics department's graduate courses. He took that department's two-semester required graduate sequence for statistics Ph.D. students. He took so many graduate courses that he earned an M.S. in Statistics while doing his Ph.D. in Chemical Engineering. Even though I knew very little about

Special Issue: In Honor of Babatunde A. Ogunnaike

Received: August 19, 2022
Revised: November 10, 2022
Accepted: November 10, 2022
Published: December 6, 2022

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

the subject, Tunde would tutor me on various fundamental issues, such as the meaning and interpretation of Bayesian statistics, and we would have long, philosophical debates about the theoretical underpinnings of the subject. Although I successfully avoided formal education in statistics as a graduate student, it was a choice I would later regret when I wanted to understand better how to model uncertainty in dynamical systems, especially when designing feedback controllers for these systems. Later, as an Assistant Professor at the University of Texas, I had to buckle down and teach myself probability and statistics. That is one of the reasons I encourage my own graduate students to take as many graduate courses as they wish during their Ph.D. studies. It is just much easier to learn a subject when an expert has organized all the material and is going to explain it to you and answer your questions.

Much later, when coauthoring a graduate engineering text on mathematical modeling, I wrote up my own take on what probability and statistics mean and why they are useful, starting with the axioms of probability, and working through random variables, sampling, conditional probability, and all the other, by then, indispensable tools of this subject that find application in engineering. Tunde was the person I would show the early drafts of the probability chapter, and he gave me lots of helpful feedback. About this time, his own text on randomness and random variables was nearing completion,[2] and he would send me drafts of various chapters. Likely the most illuminating discussions Tunde and I had during this period concerned the notion of convergence of samples of a random variable to the probability distribution of that random variable. I think we both learned something from those discussions. One illustration of those ideas is given in Figure 2 of this paper.

The content of this paper is the natural consequence of my many discussions with Tunde over many years. Nature is random. You want to control some part of nature. So how do you design a control system to contend with nature's randomness? And what precisely can you say about how well your designed control system will perform at this task?

2. INTRODUCTION

Disturbances and uncertainty are an inescapable part of any engineering system. In process control, the preferred method to address these potential disturbances and uncertainty is through *feedback*. By reacting to these disturbances as they are observed, we imbue even simple control algorithms such as PID with some margin of robustness to these disturbances. This margin of robustness allows the control algorithm to withstand and overcome the small perturbations and model uncertainty that are ubiquitous in applications. For this reason, feedback is an essential component of even advanced, optimization-based control algorithms such as model predictive control (MPC).

In nominal MPC, we use a nominal dynamical model of the system to predict the future states of the plant given a trajectory of inputs over a finite horizon. We then solve an optimization problem to select the optimal trajectory of inputs for the system given some measure of performance that we call a stage cost. We implement only the *first* input in this optimal trajectory. After a fixed time interval, we then estimate the new state of the system based on measurements (i.e., feedback) and recompute the optimal input trajectory from this new state, again implementing only the first input of this trajectory. We repeat this process at each time step and therefore construct an algorithm that includes both optimization and feedback from the system.

In a significant contribution, Grimm et al.[3] demonstrate that certain (nonlinear) nominal MPC formulations may produce suitable performance for the idealized case (i.e., no disturbances) and yet afford zero margin of robustness to disturbances. For example, arbitrarily small disturbances can produce unstable closed-loop systems without careful algorithm design. After a decade of further research, Yu et al.[4] demonstrate sufficient conditions for the MPC problem formulation that guarantee a nonzero margin of robustness to disturbances. Allan et al.[5] extend these results to a suboptimal MPC algorithm and more general terminal conditions that admit discrete-valued inputs. We refer to this property as *inherent* robustness because this robustness is afforded by feedback and does not require any disturbance information in the nominal MPC optimization problem.

While the inherent robustness of nominal MPC is often sufficient in industrial practice, there may arise applications with unusually stringent safety requirements or performance demands, or with large uncertainty that is poorly captured via a nominal dynamical model of the system. For these applications, stochastic MPC (SMPC) offers a method with the potential to improve on the robustness of nominal MPC by including a stochastic description of the disturbance directly in the problem formulation. The SMPC optimization problem typically involves minimizing the expected value of the stage cost subject to deterministic and probabilistic constraints.[6−8] SMPC still incorporates feedback via the same rolling horizon approach as nominal MPC (i.e., the stochastic optimization problem is solved at each sampling time with an updated state estimate to determine the input at that time). For more details on approximating and solving both linear and nonlinear SMPC problems see Mesbah[7] and the references therein. While solution methods for these SMPC problems are important research topics, we instead focus this work on the properties of the control law and resulting closed-loop system generated by SMPC.

For the linear SMPC problem, one can establish several important results for the closed-loop system. If we use a global stochastic Lyapunov function as the terminal cost, then one can establish that linear SMPC renders the origin asymptotically stable in probability for multiplicative disturbances[9] and stable in expectation for additive disturbances.[10] Global stochastic Lyapunov functions, however, are not available for open-loop unstable systems with input constraints. Instead, one can use a terminal constraint and a *local* Lyapunov function as the terminal cost to ensure recursive feasibility and stability in expectation for closed-loop systems subject to bounded disturbances.[11−14] We note that for open-loop unstable systems with input constraints the disturbances must be bounded to ensure that a stabilizing control law exists for the system. Goulart and Kerrigan[15] establish that linear SMPC with a quadratic, positive definite stage cost renders the origin input-to-state stable (ISS) for the closed-loop system, regardless of the disturbance's probability distribution. Lorenzen et al.[16] extend these results to include a less restrictive constraint tightening approach and establish that the minimal robust positive invariant set is asymptotically stable with probability one for linear SMPC. Hewing et al.[17] propose a linear SMPC algorithm with indirect feedback to ensure recursive feasibility of the stochastic optimization problem and establish similar stability results for their

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

algorithm. Sehr and Bitmead[18] propose a linear SMPC algorithm with output-feedback and establish a bound on the asymptotic average performance of the closed-loop system.

For *nonlinear* SMPC, the closed-loop properties are more difficult to analyze, and results are necessarily more conservative. In a significant contribution, Chatterjee and Lygeros[19] use a global stochastic Lyapunov function as the terminal cost to establish that the expected value of the optimal cost is bounded along the closed-loop trajectory. Mayne and Falugi[20] then extend this result to systems with input/state constraints and bounded disturbances via a terminal constraint. In McAllister and Rawlings,[21] the authors improve on these stability results and establish that SMPC renders the origin robustly asymptotically stable in expectation (RASiE).

All of these results, however, rely on the pivotal assumption that the stochastic model of uncertainty used in the SMPC optimization problem is equivalent to the stochastic uncertainty of the plant. Thus, these results apply to only an idealized version of SMPC. In practice, the disturbance distribution, typically identified from data, is not exact and is instead subject to its own uncertainty, often called *distributional uncertainty*. What then happens to these guarantees for idealized SMPC if the distribution used in the optimization problem is not the same as the plant? As demonstrated for nominal MPC, there is a possibility that without careful algorithm design, SMPC may provide zero margin of robustness to distributional uncertainty for some nonlinear systems, even if the idealized SMPC performance is satisfactory.

In this work, we introduce sufficient conditions that ensure SMPC provides a nonzero margin of robustness to this distributional uncertainty. We call this property *inherent distributional robustness*. We again use the term *inherent* to emphasize that this distributional robustness is the result of feedback, and we do not incorporate any measure of this distributional uncertainty in the stochastic optimization problem. These sufficient conditions provide a set of design rules for SMPC that ensure the algorithm is not fragile and therefore more suitable for industrial implementation.

The paper outline is as follows. In section 2, we consider linear unconstrained systems and the stochastic linear quadratic regulator to introduce the concept of distributional robustness. In section 3, we then introduce the nonlinear SMPC problem formulation and associated assumptions. In section 4, we define distributional robustness for closed-loop systems and note that SMPC is distributionally robust in this context for sufficiently small errors in the disturbance distribution. In section 5, we use a small example to demonstrate the implications of this distributional robustness for incorrectly modeled and unmodeled disturbances. In section 6, we discussion the significance of this result for scenario-based approximations of the SMPC optimization problem and extensions to nominal MPC. More technical details on these results can be found in McAllister and Rawlings.[22]

2.1. Notation. Let $\mathbb{I}$ and $\mathbb{R}$ denote the integers and reals, respectively. Let superscripts on these sets denote dimension, and let subscripts on these sets denotes restrictions (e.g., $\mathbb{R}^n_{\geq 0}$ for 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. Let f: $X \to Y$ denote a function that maps any $x \in X$ to a point $f(x) \in Y$. A function f: $X \to Y$ is Lipschitz continuous if there exists $L \geq 0$ such that $|f(x_1) - f(x_2)| \leq L|x_1 - x_2|$ for all $x_1, x_2 \in X$. A function f: $X \to Y$ is locally Lipschitz continuous if $f(\cdot)$ is Lipschitz continuous on any compact subset of X. Let $\mathcal{B}(\Omega)$ denote the Borel algebra of some set Ω. Let $tr(A)$ denote the trace of the matrix A. Let $\Pr(x \in S)$ denote the probability that a random variable x takes a value in the set S. Let $\mathbb{E}[x]$ denote expected value of a random variable x and $\mathbb{E}[x|y]$ denote the conditional expected value of x given y. The function α: $\mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is in class $\mathcal{K}$, written $\alpha(\cdot) \in \mathcal{K}$, if $\alpha(\cdot)$ is continuous, strictly increasing, and $\alpha(0) = 0$.

3. STOCHASTIC LINEAR QUADRATIC REGULATOR

We begin with a linear unconstrained control problem to introduce the concept of distributional robustness for closed-loop systems. The system is described by the difference equation

$$x^+ = Ax + Bu + w \tag{1}$$

in which $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, and $w \in \mathbb{W} = \mathbb{R}^q$ is the disturbance. The successor state is denoted x^+ and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ are matrices. We assume that the pair (A, B) is stabilizable. The disturbance w is assumed to be a normally distributed random variable that is independent and identically distributed (i.i.d.) in time with zero mean:

$$w \sim \mathcal{N}(0, \Sigma)$$

in which $\Sigma \succeq 0$ is the covariance of w. Let

$$\mathbf{w}_k := (w(0), w(1), ..., w(k-1))$$

denote the sequence of random variables up to $k \in \mathbb{I}_{\geq 0}$. For any Borel measurable function g: $\mathbb{W}^k \to \mathbb{R}$, let $\mathbb{E}[g(\mathbf{w}_k)]$ denote expected value with respect to this distribution of w.

In this work, however, we do not assume that the probability distribution of the plant is known. We instead assume that the stochastic optimal control problem is formulated with a *model* of this disturbance denoted $\hat{w}$. For this linear case, we assume that $\hat{w}$ is also a normally distributed random variable that is i.i.d. and zero mean:

$$\hat{w} \sim \mathcal{N}(0, \hat{\Sigma})$$

in which $\hat{\Sigma} \succeq 0$ is the covariance of $\hat{w}$. Note that $\hat{\Sigma}$ may not be equal to Σ. The corresponding dynamical model therefore evolves according to

$$x^+ = Ax + Bu + \hat{w}$$

in which $\hat{w} \sim \mathcal{N}(0, \hat{\Sigma})$. Let

$$\hat{\mathbf{w}}_k := (\hat{w}(0), \hat{w}(1), ..., \hat{w}(k-1))$$

denote the sequence of random variables up to $k \in \mathbb{I}_{\geq 0}$. For any Borel measurable function g: $\mathbb{W}^k \to \mathbb{R}$, let $\hat{\mathbb{E}}[g(\hat{\mathbf{w}}_k)]$ denote expected value with respect to this distribution of $\hat{w}$.

For this model of the probability distribution, we define the stochastic linear quadratic regulator (LQR) as follows. We do not refer to this formulation as linear quadratic Gaussian (LQG) control because we do not include the Kalman filter component of the standard LQG problem. Instead, we focus on only the regulation portion of the problem.

Since this formulation considers all possible realizations of the disturbance, we must optimize over a sequence of feedback *policies* $\Pi := (\pi_0, \pi_1, ...)$ instead of a single trajectory of inputs.

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

These feedback policies map the state of the system at time k to a specific input $u(k) = \pi_k(x(k))$ within the stochastic optimization problem. We assume, without loss of generality, that the origin is the target steady state for the controller.

We define the stage cost as

$$l(x, u) := x'Qx + u'Ru$$

in which Q, R are positive definite matrices $(Q, R \succ 0)$. We then define the infinite horizon cost function as

$$V_{\hat{\Sigma}}(x, \Pi) := \lim_{N\to\infty} \frac{1}{N}\hat{\mathbb{E}}\left[\sum_{k=0}^{N-1} l(x(k), \pi_k(x(k)))\right]$$

in which $x(k+1) = Ax(k) + B\pi_k(x(k)) + \hat{w}(k)$ is the stochastic system evolution, $x(0) = x$ is the deterministic initial condition, and Π is the infinite trajectory of control policies. We normalize the infinite horizon cost by the horizon length so that $V_{\hat{\Sigma}}(x, \Pi)$ is finite. The infinite horizon stochastic optimal control problem is then given by

$$V^0_{\hat{\Sigma}}(x) = \min_{\Pi} V_{\hat{\Sigma}}(x, \Pi) \tag{2}$$

for all $x \in \mathbb{R}^n$ in which the optimal trajectory of policies is denoted $\Pi^0_{\hat{\Sigma}} = (\pi^0_{\hat{\Sigma},0}, \pi^0_{\hat{\Sigma},1}, ...)$. We use the subscript $\hat{\Sigma}$ to indicate that the cost function and therefore the optimization problem depend on the probability distribution of $\hat{w}$.

Using dynamic programming, one can show that the solution to this optimization problem is given by the matrix $P \succ 0$ that solves the discrete-time algebraic Riccati equation (DARE):

$$P = A'PA - (A'PB)(R + B'PB)^{-1}(B'PA) + Q \tag{3}$$

such that $A + BK$ is Schur stable with

$$K = -(R + B'PB)^{-1}(B'PA)$$

(see section 3.1 in Bertsekas).[23] Note that we require (A, B) stabilizable for this solution to exist. The optimal cost is given by $V^0_{\hat{\Sigma}}(x) = tr(P\hat{\Sigma})$ and the optimal control law, defined by the first control policy in the optimal solution, is given by $\kappa_{\hat{\Sigma}}(x) = \pi^0_{\hat{\Sigma},0}(x) = Kx$. Note that the matrices P and K do not depend on the variance $\hat{\Sigma}$.

We take a moment to emphasize the significance of this result. The choice of $\hat{\Sigma}$ does not affect the optimal control law derived from this problem formulation. In fact, these are the same the matrix P and control law $\kappa(x) = Kx$ for the *nominal* LQR problem of the same system and cost matrices Q, R. This property is known as *certainty equivalence*.[24,25] Van de Water and Willems[26] provide a discussion of certainty equivalence for stochastic optimal control problems. Moreover, this equivalence means that including stochastic information in the optimization problem produces a different controller only if we consider nonquadratic stage costs, nonlinear systems, or problems with relevant input constraints. If we instead consider systems with nearly linear dynamics, inactive constraints, and quadratic stage costs, then differences between the control laws of nominal MPC and SMPC may be insignificant.

This control law gives the following closed-loop system

$$x^+ = Ax + B\kappa_{\hat{\Sigma}}(x) + w = A_K x + w \tag{4}$$

in which $A_K := A + BK$ and $w \sim \mathcal{N}(0, \Sigma)$. We use $\phi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)$ to denote the state of the closed-loop system in eq 4 at time $k \in \mathbb{I}_{\geq 0}$, given the initial condition x and disturbance sequence $\mathbf{w}_k$:

$$\phi_{\hat{\Sigma}}(k; x, \mathbf{w}_k) := A_K^k x + \sum_{i=0}^{k-1} A_K^{k-1-i} w(i)$$

We leave the subscript $\hat{\Sigma}$ on $\phi_{\hat{\Sigma}}(\cdot)$ to indicate that the control law is designed assuming the disturbance is distributed according to $\hat{\Sigma}$. While irrelevant for the stochastic LQR, this dependence is important for the (nonlinear) SMPC problem in the following sections and is therefore retained for consistency.

What then can we say about the stochastic properties of $\phi_{\hat{\Sigma}}(\cdot)$ subject to the disturbance w with the covariance Σ? We have the following result that establishes a form of distributional robustness for stochastic LQR.

Theorem 1 (Stochastic LQR). For the closed-loop system in (4), there exist $\lambda \in (0, 1)$ and $\rho, \gamma_1, \gamma_2 > 0$ such that

$$\mathbb{E}[|\phi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)|] \leq \lambda^k \rho |x| + \gamma_1\sqrt{tr(\hat{\Sigma})} + \gamma_2\sqrt{|tr(\Sigma - \hat{\Sigma})|} \tag{5}$$

for all $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, $\Sigma \succeq 0$, and $k \in \mathbb{I}_{\geq 0}$.

Proof. We define the function $V(x) := x'Px$, in which P is the solution to the DARE in eq 3 and choose any $\Sigma \succeq 0$. We then have from the fact that $\mathbb{E}[w] = 0$ and the definition of P and K that

$$\begin{aligned}
&\mathbb{E}[V(A_K x + w)|x] \\
&\quad = \mathbb{E}[(A_K x + w)' P(A_K x + w)|x] \\
&\quad = (A_K x)' P(A_K x) + tr(P\Sigma) \\
&\quad = x'Px - x'Qx - (Kx)' R(Kx) + tr(P\Sigma) \\
&\quad = V(x) - l(x, Kx) + tr(P\Sigma)
\end{aligned} \tag{6}$$

Since Q, $P \succ 0$, there exist $c_1, c_2, c_3, c_4 > 0$ such that $c_1|x|^2 \leq V(x) \leq c_2|x|^2$ and

$$\mathbb{E}[V(x^+)|x] \leq V(x) - c_3|x|^2 + c_4 tr(\Sigma) \tag{7}$$

in which $x^+ = A_K x + w$. Thus, $V(x)$ serves as a *stochastic* ISS Lyapunov function. From the upper bound $V(x) \leq c_2|x|^2$ and eq 7, we have

$$\mathbb{E}[V(x^+)] \leq \lambda_1 V(x) + c_4 tr(\Sigma) \tag{8}$$

in which $\lambda_1 = (1 - c_3/c_2) \in (0, 1)$.

Choosing $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, we let $x(k) = \phi(k; x, \mathbf{w}_k)$. For this closed-loop system, we have from eq 8 that

$$\mathbb{E}[V(x(k+1))|x(k)] \leq \lambda_1 V(x(k)) + c_4 tr(\Sigma)$$

We then apply the law of total expectation, or iterated expectation, which states $\mathbb{E}[\mathbb{E}[x(k+1)|x(k)]] = \mathbb{E}[x(k+1)]$, to give

$$\mathbb{E}[V(x(k+1))] \leq \lambda_1 \mathbb{E}[V(x(k))] + c_4 tr(\Sigma)$$

We iterate this inequality from $x(0) = x$ to give

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

$$\mathbb{E}[V(x(k))] \leq \lambda_1^k V(x) + \sum_{i=0}^{k-1} \lambda_1^i c_4 tr(\Sigma)$$
$$\leq \lambda_1^k V(x) + \frac{c_4}{1-\lambda_1} tr(\Sigma)$$

We use the bounds $c_1|x|^2 \leq V(\cdot) \leq c_2|x|^2$ to give

$$\mathbb{E}[|x(k)|^2] \leq \lambda_1^k (c_2/c_1)|x|^2 + \frac{c_4}{c_1(1-\lambda_1)} tr(\Sigma)$$

We then apply Jensen's inequality, which states that for a convex function $\phi(\cdot)$, we have $\phi(\mathbb{E}[x]) \leq \mathbb{E}[\phi(x)]$, and take the square root of both sides of this equation. Since $\sqrt{(\cdot)}$ is subadditive, we have

$$\mathbb{E}[|x(k)|] \leq \lambda_1^{k/2}(c_2/c_1)^{1/2}|x| + \left(\frac{c_4}{c_1(1-\lambda_1)} tr(\Sigma)\right)^{1/2}$$

We define $\lambda := \lambda_1^{1/2}$, $\rho := (c_2/c_1)^{1/2}$, and $\gamma := (c_4/c_1/(1-\lambda))^{1/2}$ to give

$$\mathbb{E}[|\varphi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)|] \leq \lambda^k \rho |x| + \gamma \sqrt{tr(\Sigma)} \tag{9}$$

We also have that

$$\sqrt{tr(\Sigma)} = \sqrt{tr(\hat{\Sigma}) + tr(\Sigma - \hat{\Sigma})}$$
$$\leq \sqrt{tr(\hat{\Sigma})} + \sqrt{|tr(\Sigma - \hat{\Sigma})|} \tag{10}$$

Substitute eq 10 into eq 9 to give eq 5 with $\gamma_1 = \gamma_2 = \gamma$. Since the choice of $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, and $\Sigma \succeq 0$ was arbitrary, eq 5 holds for all $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, and $\Sigma \succeq 0$. □

The right-hand side of the bound in eq 5 contains three terms. The first ensures that the effect of the initial condition vanishes as $k \to \infty$. The second term accounts for the effect of the covariance used to design the control law ($\hat{\Sigma}$). The third term accounts for the discrepancy between the covariance used to design the control law $\hat{\Sigma}$ and the covariance that characterizes the plant Σ. For this linear case, this discrepancy is quantified by the trace of the difference between the covariances for these distributions (i.e., $tr(\Sigma - \hat{\Sigma})$). Note that we use the absolute value of this term and therefore negative values of $tr(\Sigma - \hat{\Sigma})$, indicating that Σ is "smaller" than $\hat{\Sigma}$, may increase the value of $\mathbb{E}[|\varphi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)|]$. As $\Sigma \to \hat{\Sigma}$ we recover the bound for the idealized problem:

$$\mathbb{E}[|\varphi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)|] \leq \lambda^k \rho |x| + \gamma_1 \sqrt{tr(\hat{\Sigma})}$$

If $\Sigma = \hat{\Sigma} = 0$, then we recover exponential stability of the nominal closed-loop system.

We can also establish a performance bound for this closed-loop system in terms of the asymptotic average of the expected value of the stage cost.

Theorem 2 (Stochastic LQR; Performance). For the closed-loop system in eq 4, we have that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\Sigma}}(x(k)))]$$
$$\leq tr(P\hat{\Sigma}) + |tr(P(\Sigma - \hat{\Sigma}))| \tag{11}$$

in which P is the solution to the DARE in eq 3 and $x(k) = \varphi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)$ for all $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, and $\Sigma \succeq 0$.

Proof. Choose $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, $\Sigma \succeq 0$, and let $x(k) = \varphi_{\hat{\Sigma}}(k; x, \mathbf{w}_k)$. Starting with eq 6, we apply the law of total expectation and rearrange to give

$$\mathbb{E}[l(x(k), Kx(k))]$$
$$= \mathbb{E}[V(x(k))] - \mathbb{E}[V(x(k+1))] + tr(P\Sigma)$$

We sum both sides of eq 6 from $k = 0$ to $T - 1$ and divide by T to give

$$\frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[l(x(k), Kx(k))]$$
$$= \frac{V(x) - \mathbb{E}[V(x(T))]}{T} + tr(P\Sigma)$$

We note that for any $x \in \mathbb{R}^n$, the quantity $V(x) - \mathbb{E}[V(x(T))]$ is bounded. If we take the limit as $T \to \infty$, then we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\Sigma}}(x(k)))] = tr(P\Sigma) \tag{12}$$

We also have that

$$tr(P\Sigma) = tr(P\hat{\Sigma}) + tr(P(\Sigma - \hat{\Sigma}))$$
$$\leq tr(P\hat{\Sigma}) + |tr(P(\Sigma - \hat{\Sigma}))|$$

We combine this bound with eq 12 to give eq 11. Since the choice of $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, and $\Sigma \succeq 0$ was arbitrary, eq 11 holds for all $x \in \mathbb{R}^n$, $\hat{\Sigma} \succeq 0$, and $\Sigma \succeq 0$. □

The bound in eq 11 contains two terms on the right-hand side. The first is based on the covariance used in the controller design $\hat{\Sigma}$. The second is again based on the difference between Σ and $\hat{\Sigma}$. As $\hat{\Sigma} \to \Sigma$, we recover the idealized performance bound:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\Sigma}}(x(k)))] \leq tr(P\hat{\Sigma}) \tag{13}$$

In both eqs 5 and 11, we observe a similar characterization of distributional robustness. Arbitrarily small differences between the probability distributions, in terms of $tr(\Sigma - \hat{\Sigma})$, produce similarly small increases in these bounds relative to the idealized case (i.e., $\Sigma = \hat{\Sigma}$). These bounds also ensure that small differences between Σ and $\hat{\Sigma}$ do not produce unstable closed-loop systems.

This result, however, is not unexpected for the LQR problem discussed here. For this linear unconstrained system and stochastic LQR controller, the bounds in eqs 5 and 11 can actually be strengthened to eqs 9 and 12, respectively. For nonlinear systems, however, we cannot establish bounds analogous to eqs 9 or 12. We therefore focus on the weaker versions in eqs 5 and 11 instead to better illustrate the definitions of distributional robustness provided in section 4.

4. STOCHASTIC MODEL PREDICTIVE CONTROL

4.1. Stochastic System(s). We now consider nonlinear, discrete-time dynamical systems

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

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

in which $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the manipulated input, $w \in \mathbb{W} \subseteq \mathbb{R}^q$ is the disturbance, and x^+ is the successor state. We again let $\mathbf{w}_k := (w(0), w(1), ..., w(k-1))$ denote a sequence of disturbances. We consider the following assumption for the disturbances.

Assumption 3. The disturbances $w \in \mathbb{W}$ are random variables that are i.i.d. in time. The set $\mathbb{W}$ is compact and contains the origin.

In contrast to the linear example in the previous section, Assumption 3 is more general in that w can be distributed according to any valid probability distribution but more restrictive in that $\mathbb{W}$ must be compact (i.e., closed and *bounded*). Since we intend to consider systems with bounded manipulated inputs, we require that $\mathbb{W}$ is bounded to ensure that the problem is well-posed. If we instead allowed unbounded disturbances, then we may be unable to reject these disturbances with any bounded input and controller.

Given Assumption 3, each w has an equivalent probability distribution. To represent this distribution we use a Borel probability measure, denoted by the mapping $\mu: \mathcal{B}(\mathbb{W}) \to [0, 1]$, in which $\mathcal{B}(\mathbb{W})$ denotes the collection of all Borel measurable subsets of $\mathbb{W}$. The Borel probability measure has the following interpretation.

$$\mu(S) = Pr(w \in S)$$

in which both sides denote the probability that w takes a value in the set $S \in \mathcal{B}(\mathbb{W})$. Probability measures satisfy the usual Axioms of Probability (i.e., $\mu(S) \geq 0$ for all $S \in \mathcal{B}(\mathbb{W})$, $\mu(\mathbb{W}) = 1$, and μ satisfies the property of *countable additivity*).

We use a probability measure to characterize the distribution of this random variable instead of a continuous probability density function, which is more common in engineering literature, to ensure that these results can address both continuous and discrete probability distributions. To define a discrete probability distribution via a probability measure we use the *Dirac measure* that we denote δ_ω for some point $\omega \in \mathbb{W}$ and is defined as

$$\delta_\omega(S) := \begin{cases} 0, & \omega \notin S \\ 1, & \omega \in S \end{cases}$$

for any $S \in \mathcal{B}(\mathbb{W})$. For $s \in \mathbb{I}_{\geq 1}$ discrete points of equal probability, given by the sequence $\{\omega_i\}_{i=1}^s$, we have the probability measure

$$\mu_d(\cdot) = \frac{1}{s}\sum_{i=1}^{s} \delta_{\omega_i}(\cdot)$$

All of the subsequent results therefore apply for discrete, continuous, and mixed distributions. Most significantly, using probability measures allows us to define a general notion of distance between probability distributions.

Let $\mathcal{M}(\mathbb{W})$ denote the collection of all probability measures for the set $\mathbb{W}$. Moreover, we define expected value as the following Lebesgue integral with respect to this Borel probability measure.

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

Note that $\mathrm{d}\mu(w)$ replaces the probability density function $p(w)\,\mathrm{d}w$ in this definition of integration and expected value. For the sequence $\mathbf{w}_k$, expected value of the Borel measurable function $g: \mathbb{W}^k \to \mathbb{R}$ is defined as

$$\mathbb{E}[g(\mathbf{w}_k)] := \int_{\mathbb{W}^k} g(\mathbf{w}_k)\, \mathrm{d}\mu(w(0))\, \mathrm{d}\mu(w(1)) \\ ...\, \mathrm{d}\mu(w(k-1))$$

We now introduce the *model* of the disturbance distribution used in the SMPC problem. Let $\hat{\mathbb{W}} \subseteq \mathbb{R}^q$ denote the set of disturbances considered in the SMPC problem formulation and let $\hat{\mu}$ denote the probability measure for these disturbances. We therefore have the following stochastic dynamical model for the SMPC problem formulation.

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

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

We assume that $\hat{\mathbb{W}} \subseteq \mathbb{W}$ without loss of generality because we can increase the size of $\mathbb{W}$ to fit $\hat{\mathbb{W}}$ and assign these additional values of w zero probability with μ. We also, without loss of generality, define $\hat{\mu}$ on the domain $\mathcal{B}(\mathbb{W})$ by assigning zero probability to all points in $\mathbb{W}$ that are not in $\hat{\mathbb{W}}$, i.e., $\hat{\mu}: \mathcal{B}(\mathbb{W}) \to [0,1]$ such that $\hat{\mu}(\mathbb{W}/\hat{\mathbb{W}}) = 0$. Specifically, we have that

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

for all measurable functions $g: \mathbb{W} \to \mathbb{R}$. We require that μ and $\hat{\mu}$ are defined on the same domain to facilitate the comparison of these two distributions. The following assumption restates these requirements.

Assumption 4. The random variables $\hat{w} \in \hat{\mathbb{W}}$ are i.i.d. in time, with a known probability measure $\hat{\mu}$ that satisfies $\hat{\mu}(\hat{\mathbb{W}}) = 1$. The set $\hat{\mathbb{W}} \subseteq \mathbb{W}$ is compact and contains the origin.

In the following results, we assume that the set $\hat{\mathbb{W}}$ used in the SMPC algorithm is fixed, but derive bounds that hold for any $\hat{\mu}$ that satisfies Assumption 4. Let $\hat{\mathcal{M}}(\mathbb{W})$ denote the set of all probability measures for the set $\mathbb{W}$ that satisfy Assumption 4 (i.e., $\hat{\mu}(\hat{\mathbb{W}}) = 1$ for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$).

With this framework, we can consider problems in which we incorrectly model the disturbances. For example, we assume that $\hat{\mathbb{W}} = [-2, 2]$, but $\mu([-1, 1]) = 1$, i.e., $w \in [-1, 1]$ with probability one. This framework is also general enough to consider disturbances in the plant that are entirely absent from the disturbance model and SMPC optimization problem. For example, we can consider nominal MPC by defining $\hat{\mathbb{W}} = \{0\}$ and $\hat{\mu}(\{0\}) = 1$ while $\mu(\{0\}) < 1$. Moreover, we allow $\hat{\mathbb{W}}$ to be a finite set (e.g., $\hat{\mathbb{W}} = \{0,1\}$) even if $\mathbb{W}$ is uncountable (e.g., $\mathbb{W} = [-1,1]$). Thus, we can represent incorrectly modeled ($\hat{\mu} \neq \mu$), unmodeled ($\hat{\mathbb{W}} \neq \mathbb{W}$), or out-of-sample ($\hat{\mathbb{W}}$ is finite) disturbances.

For the sequence of i.i.d. random variables

$$\hat{\mathbf{w}} := (\hat{w}(0), \hat{w}(1), ..., \hat{w}(N-1))$$

and $N \in \mathbb{I}_{\geq 1}$, we have the joint distribution measure $\hat{\mu}^N: \mathcal{B}(\mathbb{W}^N) \to [0,1]$ defined as

$$\hat{\mu}^N(S) := \hat{\mu}(S_0)\hat{\mu}(S_1)...\hat{\mu}(S_{N-1})$$

for all $S = (S_0, S_1, ..., S_{N-1}) \in \mathcal{B}(\mathbb{W}^N)$. For any Borel measurable function $g: \mathbb{W}^N \to \mathbb{R}$, we define expected value with respect to $\hat{\mu}$ as

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

for all $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$. We use $\hat{\mathbb{E}}[\cdot]$ to indicate the expected value is evaluated with respect to $\hat{\mu}$ instead of μ.

Note that we do not assume that either w or $\hat{w}$ is zero mean. Thus, we use $\hat{\mathbb{E}}[|\hat{w}|]$ in the definition of distributional robustness instead of $\sqrt{tr(\hat{\Sigma})}$, in which $\hat{\Sigma}$ is the covariance of $\hat{w}$. We also note the following inequality:

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

Thus, for zero mean $\hat{w}$ (i.e., $\hat{\mathbb{E}}[\hat{w}] = 0$), we have that $\hat{\mathbb{E}}[|\hat{w}|] \leq \sqrt{tr(\hat{\Sigma})}$ and can substitute this bound into the subsequent results.

4.2. SMPC Problem Formulation. We now introduce the SMPC problem formulation with a prediction horizon $N \in \mathbb{I}_{\geq 0}$. As with the stochastic LQR problem, we optimize over a sequence of feedback policies $(\pi_0, \pi_1, ..., \pi_{N-1})$ such that the control action at time step in the optimization problem is given by $u(k) = \pi_k(x(k))$. In general, however, we cannot optimize (in real time) over an infinite dimensional object such as a continuous function.

To formulate a tractable optimization problem for this nonlinear dynamical model, we instead define a parameterized control policy $\pi: \mathbb{R}^n \times \mathbb{V} \to \mathbb{R}^m$ 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 that define the control policy. Often, we choose $\pi(x, v) = Kx + v$ in which K is a fixed (feedback gain) matrix and v is vector with the same number of elements as u. Depending on the choice of parametrization, however, $\mathbb{V}$ and $\mathbb{U}$ may not be directly related. For example, we may choose $\pi(x, v) := K(v_1)x + v_2$ with $K: \mathbb{V}_1 \to \mathbb{R}^{m \times n}$ and $v = (v_1, v_2) \in \mathbb{V} = \mathbb{V}_1 \times \mathbb{V}_2$. In this case, $\mathbb{V}$ may have a higher dimension than $\mathbb{U}$.

In any case, we optimize over the trajectory $\mathbf{v} := (v(0), v(1), ..., v(N-1))$ and thereby define a trajectory of control policies. The resulting dynamical model is therefore

$$x^+ = f(x, \pi(x, v), \hat{w}) \quad \hat{w} \in \hat{\mathbb{W}} \tag{14}$$

in which $\hat{w}$ is distributed according to $\hat{\mu}$. Let $\hat{\phi}(k; x, \mathbf{v}, \hat{\mathbf{w}})$ denote the predicted state at time $k \in \mathbb{I}_{0:N}$ based on the dynamical model in eq 14, given the initial condition $x \in \mathbb{R}^n$, the trajectory $\mathbf{v} \in \mathbb{V}^N$, and disturbance trajectory $\hat{\mathbf{w}} \in \hat{\mathbb{W}}^N$.

We allow input constraints $u \in \mathbb{U} \subseteq \mathbb{R}^m$, but do not allow deterministic or probabilistic constraints on the state (except the terminal constraint). Since we do not assume that the disturbance set used in the SMPC problem formulation is exact, the system can encounter $w \notin \hat{\mathbb{W}}$ and is therefore not guaranteed to satisfy these state constraints for the closed-loop system. If we nonetheless attempt to enforce these state constraints in the SMPC problem, then we may encounter states for which the optimization problem is infeasible. Instead, we assume that these desired state constraints are converted to penalty functions that are included in the stage cost. This procedure is often used in nominal MPC formulations for the same reason.[27−29] We do, however, include a terminal constraint $\mathbb{X}_f \subseteq \mathbb{R}^n$ in the problem formulation.

We denote the set of admissible control law parameter trajectories given $x \in \mathbb{R}^n$ as

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

The set of all initial states such that the SMPC optimization problem has a solution is denoted

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

We define the stage cost $l: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, terminal cost $V_f: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, and the function

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

in which $x(k) := \hat{\varphi}(k; x, \mathbf{v}, \hat{\mathbf{w}})$. We then define the SMPC cost function based on the expected value of $J(\cdot)$ with respect to $\hat{\mu}$:

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

As with the stochastic LQR, we note that this cost function depends on the distribution $\hat{\mu}$. The optimization problem for any $x \in \mathcal{X}$ is defined as

$$V^0_{\hat{\mu}}(x) := \min_{\mathbf{v} \in \mathcal{V}(x)} V_{\hat{\mu}}(x, \mathbf{v}) \tag{15}$$

and the optimal solution for a given distribution $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ is denoted by the function $\mathbf{v}^0_{\hat{\mu}}: \mathcal{X} \to \mathbb{V}^N$. If there are multiple solutions to this optimization problem for a given $x \in \mathcal{X}$, then we assume that some Borel measurable selection rule is applied such that $\mathbf{v}^0_{\hat{\mu}}(x)$ defines only a single optimal trajectory for any given $x \in \mathcal{X}$.

Note that the optimal control problem in eq 15 includes nominal MPC as a special case if we choose $\hat{\mathbb{W}} = \{0\}$, $\hat{\mu}(\{0\}) = 1$, $\pi(x, v) = v$, and $\mathbb{V} = \mathbb{U}$. In this case, $V_{\hat{\mu}}(x, \mathbf{v}) = J(x, \mathbf{u}, \mathbf{0})$ and

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

The subsequent results therefore apply to the nominal MPC problem as well as SMPC. We provide a more detailed comparison of SMPC and nominal MPC in section 6.2.

We implement SMPC in a *rolling horizon* framework such that only the first input in this optimal trajectory is injected into the plant. At the next time step, we observe/estimate the new state of the plant and resolve this optimization problem to again determine the input. The *control law* derived from this SMPC formulation is therefore defined as

$$\kappa_{\hat{\mu}}(x) := \pi(x, v^0_{\hat{\mu}}(0; x))$$

in which $v^0_{\hat{\mu}}(0; x)$ is the first element of $\mathbf{v}^0_{\hat{\mu}}(x)$. Note that the optimization problem in eq 15 and the control law $\kappa_{\hat{\mu}}: \mathcal{X} \to \mathbb{U}$ depend on the distribution $\hat{\mu}$.

The resulting closed-loop system is then

$$x^+ = f(x, \kappa_{\hat{\mu}}(x), w) \quad w \in \mathbb{W} \tag{16}$$

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

in which w is distributed according to μ. Let $\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ denote the state at time $k \in \mathbb{I}_{\geq 0}$ based on the dynamical system in eq 16, given the initial state $x \in \mathcal{X}$ and disturbance sequence $\mathbf{w}_k \in \mathbb{W}^k$. Note that the deterministic value of $\varphi_{\hat{\mu}}(\cdot)$ depends on the probability distribution $\hat{\mu}$ because this distribution affects the control law. The expected value of the closed-loop state trajectory, however, is evaluated based on μ. In the subsequent analysis, we discuss quantities such as

$$\mathbb{E}[|\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)|] = \int_{\mathbb{W}^k} |\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)| \, d\mu(w(0)) \, d\mu(w(1)) \dots d\mu(w(k-1))$$

that depend on both $\hat{\mu}$ and μ.

We introduce the following assumptions that characterize the dynamical system and SMPC problem formulation.

Assumption 5 (Lipschitz Continuity of System and Cost). The system $f: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^n$, control parametrization $\pi: \mathbb{R}^n \times \mathbb{V} \to \mathbb{R}^m$, stage cost $l: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and terminal cost $V_f: \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ are locally Lipschitz continuous. Furthermore, $f(0, 0, 0) = 0$, $l(0, 0) = 0$, and $V_f(0) = 0$.

We note that local Lipschitz continuity of these functions is already required if we intend to use nonlinear optimization solvers to solve the SMPC optimization problem. In fact, we often require these functions to be continuously differentiable, a stronger requirement than local Lipschitz continuity, if gradient-based optimization algorithms are used to solve these problems. For example, polynomial and exponential functions are both locally Lipschitz continuous and continuously differentiable.

Assumption 6 (Properties of the Constraint Sets). The set $\mathbb{U}$ and $\mathbb{V}$ are compact and contain the origin. The set $\mathbb{X}_f$ is defined as $\mathbb{X}_f := \{x \in \mathbb{R}^n: V_f(x) \leq \tau\}$ for some $\tau > 0$. The set $\mathcal{X}$ is bounded. The control law parametrization satisfies $\pi(x, v) \in \mathbb{U}$ for all $x \in \mathbb{R}^n$ and $v \in \mathbb{V}$.

The requirement that $\mathbb{U}$ and $\mathbb{V}$ are compact is standard to ensure that a solution to the SMPC optimization problem exists. We also require that $\mathbb{X}_f$ is defined as a sublevel set of $V_f(\cdot)$, which is a similar requirement for the robustness of nominal MPC (See Assumption 2 in Allan et al.).[5] We also require that the feasible set $\mathcal{X}$ is bounded. While this assumption is not often stated, most physical systems in engineering applications admit upper and lower bounds on the state. For example, the mole fraction of a chemical species is between zero and one, and the temperature of a reactor is lower bounded by the temperature of the coolant/inlet stream and upper bounded by the adiabatic limit. We also note that discretization of continuous ordinary differential equations produces a discrete-time system such that $f^{-1}(X) = \{(x, u) \in \mathbb{R}^n \times \mathbb{U}: f(x, u, 0) \in X\}$ is bounded for all bounded $X \subseteq \mathbb{R}^n$. Therefore, the compact sets $\mathbb{U}$ and $\mathbb{X}_f$ ensure that $\mathcal{X}$ is bounded for nominal MPC and SMPC (See Prop. 2.10(d) in Rawlings et al.).[30]

The final requirement of Assumption 6 means that $\pi(x, v) = Kx + v$ may not be a valid control law parametrization. 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 closed value the satisfies $u \in \mathbb{U}$ (i.e., $\text{sat}_{\mathbb{U}}(s) = \arg\min_{u \in \mathbb{U}} |u - s|$). This restriction is minor since we are not permitted to use control inputs outside of $\mathbb{U}$ in the SMPC optimization problem regardless of the parametrization chosen for $\pi(\cdot)$.

Assumption 7 (Terminal Control Law). There exists a locally Lipschitz continuous terminal control law $\kappa_f: \mathbb{X}_f \to \mathbb{U}$ and constant $\tilde{\tau} < \tau$ such that for all $x \in \mathbb{X}_f$,

$$V_f(f(x, \kappa_f(x), 0)) \leq V_f(x) - l(x, \kappa_f(x)) \tag{17}$$

$$f(x, \kappa_f(x), \hat{w}) \in \{x: V_f(x) \leq \tilde{\tau}\} \quad \forall \hat{w} \in \hat{\mathbb{W}} \tag{18}$$

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

The requirement in eq 17 for the terminal control law is a common stability assumption for nominal MPC (see Assumption 2.14(a) in Rawlings et al.).[30] The requirement in eq 18 ensures that the terminal control law drives any $x \in \mathbb{X}_f$ to the *interior* of $\mathbb{X}_f$ for any realization of the disturbance $\hat{w} \in \hat{\mathbb{W}}$. This assumption is therefore stronger than the assumption of robust positive invariance for the terminal set typically used in the analysis of SMPC (see Assumption 5 in Mayne and Falugi).[20] Strengthening this assumption is important to allow some nonzero difference between $\mathbb{W}$ and $\hat{\mathbb{W}}$. We also require exponential bounds on the stage and terminal cost. Note that a quadratic stage cost and quadratic terminal cost satisfy the following assumption.

Assumption 8 (Cost Bounds). There exist $c_1, c_2, a > 0$ such that

$$l(x, u) \geq c_1 |x|^a$$

$$V_f(x) \leq c_2 |x|^a$$

for all $(x, u) \in \mathbb{R}^n \times \mathbb{U}$.

We emphasize that all of these assumptions address the construction of the SMPC problem formulation and do not specify any requirements for the disturbance set $\mathbb{W}$ and distribution μ. Thus, these assumptions guide SMPC controller design. Note that in Assumption 7 we implicitly restrict the size of $\hat{\mathbb{W}}$ by requiring a terminal control law the satisfies eq 18 for any $\hat{w} \in \hat{\mathbb{W}}$. We then allow any $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$ for the chosen set $\hat{\mathbb{W}}$.

5. DISTRIBUTIONAL ROBUSTNESS

With this problem statement, we have two sets $\mathbb{W}$ and $\hat{\mathbb{W}}$ and two probability measures μ and $\hat{\mu}$. The goal of this work is to show that small discrepancies between these sets and measures, produce similarly small degradation of the performance bound that can be derived for the idealized version of SMPC, in which these sets and measures are equal. How then do we define distance between these sets and measures?

The most intuitive concept of a distance or metric is the Euclidean distance $|x - y|$ between two real vectors $x, y \in \mathbb{R}^n$. The concept of distance, however, can be generalized to address functions, sets, and even probability measures. In this section we introduce two such generalizations: The Hausdorff distance between sets and the Wasserstein distance between probability distributions.

To characterize the distance between $\mathbb{W}$ and $\hat{\mathbb{W}}$ we use the Hausdorff distance. The Hausdorff distance between two sets $X, Y \subseteq \mathbb{R}^q$ is defined as

$$d_H(X, Y) := \max\{\sup_{x \in X} |x|_Y, \sup_{y \in Y} |y|_X\}$$

in which $|x|_Y$ denotes the point-to-set distance from the point x to the set Y, i.e., $|x|_Y := \inf_{y \in Y} |x - y|$. Since $\hat{\mathbb{W}} \subseteq \mathbb{W}$ and $\mathbb{W}$ is compact by Assumption 4, the Hausdorff distance between these two sets is given by the simpler expression

$$d_H(\mathbb{W}, \hat{\mathbb{W}}) := \max_{w \in \mathbb{W}} |w|_{\hat{\mathbb{W}}}$$

(i.e., the largest distance between any point in $w \in \mathbb{W}$ and the set $\hat{\mathbb{W}}$).

5.1. Wasserstein Distance. There are several notions of distance available for probability measures, but we find that the Wasserstein distance is the most suitable for this work for two key reasons: (i) the dual formulation of this distance, which is important to establish the results in this section and (ii) the ability to compare continuous and discrete probability distributions, which is important to draw conclusions about sampling-based approximations of the SMPC optimization problem.

While the Wasserstein metric initially found application in the field of optimal transport,[31] more recent applications of this metric can be found in machine learning[32,33] and distributionally robust optimization (DRO).[34] DRO has subsequently been used to develop new Kalman filtering formulations[35] and new nonlinear stochastic optimal control formulations[36] in which distributional uncertainty is included directly in the optimization problem. In contrast to these approaches we do not include the Wasserstein distance or any notion of distributional uncertainty directly in the SMPC optimization problem. We instead use the Wasserstein distance only to quantify the distance between μ and $\hat{\mu}$ and use this distance in the definition of distributional robustness for closed-loop systems. Givens and Shortt[37] provide a further discussion of useful properties of the Wasserstein distance.

To characterize the distance between two probability distributions, one may begin with a simple proposal for this distance in the following form. Consider two random variables $w_1, w_2 \in \mathbb{W} \subseteq \mathbb{R}^q$ with probability measures $\mu_1, \mu_2 \in \mathcal{M}(\mathbb{W})$, respectively. We then define the expected value of the Euclidean distance between these two measures.

$$\mathbb{E}[|w_1 - w_2|] := \int_{\mathbb{W} \times \mathbb{W}} |w_1 - w_2| \, d\mu_1(w_1) \, d\mu_2(w_2)$$

This potential definition of distance is both simple, somewhat intuitive given our natural inclination toward the Euclidean norm, but is not, unfortunately, a proper distance or useful for this work. To see why, note that this quantity assumes that w_1 and w_2 are *independent* random variables. So even if $\mu_1 = \mu_2$, we do not know that $\mathbb{E}[|w_1 - w_2|] = 0$. For example, considering normally distributed scalar random variables $w_1, w_2 \sim \mathcal{N}(0,1)$, we have that $\mathbb{E}[|w_1 - w_2|] = \sqrt{2/\pi}$.

To compare the distributions of these random variables, we want to consider the *best* possible coupling of these distributions instead of treating both variables as independent. The Wasserstein distance (type 1) provides this comparison.

Definition 9 (Wasserstein Distance). The Wasserstein distance between $\mu_1 \in \mathcal{M}(\mathbb{W})$ and $\mu_2 \in \mathcal{M}(\mathbb{W})$ is defined as

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

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

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

for all $\gamma(\cdot) \in \Gamma(\mu_1, \mu_2)$.

In optimal transport, the measure $\gamma(\cdot)$ is called a *transport plan* for moving density or "earth" from a distribution described by μ_1 to another distribution described by μ_2. Hence, the Wasserstein distance is sometimes called the "earth mover's" distance for discrete distributions. We plot an example of $\gamma(\cdot)$ for two 1D normal distributions in Figure 1.

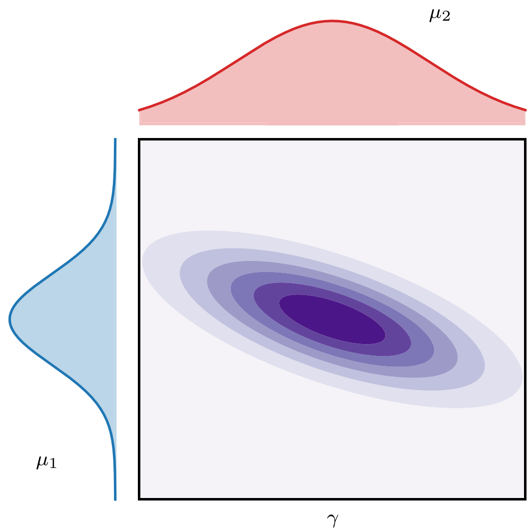


Figure 1. Probability density functions for an example of $\gamma(\cdot) \in \Gamma(\mu_1, \mu_2)$ for two normal distributions.

For 1D distributions, the Wasserstein metric admits an illustrative simplification in terms of cumulative distribution functions. Let $F_1(w_1) := \mu_1((-\infty, w_1])$ and $F_2(w_2) := \mu_1((-\infty, w_2])$ denote the cumulative distribution functions for scalar random variables $w_1, w_2 \in \mathbb{R}$. Then, we have

$$W(\mu_1, \mu_2) = \int_{\mathbb{R}} |F_1(w) - F_2(w)| \, dw$$

In Figure 2, we consider an example with μ_1 defined as a normal distribution and μ_2 defined as a discrete distribution with only four events of equal probability. We plot the probability density function $p_i(w) := \frac{dF_i}{dw}(w)$ and cumulative distribution function $F_i(w)$ for these two distributions. The arrows represent the delta functions in $p_2(w)$. In the bottom plot of Figure 2 we show the absolute difference between these two cumulative distribution functions. The integral of this absolute difference (i.e., the shaded area under this curve) is equal to the Wasserstein distance between μ_1 and μ_2 for this example.

If $w_1 \sim \mathcal{N}(0, \Sigma_1)$ and $w_2 \sim \mathcal{N}(0, \Sigma_2)$, then we also have the following upper bound for the Wasserstein distance (See Prop. 7 in Givens and Shortt).[37]

$$W(\mu_1, \mu_2)^2 \leq tr(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2})^{1/2})$$

For the case of scalar $w_1, w_2 \in \mathbb{R}$ with $w_1 \sim \mathcal{N}(0, \sigma_1)$ and $w_2 \sim \mathcal{N}(0, \sigma_2)$, we can further establish that

$$W(\mu_1, \mu_2) \leq |\sqrt{\sigma_1} - \sqrt{\sigma_2}|$$

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

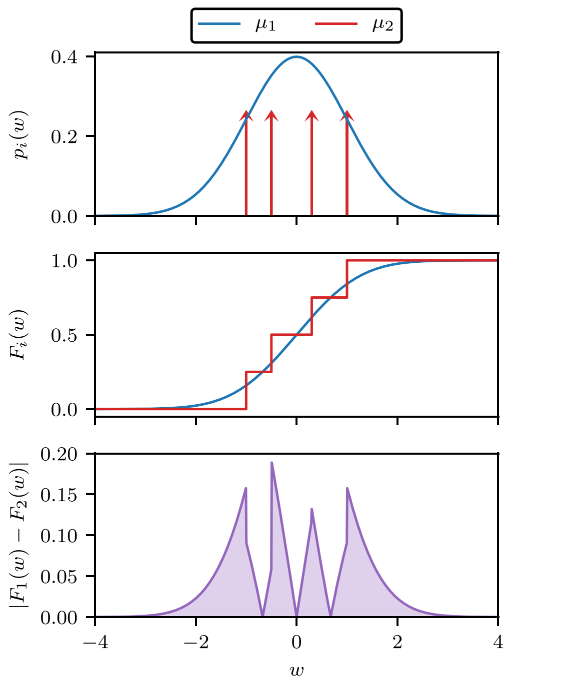


Figure 2. Probability density $p_i(w)$ and cumulative density $F_i(w)$ for two 1D probability measures μ_1, μ_2. The bottom plot is the absolute difference between these two cumulative distribution functions.

The notion of distance between normal distributions used in eqs 5 and 11 is therefore similar to that of the Wasserstein distance between normal distributions.

5.2. Main Results. Since the SMPC control law is based on an optimization problem, we must first ensure that the optimization problem remains feasible along the closed-loop trajectory. Specifically, we want to show that the feasible set for the optimization problem $\mathcal{X}$ is *robustly positive invariant* for the closed-loop system.

Definition 10 (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 $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$.

Robust positive invariance of the feasible set $\mathcal{X}$ ensures that for any state $x \in \mathcal{X}$, probability measure $\hat{\mu} \in \hat{\mathcal{M}}(\mathbb{W})$, and disturbance $w \in \mathbb{W}$, the successor state remains in $\mathcal{X}$. The optimization problem therefore remains feasible. We also use the notation $\alpha(\cdot) \in \mathcal{K}$ to indicate that the function α: $\mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is continuous, strictly increasing, and $\alpha(0) = 0$.

We now use the Wasserstein distance to define distributional robustness for closed-loop nonlinear systems.

Definition 11 (Distributionally Robust Exponential Stability in Expectation). The origin is distributionally robustly exponentially stable in expectation (DRESiE) for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ in the RPI set $\mathcal{X}$ if there exist $\lambda \in (0, 1)$, $\rho > 0$, and $\gamma_1(\cdot), \gamma_2(\cdot) \in \mathcal{K}$ such that

$$\mathbb{E}[|\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)|] \leq \lambda^k \rho |x| + \gamma_1(\hat{\mathbb{E}}[|\hat{w}|]) + \gamma_2(W(\mu, \hat{\mu})) \tag{19}$$

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

We note that eq 19 is similar to eq 5. The first term on the right-hand side of eq 19 ensures that the effect of the initial condition vanishes as $k \to \infty$. In the second term, we use $\hat{\mathbb{E}}[|\hat{w}|]$ instead of $tr(\hat{\Sigma})$ to characterize the probability measure $\hat{\mu}$ used to design the control law. If $\hat{\mathbb{E}}[\hat{w}] = 0$, then we can replace $\hat{\mathbb{E}}[|\hat{w}|]$ with $tr(\hat{\Sigma})^{1/2}$ in eq 19. In the third term, we use $W(\mu, \hat{\mu})$ instead of $tr(\Sigma - \hat{\Sigma})$ to quantify the difference between the probability distribution used to design the control law $\hat{\mu}$ and the probability distribution that characterizes the plant μ. If $\mu = \hat{\mu}$, then $\gamma_2(W(\mu, \hat{\mu})) = 0$, and we recover the same bound derived for idealized SMPC.

$$\mathbb{E}[|\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)|] \leq \lambda^k \rho |x| + \gamma_1(\mathbb{E}[|w|]) \tag{20}$$

We can now establish a main result of this work.

Theorem 12 (Distributional Robustness of SMPC). Let Assumptions 3–8 hold. Then there exists $\delta > 0$ such that for any $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$ and the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$, we have that $\mathcal{X}$ is RPI. Furthermore, the origin is DRESiE for this system in the set $\mathcal{X}$.

A proof of Theorem 12 for asymptotic, instead of exponential, decay of the initial state is available in McAllister and Rawlings.[22] Since this proof is rather technical, we do not delve into the details here. We instead focus on the practical implications of this result in the following discussion.

We note two key properties afforded by Theorem 12 for the distributional robustness of SMPC. First, Theorem 12 ensures that for small differences between $\mathbb{W}$ and $\hat{\mathbb{W}}$, the set $\mathcal{X}$ is RPI and the SMPC optimization problem remains feasible. Second, Theorem 12 ensures that the origin is DRESiE for the closed-loop system and the closed-loop system therefore satisfies eq 19. This bound ensures that small differences between $\hat{\mu}$ and μ produce similarly small deviations in upper bound derived for idealized SMPC.

In addition to set point tracking problems, an important class of applications for SMPC are *economic* problems in which the stage cost is chosen to directly represent an economic metric for the process (e.g., production cost). In certain cases, this cost may be positive definite with respect to the origin (target steady state) and satisfy the bound in Assumption 8. Often, however, this requirement is too restrictive for the class of economic cost functions of industrial interest. Thus, in economic applications of MPC (i.e., economic MPC), we often drop Assumption 8 and instead analyze the performance of closed-loop system in terms of the stage cost. Without Assumption 8, we typically obtain a weaker, but still useful, result for economic applications of MPC. For economic applications of SMPC, we have the following instructive result.

Theorem 13 (Distributional Robustness of Economic SMPC). Let Assumptions 3–7 hold. Then there exists $\delta > 0$ such that for any $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta$ and the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$ we have that $\mathcal{X}$ is RPI. Furthermore, there exist $L_1, L_2 > 0$ such that the closed-loop trajectory satisfies

$$\begin{aligned} \lim_{T \to \infty} \sup \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\mu}}(x(k)))] \\ \leq L_1 \hat{\mathbb{E}}[|\hat{w}|] + L_2 W(\mu, \hat{\mu}) \end{aligned} \tag{21}$$

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

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

A proof of Theorem 13 is available in McAllister and Rawlings.[22] The bound in eq 21 ensures that as $T \to \infty$, the time-averaged expected value of the stage cost is upper bounded by a constant proportional to $\hat{\mathbb{E}}[|\hat{w}|]$ and $W(\mu, \hat{\mu})$. If $\mu = \hat{\mu}$, then the bound in eq 21 reduces to a standard result for idealized SMPC, first derived for nonlinear systems by Chatterjee and Lygeros.[19] Specifically, we have the following bound for idealized economic SMPC.

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\mu}}(x(k)))] \le L_1\mathbb{E}[|w|] \tag{22}$$

Thus, economic applications of SMPC also provide some margin of inherent distributional robustness in terms of a suitable performance metric.

6. EXAMPLE

To demonstrate the implications of these theoretical results, we consider a small CSTR example. An irreversible, first-order reaction $A \to B$ occurs in the liquid phase. The reaction is exothermic and the reactor temperature is controller with external cooling. We assume the inlet/outlet flow rate and volume are constant. Mass and energy balances lead to the following differential equations.

$$\frac{dc}{dt} = \frac{c_f - c}{\theta} - k_0 \exp\left(\frac{-E}{T}\right) c$$

$$\frac{dT}{dt} = \frac{T_f - T}{\theta} - \frac{\Delta H}{\rho C_p} k_0 \exp\left(\frac{-E}{T}\right) c + \frac{U}{r\rho C_p}(T_c - T)$$

The state variables are the concentration of species A in the reactor c and the temperature of the reactor T. The input is the temperature of the coolant T_c. We choose the steady-state target $c_s = 0.878$ kmol/m^3, $T_s = 324.5$ K, $T_{c,s} = 300$ K and define the state and input as

$$x = \begin{bmatrix} (c - c_s)/c_s \\ (T - T_s)/T_s \end{bmatrix} \qquad u = [(T_c - T_{c,s})/T_{c,s}]$$

The input is subject to the constraint $u \in \mathbb{U} := [-0.1, 0.1]$. We discretize these differential equations via Runge–Kutta (fourth-order) with a time step of $\Delta = 0.5$ min. We also include an additive disturbance w at each time step to give the following nonlinear difference equation

$$x^+ = f(x, u, w) = \hat{f}(x, u) + w$$

For the SMPC problem, we assume that there is a disturbance in the temperature of the reactor at each time step. We model this disturbance with a discrete distribution $\hat{\mathbb{W}} := \{0\} \times \{-0.006, 0, 0.006\}$ (i.e., $\hat{w}_1 = 0$ and $\hat{w}_2 \in \{-0.006, 0, 0.006\}$ (approximately ± 2 K)).The probability measure for this disturbance is defined such that $\hat{\mu}(\{(0, -0.006)\}) = \hat{\mu}(\{(0, 0.006)\}) = 0.3$ and $\hat{\mu}(\{(0, 0)\}) = 0.4$. We choose this simple disturbance representation to ensure that these results are easy to interpret and reproduce.

We choose the quadratic stage cost

$$l(x, u) = x'Qx + u'Ru$$

with $Q = 0.1I$ and $R = 1$. We then design the terminal control law and cost by linearizing the nominal system at the target steady state and computing the LQR gain K and cost matrix P for this unconstrained linear system. We use inflated stage cost matrices $Q_{\mathrm{LQR}} = 10Q$ and $R_{\mathrm{LQR}} = R$ for this LQR calculation. We then define the terminal cost as $V_f(x) = x'Px$ and the terminal constraint as $\mathbb{X}_f = \{x \in \mathbb{R}^n\colon V_f(x) \le \tau\}$ in which $\tau = 0.001$. We define the control law parametrization as $\pi(x, v) = \mathrm{sat}_{\mathbb{U}}(Kx + v)$ and choose $\mathbb{V} = [-1, 1]$ to ensure sufficient flexibility in selecting this parameter. We choose $N = 3$. With this formulation, we satisfy Assumptions 5, 6, and 8.

We now verify Assumption 7 by sampling the terminal region and computing the successor state for each of these samples subject to the terminal control law $\kappa_f(x) = Kx$ and all possible realizations of $\hat{w}$. That is, we approximate the set

$$\hat{\mathbb{X}}_f^+ := \{f(x, Kx, \hat{w})\colon x \in \mathbb{X}_f,\ \hat{w} \in \hat{\mathbb{W}}\}$$

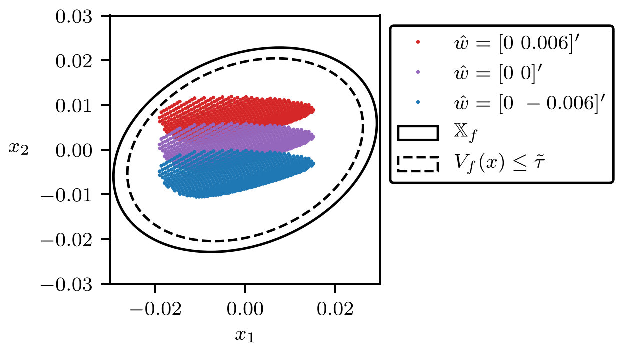


Figure 3. Terminal constraint $\mathbb{X}_f$ and a sampling-based approximation of $\hat{\mathbb{X}}_f^+$ to verify that Assumption 7 holds for the SMPC problem formulation.

We plot these points in Figure 3 along with the set $\mathbb{X}_f$. We also show in Figure 3 that all of these points are within the set defined by $\{x\colon V_f(x) \le \tilde{\tau}\}$ in which $\tilde{\tau} = 0.8\tau$. We use this same approach to confirm that eq 17 also holds. We therefore satisfy Assumption 7 with this problem formulation, within the sampling error.

We can then formulate and solve the optimization problem in eq 15 to determine the input at each time step. Specifically, the SMPC optimization problem can be written as a nonlinear program as follows. Since $\mathbb{W}$ is finite with only three possible values for the disturbance, we consider each of the 3^N disturbance trajectories $\mathbf{w}^{(s)} = (w^{(s)}(0), ..., w^{(s)}(N-1))$ for all $s \in \{1, ..., 3^N\}$ and their associated probability $p^{(s)}$ determined by the distribution $\hat{\mu}$. Each disturbance trajectory has a corresponding predicted state and input trajectory, denoted $x^{(s)}(k)$ and $u^{(s)}(k)$. With these variables, the SMPC optimization problem for the initial state x_0 is then written as follows.

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

$$\begin{aligned}
\min_{\mathbf{v}} \quad & \sum_{s=1}^{3^N}\left(p^{(s)}\sum_{k=0}^{N-1} l(x^{(s)}(k), u^{(s)}(k)) + V_f(x^{(s)}(N))\right) \\
\text{s.t.} \quad & x^{(s)}(k+1) = f(x^{(s)}(k), u^{(s)}(k), w^s(k)) \\
& \quad \forall\, s \in \mathbb{I}_{1:3^N},\ k \in \mathbb{I}_{0:N-1} \\
& u^{(s)}(k) = \mathrm{sat}_{\mathbb{U}}(Kx^{(s)}(k) + v(k)) \quad \forall\, s \in \mathbb{I}_{1:3^N},\ k \in \mathbb{I}_{0:N-1} \\
& x^{(s)}(0) = x_0,\ x^{(s)}(N) \in \mathbb{X}_f \quad \forall\, s \in \mathbb{I}_{1:3^N}
\end{aligned} \tag{23}$$

We can then approximate $\mathrm{sat}_{\mathbb{U}}(Kx^{(s)}(k) + v(k))$ via sigmoid functions or relax this constraint to $u^{(s)}(k) = Kx^{(s)}(k) + v(k)$ if the input constraints are not active for the optimal solution. We use the latter approach for the following simulations.

Realize that constructing an SMPC formulation that satisfies these assumptions is not significantly more difficult than constructing a nominal MPC formulation that satisfies the assumptions in Yu et al.[4] or Allan et al.[5] For quadratic stage costs, we typically construct a terminal cost for nominal MPC via the same procedure used in this section. The computational effort required to *solve* the SMPC optimization problem, however, may be considerably larger than the computational effort required to solve the nominal MPC problem, particularly for nonlinear systems with long horizons. In the following subsections, we investigate the robustness of SMPC for this example by simulating the closed-loop trajectory subject to various distributions for both w_1 and w_2.

6.1. Ideal SMPC. We first consider the case of idealized SMPC, in which $\mu = \hat{\mu}$ and $\mathbb{W} = \hat{\mathbb{W}}$. We simulated the closed-loop system starting from $c = 0.92$ and $T = 340$ subject to 30 realizations of the disturbance trajectory and plot the closed-loop state and input trajectories for these simulations in Figure 4. We use this same initial condition for all subsequent simulations as well.

For these trajectories, we plot the norm of the closed-loop state $|x(k)|$ in which $x(k) = \varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ in Figure 5. We also plot the sample average of these trajectories, denoted

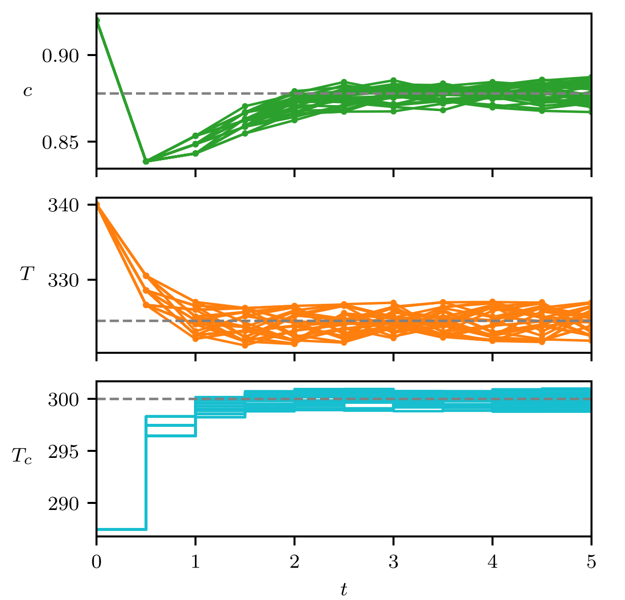


Figure 4. Closed-loop trajectories for idealized SMPC with 30 realizations of the disturbance trajectory.

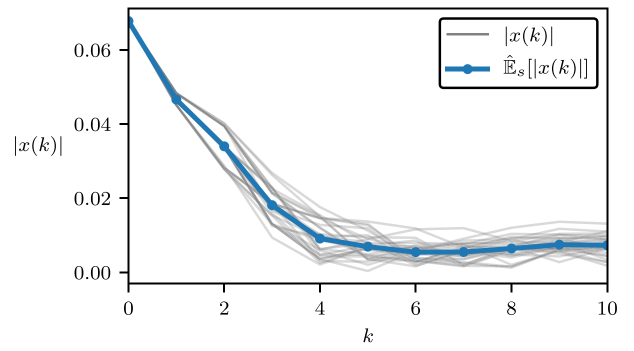


Figure 5. Norm of the closed-loop state trajectories $|x(k)|$ and the sample average of these trajectories $\hat{\mathbb{E}}_s[|x(k)|]$.

$\hat{\mathbb{E}}_s[|x(k)|]$, in Figure 5. We observe an initial, exponential decay in $\hat{\mathbb{E}}_s[|x(k)|]$, followed by a persistent deviation from the origin due to the disturbance. Note this behavior is consistent with eq 20 and therefore Theorem 12.

6.2. Incorrectly Modeled Disturbances. Next, we consider that the disturbance support is correct (i.e., $\mathbb{W} = \hat{\mathbb{W}}$), but the probability distribution is incorrect (i.e., $\mu \neq \hat{\mu}$). Let $\mu(\{(0, -0.006)\}) = \mu(\{(0, 0.006)\}) = \varepsilon_2/2$ and $\mu(\{(0, 0)\}) = 1 - \varepsilon_2$ for some $\varepsilon_2 \in [0, 1]$. For this case, we have that $W(\mu, \hat{\mu}) = 0.006|\varepsilon_2 - 0.6|$. For multiple values of ε_2 we simulate the closed-loop trajectory for 30 realizations of the disturbance according to this distribution. We plot the sample average of the norm of the closed-loop state, denoted $\mathbb{E}_s[|x(k)|]$, in Figure 6 for these different distributions.

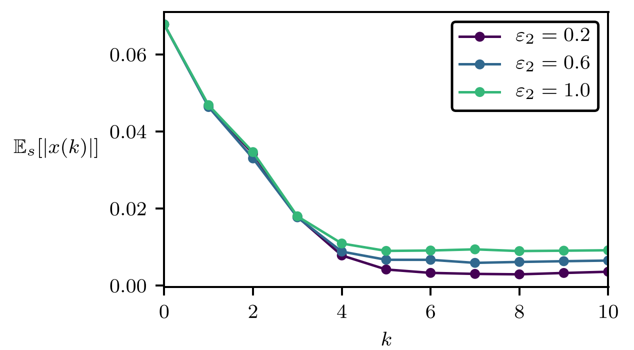


Figure 6. Sample average of the norm of the closed-loop state, denoted $\mathbb{E}_s[|x(k)|]$, for different values of ε_2 and therefore different μ.

We again observe a similar exponential decay for all of these trajectories in the first few time steps. After this initial decay ($k \geq 6$), these trajectories appear to remain within some region of the origin, based on the value of ε_2. Note that $\mathbb{E}_s[|x(k)|]$ for $\varepsilon_2 = 1$ is larger than for ideal SMPC ($\varepsilon_2 = 0.6$) for $k \geq 6$, as suggested by eq 19. We also observe that $\mathbb{E}_s[|x(k)|]$ is lower for $\varepsilon_2 = 0.2$ than ideal SMPC for $k \geq 6$. This result, however, does not conflict with eq 19. The bound in eq 19 must account for any possible probability distribution and is therefore conservative. In either case, SMPC is robust to incorrectly modeled distributions.

6.3. Out-of-Sample Disturbances. We now consider a *continuous* distribution for the temperature disturbance in the plant and therefore $\mathbb{W} \neq \hat{\mathbb{W}}$, $\hat{\mu} \neq \mu$. Specifically, we assume that w_2 is distributed according to a truncated normal

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

distribution with zero mean, a variance of 0.006^2, and truncated such that $w_2 \in [-0.012, 0.012]$. Disturbances drawn from this truncated normal distribution are *outside* of the set of samples used in the SMPC optimization problem, hence the term *out-of-sample* disturbances. We simulate the closed-loop trajectory for 30 realizations of the disturbance drawn from this truncated normal distribution. In Figure 7, we compare the sample average of the norm of the closed-loop trajectory for this out-of-sample distribution to the idealized SMPC result.

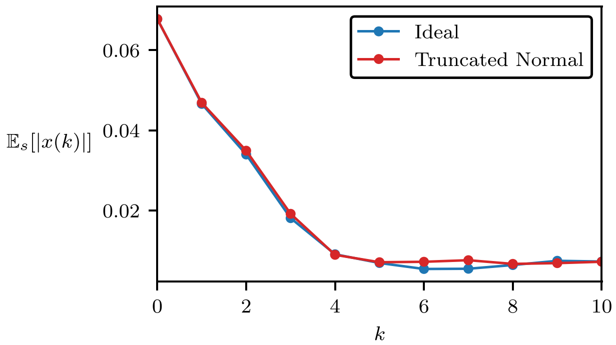


Figure 7. Sample average of the norm of the closed-loop state, $\mathbb{E}_s[|x(k)|]$, for ideal SMPC and SMPC subject to a truncated normal distribution for w_2.

Theorem 12 also addresses these out-of-sample disturbances. Note that the SMPC optimization problem remained feasible along all of these closed-loop trajectories, even though we considered out-of-sample disturbances that were up to twice the size of the disturbances considered in the SMPC optimization problem, i.e., $d_H(\mathbb{W}, \hat{\mathbb{W}}) = 0.006$. Moreover, the trajectories for ideal SMPC and SMPC subject to the truncated normal distribution are very similar. These results are consistent with Theorem 12 and are indicative of the fact that SMPC is robust to out-of-sample disturbances.

6.4. Unmodeled Disturbances. Instead of incorrect distributions for w_2, we now consider the possibility of nonzero w_1, i.e., an unmodeled disturbance. We assume that w is described by a discrete distribution in which

$$\mathbb{W} := \{-0.01, 0, 0.01\} \times \{-0.006, 0, 0.006\}$$

(i.e., $w_1 \in \{-0.01, 0, 0.01\}$ and $w_2 \in \{-0.006, 0, 0.006\}$). We therefore have that $d_H(\mathbb{W}, \hat{\mathbb{W}}) = 0.01$. The probability distribution is given by $\mu(\{(w_1, w_2)\}) = \mu_1(\{w_1\})\mu_2(\{w_2\})$ in which $\mu_1(\{-0.01\}) = \mu_1(\{0.01\}) = \varepsilon_1/2$, $\mu_1(\{0\}) = 1 - \varepsilon_1$, $\mu_2(\{-0.006\}) = \mu_2(\{0.006\}) = \varepsilon_2/2$, and $\mu_2(\{0\}) = 1 - \varepsilon_2$. We assume that the distribution for w_2 is chosen correctly in the SMPC optimization problem and therefore $\varepsilon_2 = 0.6$. We consider several values of ε_1 to investigate the robustness of SMPC to this unmodeled disturbance. Note that $W(\mu, \hat{\mu}) = 0.01\varepsilon_1$ for this example.

We simulate the closed-loop trajectory for 30 realizations of the disturbance trajectory for several values of ε_1. Note that the SMPC optimization problem remained feasible for all of these simulations despite this unmodeled disturbance. We then evaluate the sample average norm of the closed-loop state trajectories for each ε_1 and plot these trajectories in Figure 8. As ε_1 and therefore $W(\mu, \hat{w})$ increase, $\mathbb{E}_s[|x(k)|]$ increases as well for $k \geq 6$.

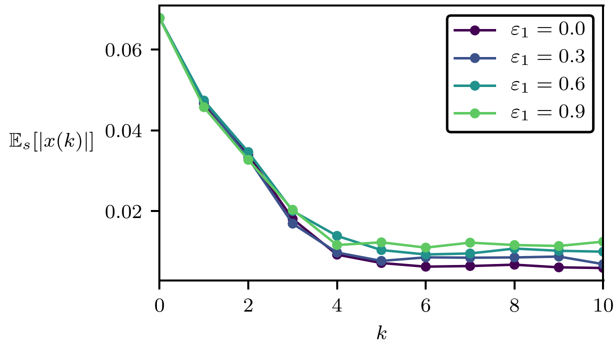


Figure 8. Sample average of the norm of the closed-loop state, $\mathbb{E}_s[|x(k)|]$, for multiple values of ε_1.

7. DISCUSSION AND CONCLUSIONS

7.1. Scenario-Based Approximations. Scenario optimization methods or the *sample average approximation* (SAA) are often used to approximate and solve the SMPC optimization problem for nonlinear systems.[7] In these approximations, a finite set of samples is chosen from a given disturbance distribution. The expected value of the cost function is approximated by the average cost of these scenarios and the constraints in the optimization problem are required to hold for all of these scenarios. Thus, the stochastic optimization problem can be solved with deterministic methods (e.g., as a nonlinear program). The quality of this approximation is often characterized by the distance between the optimal cost/argument for the approximate problem and the optimal cost/argument for the original stochastic optimization problem. Unfortunately, the quality of this approximation is meaningless if near exact approximations still produce poor controllers. The result in Theorem 12, however, establishes that the distributional robustness of SMPC also extends to these scenario-based approximations.

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 finite set $\hat{\mathbb{W}}_s = \{\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_s\}$ and empirical (discrete) distribution

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

The set $\hat{\mathbb{W}}_s$ and distribution $\hat{\mu}_s$ satisfy Assumption 4. If Assumption 7 holds for $\hat{\mathbb{W}}$, then Assumption 7 also holds for $\hat{\mathbb{W}}_s \subseteq \hat{\mathbb{W}}$. We can therefore substitute $\hat{\mathbb{W}}_s$ and $\hat{\mu}_s$ in place of $\hat{\mu}$ and $\hat{\mathbb{W}}$ for all algorithms and results in this work including Theorems 12 and 13. This algorithm is closely related to that of multistage MPC.[38,39]

We can thereby draw two important conclusions for these scenario-based approximations. Both of these conclusions are the result of the triangle inequality, which applies to even generalized notions of distance such as the Hausdorff and Wasserstein distance.

First, we have the following conclusion for the robust recursive feasibility of the scenario-based approximation of the SMPC problem. If $\hat{\mathbb{W}}$ is close to ($d_H(\mathbb{W}, \hat{\mathbb{W}}) \leq \delta/2$) and the sampling of $\hat{\mathbb{W}}$ is sufficiently dense ($d_H(\hat{\mathbb{W}}, \hat{\mathbb{W}}_s) \leq \delta/2$), then we have that

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190

$$d_H(\mathbb{W}, \hat{\mathbb{W}}_s) \le d_H(\mathbb{W}, \hat{\mathbb{W}}) + d_H(\hat{\mathbb{W}}, \hat{\mathbb{W}}_s) \le \delta$$

and therefore $\mathcal{X}$ is RPI. In other words, scenario optimization can ensure robust recursive feasibility of SMPC, if a sufficient density of samples are used and the disturbance model is sufficiently accurate.

Second, the performance, in terms of norm of the state or stage cost, of the closed-loop system is bounded by the distance between $\hat{\mu}_s$ and μ. We again use the triangle inequality to give

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

We can further establish that random sampling of the probability distribution ensures that $\hat{\mu}_s \to \hat{\mu}$ as $s \to \infty$,[40] and therefore

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

with probability one as $s \to \infty$ (See Theorem 6.9 in Villani).[31] The bounds in eqs 19 and 21 therefore converge to their values for the original stochastic optimization problem as the number of samples increases.

7.2. Comparison to Nominal MPC.

Perhaps the most interesting feature of the results in Theorems 12 and 13, is that they unify the notions of stochastic robustness across nominal and stochastic MPC. With the framework introduced in this work, we can treat nominal MPC as a special case of SMPC. By choosing $\hat{\mathbb{W}} = \{0\}$ and $\hat{\mu}(\{0\}) = 1$, the SMPC problem in eq 15 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 parametrization is sometimes used in nominal MPC to ensure the optimization problem is well conditioned for open-loop unstable systems.[41,42]

If we also choose $\pi(x, v) = v$ and $\mathbb{V} = \mathbb{U}$, then we have that

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

and the optimization problem reduces to nominal MPC exactly.

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

With this choice of $\pi(x, v) = v$ and $\mathbb{V} = \mathbb{U}$, Assumptions 5, 6, and 8 are equivalent to the assumptions used in Allan et al.[5] to establish the inherent robustness of nominal MPC. In Assumption 7, the nominal cost decrease condition in eq 17 is also required in Allan et al.[5] and is standard in MPC literature. The requirement in eq 18 reduces to

$$f(x, \kappa_f(x), 0) \in \{x \in \mathbb{R}^n : V_f(x) \le \tilde{\tau}\} \tag{24}$$

for some $\tilde{\tau} < \tau$. If Assumption 8 also holds, then eq 18 is implied by eq 17 and is therefore not an additional requirement for the terminal cost, constraint, and control law. In summary, the assumptions used in this work reduce to their nominal MPC counterparts for this choice of $\hat{\mathbb{W}} = \{0\}$, $\hat{\mu}(\{0\}) = 1$, $\pi(x, v) = v$, and $\mathbb{V} = \mathbb{U}$.

For $\hat{\mu}(\{0\}) = 1$, we have that $\hat{\mathbb{E}}[|\hat{w}|] = 0$ and

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

Moreover, $d_H(\mathbb{W}, \hat{\mathbb{W}}) = \max_{w \in \mathbb{W}} |w|$ for $\hat{\mathbb{W}} = \{0\}$. We therefore have the following corollary of Theorems 12 and 13.

Corollary 14 (Nominal MPC). Let Assumptions 3−7 hold with $\hat{\mathbb{W}} = \{0\}$ and $\hat{\mu}(\{0\}) = 1$. Then there exists $\delta > 0$ such that for any set $\mathbb{W} \subseteq \mathbb{R}^q$ satisfying $\max_{w \in \mathbb{W}} |w| \le \delta$, the set $\mathcal{X}$ is RPI for the system $x^+ = f(x, \kappa_{\hat{\mu}}(x), w)$, $w \in \mathbb{W}$. Furthermore, there exists $L_2 \ge 0$ such that the closed-loop trajectory satisfies

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{k=0}^{T-1} \mathbb{E}[l(x(k), \kappa_{\hat{\mu}}(x(k)))] \le L_2 \mathbb{E}[|w|] \tag{25}$$

in which $x(k) = \varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)$ for all $x \in \mathcal{X}$ and $\mu \in \mathcal{M}(\mathbb{W})$. If Assumption 8 also holds, then there exist $\lambda \in (0, 1)$, $\rho > 0$, and $\gamma_2(\cdot) \in \mathcal{K}$ such that

$$\mathbb{E}[|\varphi_{\hat{\mu}}(k; x, \mathbf{w}_k)|] \le \lambda^k \rho |x| + \gamma_2(\mathbb{E}[|w|]) \tag{26}$$

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

Thus, Corollary 14 ensures that nominal MPC confers some margin of inherent *stochastic* robustness to sufficiently small disturbances. Note the similarities between the bounds in eqs 25, 26 and eqs 20, 22. We emphasize, however, that the function $\gamma_2(\cdot)$ and constant L_2 in eqs 25, 26, respectively, are *not* the same as the function $\gamma_1(\cdot)$ and constant L_1 in eqs 20, 22. This observations suggests that the performance of SMPC and nominal MPC may differ quantitatively, but their qualitative behavior is likely similar for an otherwise equivalent problem. In general, the functions $\gamma_1(\cdot)$, $\gamma_2(\cdot)$ and constants L_1, L_2, δ are often too conservative to provide useful information about the performance of these systems. Thus, efforts to compare the performance of nominal MPC and SMPC via these bounds are likely misplaced. Simulation studies, such as Kumar et al.,[43] remain the best means to evaluate the potential benefits of SMPC relative to nominal MPC.

7.3. Potential Extensions.

7.3.1. State and Probabilistic Constraints. Since we are considering systems in which $\mathbb{W} \neq \hat{\mathbb{W}}$, we cannot guarantee that state constraints (probabilistic or deterministic) are satisfied for all realizations of the disturbance $w \in \mathbb{W}$. Instead, we propose "softening" these constraints by including them as penalty functions in the SMPC stage cost. We note that for probabilistic constraints of the form

$$Pr(f(x, u, \hat{w}) \in \tilde{\mathbb{X}}) \ge 1 - \varepsilon \tag{27}$$

we can rewrite eq 27 as

$$(x, u) \in \mathbb{Z} := \{(x, u) : g(x, u) \le 0\}$$

in which

$$g(x, u) = 1 - \varepsilon - \int_{\hat{\mathbb{W}}} I_{\tilde{\mathbb{X}}}(f(x, u, \hat{w}))\, d\hat{\mu}(\hat{w})$$

and $I_{\tilde{\mathbb{X}}}$ is the indicator function for the set $\tilde{\mathbb{X}}$ (i.e., $I_{\tilde{\mathbb{X}}}(x) = 1$ if $x \in \tilde{\mathbb{X}}$ and zero otherwise). Note that $\mathbb{Z}$ is a closed set if $\tilde{\mathbb{X}}$ is closed and $f(\cdot)$ is continuous (See Lemma 1 in McAllister and Rawlings).[21] In practice, the set $\mathbb{Z}$ and/or function $g(\cdot)$ is often approximated (offline). A potential approach to soften this constraint is to include a penalty on the violation of this constraint in the stage cost. Closed-loop properties for such a reformulation, such as guarantees on closed-loop constraint satisfaction under distributional uncertainty, have not been investigated.

7.3.2. Suboptimal SMPC Algorithms. Note that for nonlinear SMPC formulations, we require the solution to a nonconvex (stochastic) optimization problem (e.g., eq 23). For problems of industrially relevant size, however, we are typically unable to determine a *global* optimum for this problem, as required by the analysis in this work, within the computation time required to implement this controller in real-time. Thus, suboptimal SMPC algorithms that guarantee some nonzero margin of distributional robustness without requiring an exact solution to the SMPC optimization problem in eq 15 are of significant practical interest.

7.3.3. State Estimation. In this work, we assumed perfect knowledge of the current state of the system, but in practice this state is estimated via methods such as Kalman filtering (standard, extended, or unscented varieties) or moving horizon estimation (MHE). Some of these state estimation methods, such as Kalman filtering, provide a description of the uncertainty in the state estimate. Including this information in the SMPC problem formulation may therefore provide some benefit to the closed-loop performance. Sehr and Bitmead,[18] for example, present a framework for combining the state estimation and regulation part of the stochastic optimal control problem via stochastic dynamic programming (SDP). Their results, however, are limited by the considerable computational requirements required to solve the SDP problem for systems of large state dimension. We can likely extend the problem formulation in (eq 15) to include a distribution for the initial state (i.e., the initial state x is distributed according to μ_x). With this reformulation, however, the control action depends on the distribution μ_x instead of just the best estimate of the state $\hat{x}$. Thus, it remains unclear how to extend the analysis used in this work to address SMPC controllers that include the state estimate uncertainty in their problem formulation. Another approach is to establish that SMPC is (distributionally) robust to state estimate error as an unmodeled disturbance.

7.3.4. Forecasts. Often in economic SMPC applications, we consider forecasts of disturbances, such as weather and real-time electricity prices. To address these problems, we first require that the results in this work are extended to time-varying systems to account for the time-varying nature of these forecasts. While we focused on the time-invariant case in this work to avoid the additional notation of time-varying systems, we do not anticipate special difficulties in extending these theoretical results to time-varying systems. Often, these forecasts are updated in real-time as more accurate data become available. In general, however, updates to these forecasts, and the forecasts themselves, may not be independent and identically distributed. If the disturbances are independent but not identically distributed, then an extension of the theory presented in this paper to time-varying systems may be sufficient to address these disturbances.

7.3.5. Disturbances That Are Correlated in Time. Disturbances that are correlated in time are sometimes considered in stochastic optimal control problem formulations, but are rarely considered in the closed-loop analysis of these problems. Recently, some theoretical results consider the problem of disturbances that are correlated in time, but these results are still limited in scope. Hewing et al.,[17] for example, consider disturbances that are correlated in time to construct probabilistic reachable sets and establish recursive feasibility, but derive closed-loop performance bounds for only the case of i.i.d. disturbances. For the idealized SMPC problem, we are unaware of any results that provide closed-loop performance guarantees for systems with disturbances that are correlated in time. There does not appear to be a simple extension of the results in this work to address disturbances that are correlated in time.

7.4. Conclusions. Even our best attempts to model nature's randomness are subject to distributional uncertainty. If properly designed, however, control algorithms that include these models of randomness in their problem formulation are afforded *distributional robustness* by feedback from the plant. In this paper, we defined distributional robustness for closed-loop systems and established sufficient conditions that ensure this property for SMPC. Through a small example, we showed that constructing an SMPC algorithm that satisfies these sufficient conditions is not significantly more difficult than constructing a nominal MPC algorithm. We then demonstrated the implications of distributional robustness for incorrectly modeled, out-of-sample, and unmodeled disturbances with this example. We further established that the distributional robustness of SMPC also addresses scenario-based approximations of the stochastic optimal control problem. Moreover, these results allowed us to characterize the stochastic robustness of nominal MPC as well, and thereby unified the analysis of these two problem formulations. We note that this definition of distributional robustness for closed-loop systems, Definition 11, is general and therefore applicable to the larger field of stochastic optimal control.

AUTHOR INFORMATION

Corresponding Author

Robert D. McAllister − *Department of Chemical Engineering, University of California, Santa Barbara, Santa Barbara, California 93106, United States;* orcid.org/0000-0002-5687-6875; Email: rdmcallister@ucsb.edu

Author

James B. Rawlings − *Department of Chemical Engineering, University of California, Santa Barbara, Santa Barbara, California 93106, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.iecr.2c02970

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the National Science Foundation (NSF) under Grant No. 2027091.

REFERENCES

(1) Ogunnaike, B. A.; Ray, W. H.*Process Dynamics, Modeling, and Control*; Oxford University Press: New York, NY, 1994.

(2) Ogunnaike, B. A.*Random Phenomena*; CRC Press: Boca Raton, FL, 2010.

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

(4) Yu, S.; Reble, M.; Chen, H.; Allgöwer, F. Inherent robustness properties of quasi-infinite horizon nonlinear model predictive control. *Automatica* **2014**, *50*, 2269−2280.

(5) Allan, D. A.; Bates, C. N.; Risbeck, M. J.; Rawlings, J. B. On the inherent robustness of optimal and suboptimal nonlinear MPC. *Sys. Cont. Lett.* **2017**, *106*, 68−78.

(6) Farina, M.; Giulioni, L.; Scattolini, R. Stochastic linear model predictive control with chance constraints—a review. *J. Proc. Cont.* **2016**, *44*, 53−67.
(7) Mesbah, A. Stochastic Model Predictive Control. *IEEE Ctl. Sys. Mag.* **2016**, 30−44.
(8) Mayne, D. Q. Robust and stochastic model predictive control: Are we going in the right direction? *Annual Rev. Control* **2016**, *41*, 184−192.
(9) Primbs, J. A.; Sung, C. H. Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise. *IEEE Trans. Auto. Cont.* **2009**, *54*, 221−230.
(10) Paulson, J. A.; Streif, S.; Mesbah, A.Stability for receding-horizon stochastic model predictive control. *Proc. 2015 Am. Control Conf.***2015**, 937−943.
(11) Cannon, M.; Kouvaritakis, B.; Ng, D. Probabilistic tubes in linear stochastic model predictive control. *Sys. Cont. Lett.* **2009**, *58*, 747−753.
(12) Cannon, M.; Kouvaritakis, B.; Wu, X. Probabilistic constrained MPC for multiplicative and additive stochastic uncertainty. *IEEE Trans. Automat. Contr.* **2009**, *54*, 1626−1632.
(13) Cannon, M.; Kouvaritakis, B.; Rakovic, S. V.; Cheng, Q. Stochastic tubes in model predictive control with probabilistic constraints. *IEEE Trans. Auto. Cont.* **2011**, *56*, 194−200.
(14) Kouvaritakis, B.; Cannon, M.; Raković, S. V.; Cheng, Q. Explicit use of probabilistic distributions in linear predictive control. *Automatica* **2010**, *46*, 1719−1724.
(15) Goulart, P. J.; Kerrigan, E. C. Input-to-state stability of robust receding horizon control with an expected value cost. *Automatica* **2008**, *44*, 1171−1174.
(16) Lorenzen, M.; Dabbene, F.; Tempo, R.; Allgöwer, F. Constraint-tightening and stability in stochastic model predictive control. *IEEE Trans. Auto. Cont.* **2017**, *62*, 3165−3177.
(17) Hewing, L.; Wabersich, K. P.; Zeilinger, M. N. Recursively feasible stochastic model predictive control using indirect feedback. *Automatica* **2020**, *119*, 109095.
(18) Sehr, M. A.; Bitmead, R. R. Stochastic output-feedback model predictive control. *Automatica* **2018**, *94*, 315−323.
(19) Chatterjee, D.; Lygeros, J. On stability and performance of stochastic predictive control techniques. *IEEE Trans. Auto. Cont.* **2015**, *60*, 509−514.
(20) Mayne, D. Q.; Falugi, P. Stabilizing conditions for model predictive control. *Int. J. Robust and Nonlinear Control* **2019**, *29*, 894−903.
(21) McAllister, R. D.; Rawlings, J. B. Nonlinear Stochastic Model Predictive Control: Existence, Measurability, and Stochastic Asymptotic Stability. *IEEE Trans. Automat. Contr.* **2022**, 1.
(22) McAllister, R. D.; Rawlings, J. B. On the inherent distributional robustness of stochastic and nominal model predictive control. *IEEE Trans. Automat. Contr.* **2022**, *67*, 5166.
(23) Bertsekas, D.*Dynamic programming and Optimal Control*; Athena Scientific: Belmont, MA, 2017; Vol. *1*.
(24) Simon, H. A. Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica* **1956**, *24*, 74−81.
(25) Theil, H. A note on certainty equivalence in dynamic planning. *Econometrica* **1957**, *25*, 346−349.
(26) Van de Water, H.; Willems, J. The certainty equivalence property in stochastic control theory. *IEEE Trans. Auto. Cont.* **1981**, *26*, 1080−1087.
(27) Zheng, A.; Morari, M.Control of linear unstable systems with constraints. *Proc. Am. Control Conf.*1995, 3704−3708.
(28) Scokaert, P. O. M.; Rawlings, J. B. Feasibility Issues in Linear Model Predictive Control. *AIChE J.* **1999**, *45*, 1649−1659.
(29) Kerrigan, E. C.; Maciejowski, J. M.Soft constraints and exact penalty functions in model predictive control. *Control 2000 Conf.*2000, 2319−2327.
(30) Rawlings, J. B.; Mayne, D. Q.; Diehl, M. M.*Model Predictive Control: Theory, Design, and Computation*, 2nd ed.; Nob Hill Publishing: Santa Barbara, CA, 2020.
(31) Villani, C.*Optimal Transport: Old and New*; Springer, 2009; Vol. *338*.
(32) Frogner, C.; Zhang, C.; Mobahi, H.; Araya, M.; Poggio, T. A.Learning with a Wasserstein loss. *Adv. Neural Inf. Process. Syst.***2015**, 28.
(33) Arjovsky, M.; Chintala, S.; Bottou, L.Wasserstein generative adversarial networks. *Int. Conf. Mach. Learn.***2017**, 214−223.
(34) Goh, J.; Sim, M. Distributionally robust optimization and its tractable approximations. *Oper. Res.* **2010**, *58*, 902−917.
(35) Shafieezadeh Abadeh, S.; Nguyen, V. A.; Kuhn, D.; Mohajerin Esfahani, P. M. Wasserstein distributionally robust Kalman filtering. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 8067.
(36) Yang, I. Wasserstein distributionally robust stochastic control: A data-driven approach. *IEEE Trans. Auto. Cont.* **2021**, *66*, 3863−3870.
(37) Givens, C. R.; Shortt, R. M. A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **1984**, *31*, 231−240.
(38) Lucia, S.; Finkler, T.; Engell, S. Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *J. Proc. Cont.* **2013**, *23*, 1306−1319.
(39) Bonzanini, A. D.; Paulson, J. A.; Makrygiorgos, G.; Mesbah, A. Fast approximate learning-based multistage nonlinear model predictive control using Gaussian processes and deep neural networks. *Comput. Chem. Eng.* **2021**, *145*, 107174.
(40) Varadarajan, V. S. On the convergence of sample probability distributions. *Sankhya* **1958**, *19*, 23−26.
(41) Jerez, J. L.; Kerrigan, E. C.; Constantinides, G. A. A condensed and sparse QP formulation for predictive control. *2011 50th IEEE Conf. Decision Control Eur. Control Conf.* **2011**, 5217−5222.
(42) Rossiter, J. A.; Kouvaritakis, B.; Rice, M. J. A numerically robust state-space approach to stable-predictive control strategies. *Automatica* **1998**, *34*, 65−73.
(43) Kumar, R.; Jalving, J.; Wenzel, M. J.; Ellis, M. J.; ElBsat, M. N.; Drees, K. H.; Zavala, V. M. Benchmarking stochastic and deterministic MPC: A case study in stationary battery systems. *AIChE J.* **2019**, *65*, No. e16551.

https://doi.org/10.1021/acs.iecr.2c02970
Ind. Eng. Chem. Res. 2023, 62, 2175−2190