

A Framework for Time-Consistent, Risk-Averse Model Predictive Control: Theory and Algorithms

Yin-Lam Chow, Marco Pavone

Abstract—In this paper we present a framework for risk-averse model predictive control (MPC) of linear systems affected by multiplicative uncertainty. Our key innovation is to consider time-consistent, dynamic risk metrics as objective functions to be minimized. This framework is axiomatically justified in terms of time-consistency of risk preferences, is amenable to dynamic optimization, and is unifying in the sense that it captures a full range of risk assessments from risk-neutral to worst case. Within this framework, we propose and analyze an online risk-averse MPC algorithm that is provably stabilizing. Furthermore, by exploiting the dual representation of time-consistent, dynamic risk metrics, we cast the computation of the MPC control law as a convex optimization problem amenable to implementation on embedded systems. Simulation results are presented and discussed.

I. INTRODUCTION

Model predictive control (MPC) is one of the most popular methods to address optimal control problems in an online setting [1], [2]. The key idea behind MPC is to obtain the control action by repeatedly solving, at each sampling instant, a finite horizon open-loop optimal control problem, using the current state of the plant as the initial state; the result of the optimization is an (open-loop) control sequence, whose first element is applied to control the system [3].

The classic MPC framework does not provide a systematic way to address model uncertainties and disturbances [4]. Accordingly, one of the main research thrusts for MPC is to find techniques to guarantee persistent feasibility and stability in the presence of disturbances. Essentially, current techniques fall into two categories: (1) min-max (or worst-case) formulations, where the performance indices to be minimized are computed with respect to the worst possible disturbance realization [5], [6], [7], and (2) stochastic formulations, where *risk-neutral expected* values of performance indices (and possibly constraints) are considered [4], [8].

The main drawback of the worst-case approach is that the control law may be too conservative, since the MPC law is required to guarantee stability and constraint fulfillment under the worst-case scenario. On the other hand, stochastic formulations whereby the assessment of future random outcomes is accomplished through a risk-neutral expectation may be unsuitable in scenarios where one desires to protect the system from large deviations.

The objective of this paper is to introduce a systematic method to include risk-aversion in MPC. The inclusion of risk aversion is important for several reasons. First, in uncertain environments, a guaranteed-feasible solution may not exist and the issue becomes how to properly balance between

planner conservatism and the risk of infeasibility (clearly this can not be achieved with a worst-case approach). Second, risk-aversion allows the control designer to increase policy robustness by limiting confidence in the model. Third, risk-aversion serves to prevent rare undesirable events. Finally, in a reinforcement learning framework when the world model is not accurately known, a risk-averse agent can cautiously balance exploration versus exploitation for fast convergence and to avoid “bad” states that could potentially lead to a catastrophic failure [9], [10].

Inclusion of risk-aversion in MPC is difficult for two main reasons. First, it appears to be difficult to model risk in multi-period settings in a way that matches intuition [10]. In particular, a common strategy to include risk-aversion in multi-period contexts is to apply a *static* risk metric, which assesses risk from the perspective of a single point in time, to the total cost of the future stream of random outcomes. However, using static risk metrics in multi-period decision problems can lead to an over or under-estimation of the true dynamic risk, as well as to a potentially “inconsistent” behavior, see [11] and references therein. Second, optimization problems involving risk metrics tend to be computationally intractable, as they do not allow a recursive estimation of risk. In practice, risk-averse MPC often resolves into the minimization of the expected value of an aggregate, risk-weighted objective function [12], [13].

In this paper, as a radical departure from traditional approaches, we leverage recent strides in the theory of *dynamic* risk metrics developed by the operations research community [14], [15] to include risk aversion in MPC. The key property of *dynamic* risk metrics is that, by assessing risk at multiple points in time, one can guarantee *time-consistency* of risk preferences over time [14], [15]. In particular, the essential requirement for time consistency is that if a certain outcome is considered less risky in all states of the world at stage $k + 1$, then it should also be considered less risky at stage k . Remarkably, in [15], it is proven that any risk measure that is time consistent can be represented as a *composition* of one-step risk metrics, in other words, in multi-period settings, risk (as expected) should be compounded over time.

The contribution of this paper is threefold. First, we introduce a notion of dynamic risk metric, referred to as Markov dynamic polytopic risk metric, that captures a full range of risk assessments and enjoys a geometrical structure that is particularly favorable from a computational standpoint. Second, we present and analyze a *risk-averse* MPC algorithm that minimizes in a receding-horizon fashion a Markov dynamic polytopic risk metric, under the assumption that the system’s model is linear and is affected by multiplicative uncertainty. Finally, by exploring the “geometrical” structure of Markov dynamic polytopic risk metrics, we present a convex programming formulation for risk-averse MPC that is

Y.-L. Chow is with the Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA. Email: ychow@stanford.edu.

M. Pavone is with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA. Email: pavone@stanford.edu.

amenable to real-time implementation (for moderate horizon lengths). Our framework has three main advantages: (1) it is axiomatically justified, in the sense that risk, by construction, is assessed in a time-consistent fashion; (2) it is amenable to dynamic and convex optimization, thanks to the compositional form of Markov dynamic polytopic risk metrics and their geometry; and (3) it is general, in the sense that it captures a full range of risk assessments from risk-neutral to worst case. In this respect, our formulation represents a *unifying* approach for risk-averse MPC.

The rest of the paper is organized as follows. In Section II we provide a review of the theory of dynamic risk metrics. In Section III we discuss the stochastic model we will consider in this paper. In Section IV we introduce and discuss the notion of Markov dynamic polytopic risk metrics. In Section V we state the infinite horizon optimal control problem we wish to address and in Section VI we derive conditions for risk-averse closed-loop stability. In Section VII and VIII we present, respectively, a risk-averse model predictive control law and approaches for its computation. Numerical experiments are presented and discussed in Section IX. Finally, in Section X we draw some conclusions and we discuss directions for future work.

II. REVIEW OF DYNAMIC RISK THEORY

In this section, we briefly describe the theory of coherent and dynamic risk metrics, on which we will rely extensively later in the paper. The material presented in this section summarizes several novel results in risk theory achieved in the past 10 years. Our presentation strives to present this material in a intuitive fashion and with a notation tailored to control applications.

A. Static, coherent measures of risk

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of outcomes (sample space), \mathcal{F} is a σ -algebra over Ω representing the set of events we are interested in, and \mathbb{P} is a probability measure over \mathcal{F} . In this paper we will focus on disturbance models characterized by probability *mass* functions, hence we restrict our attention to finite probability spaces (i.e., Ω has a finite number of elements or, equivalently, \mathcal{F} is a finitely generated algebra). Denote with \mathcal{Z} the space of random variables $Z : \Omega \mapsto (-\infty, \infty)$ defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this paper a random variable $Z \in \mathcal{Z}$ is interpreted as a cost, i.e., the smaller the realization of Z , the better. For Z, W , we denote by $Z \leq W$ the point-wise partial order, i.e., $Z(\omega) \leq W(\omega)$ for all $\omega \in \Omega$.

By a *risk measure* (or *risk metric*, we will use these terms interchangeably) we understand a function $\rho(Z)$ that maps an uncertain outcome Z into the extended real line $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. In this paper we restrict our analysis to *coherent risk measures*, defined as follows:

Definition II.1 (Coherent Risk Measures). *A coherent risk measure is a mapping $\rho : \mathcal{Z} \rightarrow \mathbb{R}$, satisfying the following four axioms:*

- A1 *Convexity:* $\rho(\lambda Z + (1 - \lambda)W) \leq \lambda\rho(Z) + (1 - \lambda)\rho(W)$, for all $\lambda \in [0, 1]$ and $Z, W \in \mathcal{Z}$;
- A2 *Monotonicity:* if $Z \leq W$ and $Z, W \in \mathcal{Z}$, then $\rho(Z) \leq \rho(W)$;
- A3 *Translation invariance:* if $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$;

- A4 *Positive homogeneity:* if $\lambda \geq 0$ and $Z \in \mathcal{Z}$, then $\rho(\lambda Z) = \lambda\rho(Z)$.

These axioms were originally conceived in [16] and ensure the “rationality” of single-period risk assessments (we refer the reader to [16] for a detailed motivation of these axioms). One of the main properties for coherent risk metrics is a universal representation theorem for coherent risk metrics, which in the context of *finite* probability spaces takes the following form:

Theorem II.2 (Representation Theorem for Finite Probability Spaces [17, page 265]). *Consider the probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ where Ω is finite, i.e., $\Omega = \{\omega_1, \dots, \omega_L\}$, \mathcal{F} is the σ -algebra of all subsets (i.e., $\mathcal{F} = 2^\Omega$), and $\mathbb{P} = (p(1), \dots, p(L))$, with all probabilities positive. Let \mathcal{B} be the set of probability density functions:*

$$\mathcal{B} := \left\{ \zeta \in \mathbb{R}^L : \sum_{j=1}^L p(j)\zeta(j) = 1, \zeta \geq 0 \right\}.$$

The risk measure $\rho : \mathcal{Z} \rightarrow \mathbb{R}$ is a coherent risk measure if and only if there exists a convex bounded and weakly closed set $\mathcal{U} \subset \mathcal{B}$ such that $\rho(Z) = \max_{\zeta \in \mathcal{U}} \mathbb{E}_\zeta[Z]$.*

The result essentially states that any coherent risk measure is an expectation with respect to a worst-case density function ζ , chosen adversarially from a suitable *set* of test density functions (referred to as *risk envelope*).

B. Dynamic, time-consistent measures of risk

This section provides a multi-period generalization of the concepts presented in Section II-A and follows closely the discussion in [15]. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}_N \subset \mathcal{F}$, and an adapted sequence of real-valued random variables $Z_k, k \in \{0, \dots, N\}$. We assume that $\mathcal{F}_0 = \{\Omega, \emptyset\}$, i.e., Z_0 is deterministic. The variables Z_k can be interpreted as stage-wise costs. For each $k \in \{0, \dots, N\}$, denote with \mathcal{Z}_k the space of random variables defined over the probability space $(\Omega, \mathcal{F}_k, \mathbb{P})$; also, let $\mathcal{Z}_{k,N} := \mathcal{Z}_k \times \cdots \times \mathcal{Z}_N$. Given sequences $Z = \{Z_k, \dots, Z_N\} \in \mathcal{Z}_{k,N}$ and $W = \{W_k, \dots, W_N\} \in \mathcal{Z}_{k,N}$, we interpret $Z \leq W$ component-wise, i.e., $Z_j \leq W_j$ for all $j \in \{k, \dots, N\}$.

The fundamental question in the theory of dynamic risk measures is the following: how do we evaluate the risk of the sequence $\{Z_k, \dots, Z_N\}$ from the perspective of stage k ? The answer, within the modern theory of risk, relies on two key intuitive facts [15]. First, in dynamic settings the specification of risk preferences should no longer entail constructing a single risk metric but rather a *sequence* of risk metrics $\{\rho_{k,N}\}_{k=0}^N$, each mapping a future stream of random costs into a risk metric/assessment at time k . This motivates the following definition.

Definition II.3 (Dynamic Risk Measure). *A dynamic risk measure is a sequence of mappings $\rho_{k,N} : \mathcal{Z}_{k,N} \rightarrow \mathcal{Z}_k, k \in \{0, \dots, N\}$, obeying the following monotonicity property:*

$$\rho_{k,N}(Z) \leq \rho_{k,N}(W) \text{ for all } Z, W \in \mathcal{Z}_{k,N} \text{ such that } Z \leq W.$$

The above monotonicity property (analogous to axiom A2 in Definition II.1) is, arguably, a natural requirement for any meaningful dynamic risk measure.

Second, the sequence of metrics $\{\rho_{k,N}\}_{k=0}^N$ should be constructed so that the risk preference profile is *consistent* over time [18], [19], [11]. A widely accepted notion of time-consistency is as follows [15]: if a certain outcome is considered less risky in all states of the world at stage $k+1$, then it should also be considered less risky at stage k .

The following example (adapted from [20]) shows how dynamic risk measures as defined above might indeed result in *time-inconsistent*, and ultimately undesirable, behaviors.

Example II.4. Consider the simple setting whereby there is a final cost Z and one seeks to evaluate such cost from the perspective of earlier stages. Consider the three-stage scenario tree in Figure 1, with the elementary events $\Omega = \{UU, UD, DU, DD\}$, and the filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \{\emptyset, \{U\}, \{D\}, \Omega\}$, and $\mathcal{F}_2 = 2^\Omega$. Consider the dynamic risk measure:

$$\rho_{k,N}(Z) := \max_{q \in \mathcal{U}} \mathbb{E}_q[Z | \mathcal{F}_k], \quad k = 0, 1, 2$$

where \mathcal{U} contains two probability measures, one corresponding to $p = 0.4$, and the other one to $p = 0.6$. Assume that the random cost is $Z(UU) = Z(DD) = 0$, and $Z(UD) = Z(DU) = 100$. Then, one has $\rho_1(Z)(\omega) = 60$ for all ω , and $\rho_0(Z)(\omega) = 48$. Therefore, Z is deemed strictly riskier than a deterministic cost $W = 50$ in all states of nature at time $k = 1$, but nonetheless W is deemed riskier than Z at time $k = 0$ – a paradox!

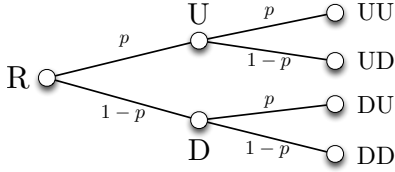


Fig. 1. Scenario tree for example II.4.

It is important to note that there is nothing special about the selection of this example, similar paradoxical results could be obtained with other risk metrics. We refer the reader to [15], [19], [11] for further insights into the notion of time consistency and its practical relevance. The issue then is what additional “structural” properties are required for a dynamic risk measure to be time consistent. We first provide a rigorous version of the previous definition of time-consistency.

Definition II.5 (Time Consistency ([15])). A dynamic risk measure $\{\rho_{k,N}\}_{k=0}^N$ is time-consistent if, for all $0 \leq l < k \leq N$ and all sequences $Z, W \in \mathcal{Z}_{l,N}$, the conditions

$$\begin{aligned} Z_i &= W_i, \quad i = l, \dots, k-1, \text{ and} \\ \rho_{k,N}(Z_k, \dots, Z_N) &\leq \rho_{k,N}(W_k, \dots, W_N), \end{aligned} \quad (1)$$

imply that

$$\rho_{l,N}(Z_l, \dots, Z_N) \leq \rho_{l,N}(W_l, \dots, W_N).$$

As we will see in Theorem II.7, the notion of time-consistent risk measures is tightly linked to the notion of coherent risk measures, whose generalization to the multi-period setting is given below:

Definition II.6 (Coherent One-step Conditional Risk Measures ([15])). A coherent one-step conditional risk measure

is a mapping $\rho_k : \mathcal{Z}_{k+1} \rightarrow \mathcal{Z}_k$, $k \in \{0, \dots, N-1\}$, with the following four properties:

- *Convexity*: $\rho_k(\lambda Z + (1-\lambda)W) \leq \lambda \rho_k(Z) + (1-\lambda) \rho_k(W)$, $\forall \lambda \in [0, 1]$ and $Z, W \in \mathcal{Z}_{k+1}$;
- *Monotonicity*: if $Z \leq W$ then $\rho_k(Z) \leq \rho_k(W)$, $\forall Z, W \in \mathcal{Z}_{k+1}$;
- *Translation invariance*: $\rho_k(Z+W) = Z + \rho_k(W)$, $\forall Z \in \mathcal{Z}_k$ and $W \in \mathcal{Z}_{k+1}$;
- *Positive homogeneity*: $\rho_k(\lambda Z) = \lambda \rho_k(Z)$, $\forall Z \in \mathcal{Z}_{k+1}$ and $\lambda \geq 0$.

We are now in a position to state the main result of this section.

Theorem II.7 (Dynamic, Time-consistent Risk Measures ([15])). Consider, for each $k \in \{0, \dots, N\}$, the mappings $\rho_{k,N} : \mathcal{Z}_{k,N} \rightarrow \mathcal{Z}_k$ defined as

$$\begin{aligned} \rho_{k,N} &= Z_k + \rho_k(Z_{k+1} + \rho_{k+1}(Z_{k+2} + \dots + \\ &\rho_{N-2}(Z_{N-1} + \rho_{N-1}(Z_N)) \dots)), \end{aligned} \quad (2)$$

where the ρ_k 's are coherent one-step conditional risk measures. Then, the ensemble of such mappings is a dynamic, time-consistent risk measure.

Remarkably, Theorem 1 in [15] shows (under weak assumptions) that the “multi-stage composition” in equation (2) is indeed *necessary for time consistency*. Accordingly, in the remainder of this paper, we will focus on the *dynamic, time-consistent risk measures* characterized in Theorem II.7.

III. MODEL DESCRIPTION

Consider the discrete time system:

$$x_{k+1} = A(w_k)x_k + B(w_k)u_k, \quad (3)$$

where $k \in \mathbb{N}$ is the time index, $x_k \in \mathbb{R}^{N_x}$ is the state, $u_k \in \mathbb{R}^{N_u}$ is the (unconstrained) control input, and $w_k \in \mathcal{W}$ is the process disturbance. We assume that the initial condition x_0 is deterministic. We assume that \mathcal{W} is a finite set of cardinality L , i.e., $\mathcal{W} = \{w^{[1]}, \dots, w^{[L]}\}$. For each stage k and state-control pair (x_k, u_k) , the process disturbance w_k is drawn from set \mathcal{W} according to the probability mass function

$$p = [p(1), p(2), \dots, p(L)]^T,$$

where $p(j) = \mathbb{P}(w_k = w^{[j]})$, $j \in \{1, \dots, L\}$. Without loss of generality, we assume that $p(j) > 0$ for all j . Note that the probability mass function for the process disturbance is time-invariant, and that the process disturbance is *independent* of the process history and of the state-control pair (x_k, u_k) . Under these assumptions, the stochastic process $\{x_k\}$ is clearly a Markov process.

By enumerating all L realizations of the process disturbance w_k , system (3) can be rewritten as:

$$x_{k+1} = \begin{cases} A_1 x_k + B_1 u_k & \text{if } w_k = w^{[1]}, \\ \vdots & \vdots \\ A_L x_k + B_L u_k & \text{if } w_k = w^{[L]}, \end{cases}$$

where $A_j := A(w^{[j]})$ and $B_j := B(w^{[j]})$, $j \in \{1, \dots, L\}$.

The results presented in this paper can be immediately extended to the time-varying case (i.e., where the probability mass function for the process disturbance is time-varying). To simplify notation, however, we prefer to focus this paper on the time-invariant case.

IV. MARKOV POLYTOPIC RISK MEASURES

In this section we *refine* the notion of dynamic time-consistent risk metrics (as defined in Theorem II.7) in two ways: (1) we add a polytopic structure to the dual representation of coherent risk metrics, and (2) we add a Markovian structure. This will lead to the definition of Markov dynamic polytopic risk metrics, which enjoy favorable computational properties and, at the same time, maintain most of the generality of dynamic time-consistent risk metrics.

A. Polytopic risk measures

According to the discussion in Section III, the probability space for the process disturbance has a finite number of elements. Accordingly, consider Theorem II.2; by definition of expectation, one has $\mathbb{E}_\zeta[Z] = \sum_{j=1}^L Z(j)p(j)\zeta(j)$. In our framework (inspired by [21]), we consider coherent risk measures where the risk envelope \mathcal{U} is a *polytope*, i.e., there exist matrices S^I, S^E and vectors T^I, T^E of appropriate dimensions such that

$$\mathcal{U}^{\text{poly}} = \{ \zeta \in \mathcal{B} \mid S^I \zeta \leq T^I, S^E \zeta = T^E \}.$$

We will refer to coherent risk measures representable with a polytopic risk envelope as *polytopic risk measures*. Consider the bijective map $q(j) := p(j)\zeta(j)$ (recall that, in our model, $p(j) > 0$). Then, by applying such map, one can easily rewrite a polytopic risk measure as

$$\rho(Z) = \max_{q \in \mathcal{U}^{\text{poly}}} \mathbb{E}_q[Z],$$

where q is a *probability mass function* belonging to a polytopic subset of the standard simplex, i.e.:

$$\mathcal{U}^{\text{poly}} = \left\{ q \in \Delta^L \mid S^I q \leq T^I, S^E q = T^E \right\}, \quad (4)$$

where $\Delta^L := \{ q \in \mathbb{R}^L : \sum_{j=1}^L q(j) = 1, q \geq 0 \}$. Accordingly, one has $\mathbb{E}_q[Z] = \sum_{j=1}^L Z(j)q(j)$ (note that, with a slight abuse of notation, we are using the same symbols as before for $\mathcal{U}^{\text{poly}}, S^I$, and S^E).

The class of polytopic risk measures is large: we give below some examples (also note that any comonotonic risk measure is a polytopic risk measure [11]).

Example IV.1. (*Examples of Polytopic Risk Measures*) The expected value of a random variable Z can be represented according to equation (4) with the polytopic risk envelope $\mathcal{U}^{\text{poly}} = \{ q \in \Delta^L \mid q(j) = p(j) \text{ for all } j \in \{1, \dots, L\} \}$.

A second example is represented by the average upper semi-deviation risk metric, defined as

$$\rho_{\text{AUS}}(Z) := \mathbb{E}[Z] + c \mathbb{E}[(Z - \mathbb{E}[Z])^+],$$

where $0 \leq c \leq 1$. This metric can be represented according to equation (4) with polytopic risk envelope ([22], [17]): $\mathcal{U}^{\text{poly}} = \{ q \in \Delta^L \mid q(j) = p(j)(1 + h(j) - \sum_{j=1}^L h(j)p(j)), 0 \leq h(j) \leq c, j \in \{1, \dots, L\} \}$.

A third example is represented by the worst case risk, defined as

$$\text{WCR}(Z) := \max\{ Z(j) : j \in \{1, \dots, L\} \}.$$

Such risk metric can be trivially represented according to equation (4) with polytopic risk envelope $\mathcal{U}^{\text{poly}} = \Delta^L$.

Other important examples include the Conditional Value-at-Risk [23], mean absolute semi-deviation [22], the spectral risk measures [24], [11], the optimized certainty equivalent and expected utility [25], [17], [21], and the distributionally-robust risk [4]. The key point is that the notion of polytopic risk metric *covers a full gamut of risk assessments*, ranging from risk-neutral to worst case.

B. Markov dynamic polytopic risk metrics

Note that in the definition of dynamic, time-consistent risk measures, since at stage k the value of ρ_k is \mathcal{F}_k -measurable, the evaluation of risk can depend on the *whole* past, see [15, Section IV]. For example, the weight c in the definition of the average upper mean semi-deviation risk metric can be an \mathcal{F}_k -measurable random variable (see [15, Example 2]). This generality, which appears of little practical value in many cases, leads to optimization problems that are intractable. This motivates us to add a *Markovian structure* to dynamic, time-consistent risk measures (similarly as in [15]). We start by introducing the notion of Markov polytopic risk measure (similar to [15, Definition 6]).

Definition IV.2 (Markov Polytopic Risk Measures). *Consider the Markov process $\{x_k\}$ that evolves according to equation (3). A coherent one-step conditional risk measure $\rho_k(\cdot)$ is a Markov polytopic risk measure with respect to $\{x_k\}$ if it can be written as*

$$\rho_k(Z(x_{k+1})) = \max_{q \in \mathcal{U}_k^{\text{poly}}(x_k, p)} \mathbb{E}_q[Z(x_{k+1})]$$

where

$$\mathcal{U}_k^{\text{poly}}(x_k, p) = \left\{ q \in \Delta^L \mid \begin{array}{l} S_k^I(x_k, p)q \leq T_k^I(x_k, p), \\ S_k^E(x_k, p)q = T_k^E(x_k, p) \end{array} \right\}.$$

In other words, a Markov polytopic risk measure is a coherent one-step conditional risk measure where the evaluation of risk is not allowed to depend on the whole past (for example, the weight c in the definition of the average upper mean semi-deviation risk metric can depend on the past only through x_k), and the risk envelope is a polytope. Correspondingly, we define a Markov dynamic polytopic risk metric as follows.

Definition IV.3 (Markov Dynamic Polytopic Risk Measures). *Consider the Markov process $\{x_k\}$ that evolves according to equation (3). A Markov dynamic polytopic risk measure is a set of mappings $\rho_{k,N} : \mathcal{Z}_{k,N} \rightarrow \mathcal{Z}_k$ defined as*

$$\rho_{k,N} = Z(x_k) + \rho_k(Z(x_{k+1}) + \dots + \rho_{N-2}(Z(x_{N-1}) + \rho_{N-1}(Z(x_N)))) \dots),$$

for $k \in \{0, \dots, N\}$, where the single-period risk measures ρ_k are Markov polytopic risk measures.

Clearly, a Markov dynamic polytopic risk metric is time consistent. Definition IV.3 can be extended to the case where the probability distribution for the disturbance depends on the current state and control action. We avoid this generalization to keep the exposition simple and consistent with model (3).

V. PROBLEM FORMULATION

In light of Sections III and IV, we are now in a position to state the risk-averse MPC problem we wish to solve in this paper. Our problem formulation relies on Markov dynamic

polytopic risk metrics that satisfy the following stationarity assumption.

Assumption V.1. (*Time-Invariance of Risk*) The polytopic risk envelopes $\mathcal{U}_k^{\text{poly}}$ do not depend explicitly on time and are independent of the state x_k , i.e. $\mathcal{U}_k^{\text{poly}}(x_k, p) = \mathcal{U}^{\text{poly}}(p)$ for all k .

This assumption is crucial for the well-posedness of our formulation and in order to devise a tractable MPC algorithm that relies on linear matrix inequalities. We next introduce a notion of stability tailored to our risk-averse context.

Definition V.2 (Uniform Global Risk-Sensitive Exponential Stability). System (3) is said to be Uniformly Globally Risk-Sensitive Exponentially Stable (UGRSES) if there exist constants $c \geq 0$ and $\lambda \in [0, 1)$ such that for all initial conditions $x_0 \in \mathbb{R}^{N_x}$,

$$\rho_{0,k}(0, \dots, 0, x_k^T x_k) \leq c \lambda^k x_0^T x_0, \quad \text{for all } k \in \mathbb{N}, \quad (5)$$

where $\{\rho_{0,k}\}$ is a Markov dynamic polytopic risk measure satisfying Assumption V.1.

One can easily show that UGRSES is a *more restrictive* stability condition than mean-square stability (considered, for example, in [4]).

Consider the MDP described in Section III and let Π be the set of all stationary feedback control policies, i.e., $\Pi := \{\pi : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_u}\}$. Consider the quadratic cost function $c : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \rightarrow \mathbb{R}_{\geq 0}$ defined as

$$c(x, u) := x^T Q x + u^T R u,$$

where $Q = Q^T \succ 0$ and $R = R^T \succ 0$ are given state and control penalties. The problem we wish to address is as follows.

Optimization Problem \mathcal{OPT} — Given an initial state $x_0 \in \mathbb{R}^{N_x}$, solve

$$\begin{aligned} & \inf_{\pi \in \Pi} \quad \limsup_{k \rightarrow \infty} J_{0,k}(x_0, \pi) \\ & \text{such that } x_{k+1} = A(w_k)x_k + B(w_k)\pi(x_k) \\ & \quad \text{System is UGRSES} \end{aligned}$$

where

$$J_{0,k}(x_0, \pi) = \rho_{0,k}(c(x_0, \pi(x_0)), \dots, c(x_k, \pi(x_k))),$$

and $\{\rho_{0,k}\}$ is a Markov dynamic polytopic risk measure satisfying Assumption V.1.

We will denote the optimal cost function as $J_{\infty}^*(x_0)$. Note that the risk measure in the definition of UGRSES is assumed to be identical to the risk measure used to evaluate the cost of a policy. Also, note that, by Assumption V.1, the single-period risk metrics are time-invariant, hence one can write

$$\begin{aligned} & \rho_{0,k}(c(x_0, \pi(x_0)), \dots, c(x_k, \pi(x_k))) = c(x_0, \pi(x_0)) \\ & + \rho(c(x_1, \pi(x_1)) + \dots + \rho(c(x_k, \pi(x_k)))) \dots, \end{aligned} \quad (6)$$

where ρ is a given Markov polytopic risk metric that models the ‘‘amount’’ of risk aversion. This paper addresses problem \mathcal{OPT} along two main dimensions:

- 1) Find sufficient conditions for *risk-sensitive* stability (i.e., for UGRSES).

- 2) Design a model predictive control algorithm to efficiently compute a suboptimal state-feedback control policy.

VI. RISK-SENSITIVE STABILITY

In this section we provide a sufficient condition for system (3) to be UGRSES, under the assumptions of Section V. This condition relies on Lyapunov techniques and is inspired by [4] (Lemma VI.1 indeed reduces to Lemma 1 in [4] when the risk measure is simply an expectation).

Lemma VI.1. Consider a policy $\pi \in \Pi$ and the corresponding closed-loop dynamics for system (3), denoted by $x_{k+1} = f(x_k, w_k)$. The closed-loop system is UGRSES if there exists a function $V(x) : \mathbb{R}^{N_x} \rightarrow \mathbb{R}$ and scalars $b_1, b_2, b_3 > 0$ such that for all $x \in \mathbb{R}^{N_x}$ the following conditions hold:

$$\begin{aligned} b_1 \|x\|^2 & \leq V(x) \leq b_2 \|x\|^2, \\ \rho(V(f(x, w))) - V(x) & \leq -b_3 \|x\|^2. \end{aligned} \quad (7)$$

Proof. From the time consistency, monotonicity, translational invariance, and positive homogeneity of Markov dynamic polytopic risk measures, condition (7) implies $b_1 \rho_{0,k+1}(0, \dots, 0, \|x_{k+1}\|^2) \leq \rho_{0,k+1}(0, \dots, 0, V(x_{k+1})) = \rho_{0,k}(0, \dots, 0, V(x_k) + \rho(V(x_{k+1}) - V(x_k))) \leq \rho_{0,k}(0, \dots, 0, V(x_k) - b_3 \|x_k\|^2) \leq \rho_{0,k}(0, \dots, 0, (b_2 - b_3) \|x_k\|^2)$. Also, since $\rho_{0,k+1}$ is monotonic, one has $\rho_{0,k+1}(0, \dots, 0, \|x_{k+1}\|^2) \geq 0$, which implies $b_2 \geq b_3$ and in turn $(1 - b_3/b_2) \in [0, 1)$.

Since $V(x_k)/b_2 \leq \|x_k\|^2$, and using the previous inequalities, one can write:

$$\begin{aligned} \rho_{0,k+1}(0, \dots, 0, V(x_{k+1})) & \leq \rho_{0,k}(0, \dots, 0, V(x_k) - b_3 \|x_k\|^2) \\ & \leq \left(1 - \frac{b_3}{b_2}\right) \rho_{0,k}(0, \dots, 0, V(x_k)). \end{aligned}$$

Repeating this bounding process, one obtains:

$$\begin{aligned} \rho_{0,k+1}(0, \dots, 0, V(x_{k+1})) & \leq \left(1 - \frac{b_3}{b_2}\right)^k \rho_{0,1}(V(x_1)) \\ & = \left(1 - \frac{b_3}{b_2}\right)^k \rho(V(x_1)) \leq \left(1 - \frac{b_3}{b_2}\right)^k (V(x_0) - b_3 \|x_0\|^2) \\ & \leq b_2 \left(1 - \frac{b_3}{b_2}\right)^{k+1} \|x_0\|^2. \end{aligned}$$

Again, by monotonicity, the above result implies

$$\rho_{0,k+1}(0, \dots, 0, x_{k+1}^T x_{k+1}) \leq \frac{b_2}{b_1} \left(1 - \frac{b_3}{b_2}\right)^{k+1} x_0^T x_0.$$

By setting $c = b_2/b_1$ and $\lambda = (1 - b_3/b_2) \in [0, 1)$, the claim is proven. \square

VII. MODEL PREDICTIVE CONTROL PROBLEM

In this section we set up the receding horizon version of problem \mathcal{OPT} . This will lead to a model predictive control algorithm for the (suboptimal) solution of problem \mathcal{OPT} . Consider the following receding-horizon cost function for $N \geq 1$:

$$\begin{aligned} J(x_{k|k}, \pi_{k|k}, \dots, \pi_{k+N-1|k}, P) & := \rho_{k,k+N}(c(x_{k|k}, \pi_{k|k}(x_{k|k})), \\ & \dots, c(x_{k+N-1|k}, \pi_{k+N-1|k}(x_{k+N-1|k})), x_{k+N}^T P x_{k+N}), \end{aligned} \quad (8)$$

where $x_{h|k}$ is the state at time h predicted at stage k (a discrete random variable), $\pi_{h|k}$ is the control policy to be applied at time h as determined at stage k (i.e., $\pi_{h|k} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_u}$), and $P = P^T \succ 0$ is a terminal weight matrix. We are now in a position to state the model predictive control problem.

Optimization problem \mathcal{MPC} — Given an initial state $x_{k|k} \in \mathbb{R}^{N_x}$ and a prediction horizon $N \geq 1$, solve

$$\begin{aligned} \min_{\pi_{k|k}, \dots, \pi_{k+N-1|k}} \quad & J(x_{k|k}, \pi_{k|k}, \dots, \pi_{k+N-1|k}, P) \\ \text{such that} \quad & x_{k+h+1|k} = A(w_{k+h})x_{k+h|k} + \\ & B(w_{k+h})\pi_{k+h|k}(x_{k+h|k}) \end{aligned}$$

for $h \in \{0, \dots, N-1\}$.

Note that a Markov policy is guaranteed to be optimal for problem \mathcal{MPC} (see [15, Theorem 2]). The optimal cost function for problem \mathcal{MPC} is denoted by $J_k^*(x_{k|k})$, and the set of minimizers is denoted by $\{\pi_{k|k}^*, \dots, \pi_{k+N-1|k}^*\}$. For each state x_k , we set $x_{k|k} = x_k$ and the (time-invariant) model predictive control law is then defined as

$$\pi^{MPC}(x_k) = \pi_{k|k}^*(x_{k|k}). \quad (9)$$

Note that the model predictive control problem \mathcal{MPC} involves an optimization over *time-varying closed-loop policies*, as opposed to the classical deterministic case where the optimization is over open-loop sequences. A similar approach is taken in [8], [4]. We will show in Section VIII how to solve problem \mathcal{MPC} efficiently.

The following theorem shows that the model predictive control law (9), with a proper choice of the terminal weight P , is risk-sensitive stabilizing, i.e., the closed-loop system (3) is UGRSES.

Theorem VII.1. (Stochastic Stability for Model Predictive Control Law) Consider the model predictive control law in equation (9) and the corresponding closed-loop dynamics for system (3) with initial condition $x_0 \in \mathbb{R}^{N_x}$. Suppose that $P = P^T \succ 0$, and there exists a matrix F such that:

$$\sum_{j=1}^L q_l(j) (A_j + B_j F)^T P (A_j + B_j F) - P + (F^T R F + Q) \prec 0, \quad (10)$$

for all $l \in \{1, \dots, \text{cardinality}(\mathcal{U}^{\text{poly}, V}(p))\}$, where $\mathcal{U}^{\text{poly}, V}(p)$ is the set of vertices of polytope $\mathcal{U}^{\text{poly}}(p)$, q_l is the l th element in set $\mathcal{U}^{\text{poly}, V}(p)$, and $q_l(j)$ denotes the j th component of vector q_l , $j \in \{1, \dots, L\}$. Then, the closed loop system (3) is UGRSES.

Proof. The strategy of this proof is to show that J_k^* is a valid Lyapunov function in the sense of Lemma VI.1. Specifically, we want to show that J_k^* satisfies the two inequalities in equation (7); the claim then follows by simply noting that, in our time-invariant setup, J_k^* does not depend on k .

We start by focusing on the bottom inequality in equation (7). Consider a time k and an initial condition $x_{k|k} \in \mathbb{R}^{N_x}$ for problem \mathcal{MPC} . The sequence of optimal control policies is given by $\{\pi_{k+h|k}^*\}_{h=0}^{N-1}$. Let us define a sequence of control policies from time $k+1$ to N according to

$$\pi_{k+h|k+1}(x_{k+h|k}) := \begin{cases} \pi_{k+h|k}^*(x_{k+h|k}) & \text{if } h \in [1, N-1], \\ F x_{k+N|k} & \text{if } h = N. \end{cases} \quad (11)$$

This sequence of control policies is essentially the concatenation of the sequence $\{\pi_{k+h|k}^*\}_{h=1}^{N-1}$ with a linear feedback control law for stage N (the reason why we refer to this policy with the subscript “ $k+h|k+1$ ” is that we will use this policy as a feasible policy for problem \mathcal{MPC} starting at stage $k+1$).

Consider the \mathcal{MPC} problem at stage $k+1$ with initial condition given by $x_{k+1|k+1} = A(w_k)x_{k|k} + B(w_k)\pi_{k|k}^*(x_{k|k})$, and denote with $\bar{J}_{k+1}(x_{k+1|k+1})$ the cost of the objective function assuming that the sequence of control policies is given by $\{\pi_{k+h|k+1}^*\}_{h=1}^N$. Note that $x_{k+1|k+1}$ (and therefore $\bar{J}_{k+1}(x_{k+1|k+1})$) is a random variable with L possible realizations. Define:

$$\begin{aligned} Z_{k+N} &:= -x_{k+N|k}^T P x_{k+N|k} + x_{k+N|k}^T Q x_{k+N|k} \\ &\quad + (F x_{k+N|k})^T R F x_{k+N|k}, \\ Z_{k+N+1} &:= \left((A(w_{k+N|k}) + B(w_{k+N|k})F)x_{k+N|k} \right)^T \\ &\quad P \left((A(w_{k+N|k}) + B(w_{k+N|k})F)x_{k+N|k} \right). \end{aligned}$$

By exploiting the dual representation of Markov polytopic risk metrics, one can easily show that equation (10) implies

$$Z_{k+N} + \rho(Z_{k+N+1}) \leq 0. \quad (12)$$

One can then write the following chain of inequalities:

$$\begin{aligned} J_k^*(x_{k|k}) &= x_{k|k}^T Q x_{k|k} + (\pi_{k|k}^*(x_{k|k}))^T R \pi_{k|k}^*(x_{k|k}) + \\ &\quad \rho \left(\rho_{k+1, N} \left(c(x_{k+1|k}, \pi_{k+1}^*(x_{k+1|k})), \dots, x_{k+N|k}^T Q x_{k+N|k} + \right. \right. \\ &\quad \left. \left. (F x_{k+N|k})^T R F x_{k+N|k} + \rho(Z_{k+N+1}) - Z_{k+N} - \rho(Z_{k+N+1}) \right) \right) \geq \\ &\quad x_{k|k}^T Q x_{k|k} + (\pi_{k|k}^*(x_{k|k}))^T R \pi_{k|k}^*(x_{k|k}) + \\ &\quad \rho \left(\rho_{k+1, N} \left(c(x_{k+1|k}, \pi_{k+1}^*(x_{k+1|k})), \dots, x_{k+N|k}^T Q x_{k+N|k} + \right. \right. \\ &\quad \left. \left. (F x_{k+N|k})^T R F x_{k+N|k} + \rho(Z_{k+N+1}) \right) \right) = \\ &\quad x_{k|k}^T Q x_{k|k} + (\pi_{k|k}^*(x_{k|k}))^T R \pi_{k|k}^*(x_{k|k}) + \rho(\bar{J}_{k+1}(x_{k+1|k+1})) \geq \\ &\quad x_{k|k}^T Q x_{k|k} + (\pi_{k|k}^*(x_{k|k}))^T R \pi_{k|k}^*(x_{k|k}) + \rho(J_{k+1}^*(x_{k+1|k+1})), \end{aligned} \quad (13)$$

where the first equality follows from our definitions, the second inequality follows from equation (12) and the monotonicity property of Markov polytopic risk metrics (see also [15, Page 242]), the third equality follows from the fact that the sequence of control policies $\{\pi_{k+h|k+1}^*\}_{h=1}^N$ is a feasible sequence for the \mathcal{MPC} problem starting at stage $k+1$ with initial condition $x_{k+1|k+1} = A(w_k)x_{k|k} + B(w_k)\pi_{k|k}^*(x_{k|k})$, and the fourth inequality follows from the definition of J_{k+1}^* and the monotonicity of Markov polytopic risk metrics.

We now turn our focus to the top inequality in equation (7). One can easily bound $J_k^*(x_{k|k})$ from below according to:

$$J_k^*(x_{k|k}) \geq x_{k|k}^T Q x_{k|k} \geq \lambda_{\min}(Q) \|x_{k|k}\|^2, \quad (14)$$

where $\lambda_{\min}(Q) > 0$ by assumption. To bound $J_k^*(x_{k|k})$ from above, define:

$$M_A := \max_{r \in \{0, \dots, N-1\}} \max_{j_0, \dots, j_r \in \{1, \dots, L\}} \|A_{j_r} \dots A_{j_1} A_{j_0}\|_2.$$

One can write:

$$\begin{aligned} J_k^*(x_{k|k}) &\leq c(x_{k|k}, 0) + \rho(c(x_{k+1|k}, 0) + \rho(c(x_{k+2|k}, 0) \\ &\quad + \dots + \rho(x_{k+N|k}^T P x_{k+N|k}) \dots)) \\ &\leq \|Q\|_2 \|x_{k|k}\|_2^2 + \rho(\|Q\|_2 \|x_{k+1|k}\|_2^2 + \rho(\|Q\|_2 \|x_{k+2|k}\|_2^2 + \\ &\quad \dots + \rho(\|P\|_2 \|x_{k+N}\|_2^2) \dots)). \end{aligned}$$

Therefore, by using the translational invariance and monotonicity property of Markov polytopic risk metrics, one obtains the upper bound:

$$0 \leq J_k^*(x_{k|k}) \leq (N \|Q\|_2 + \|P\|_2) M_A \|x_{k|k}\|_2^2. \quad (15)$$

Combining the results in equations (13), (14), (15), and given the time-invariance of our problem setup, one concludes that $J_k^*(x_{k|k})$ is a ‘‘risk-sensitive’’ Lyapunov function for the closed-loop system (3), in the sense of Lemma VI.1. This concludes the proof. \square

The study of the conservatism of the LMI condition (10) is left for future work. We note here that out of 100 randomly generated 5-state-2-action-3-scenario examples, 83 of them satisfied condition (10).

VIII. SOLUTION ALGORITHMS

In this section we discuss two solution approaches, the first one via dynamic programming, the second one via convex programming.

A. Dynamic programming approach

Problem MPC can be solved via dynamic programming, see [15, Theorem 2]. However, one would first need to find a matrix P that satisfies equation (10). The next theorem presents a linear matrix inequality characterization of condition (10).

Theorem VIII.1. Define $\bar{A} = [A_1^T \dots A_L^T]^T$, $\bar{B} = [B_1^T \dots B_L^T]^T$ and $\Sigma_l = \text{diag}(q_l(1)I, \dots, q_l(L)I) \succeq 0$, for all $q_l \in \mathcal{U}^{\text{poly}, V}(p)$. Consider the following set of linear matrix inequalities with respect to decision variables Y , G , $\bar{Q} = \bar{Q}^T \succ 0$:

$$\begin{bmatrix} I_{L \times L} \otimes \bar{Q} & 0 & 0 & -\Sigma_l^{\frac{1}{2}}(\bar{A}G + \bar{B}Y) \\ * & R^{-1} & 0 & -Y \\ * & * & I & -Q^{\frac{1}{2}}G \\ * & * & * & -\bar{Q} + G + G^T \end{bmatrix} \succ 0, \quad (16)$$

where $l \in \{1, \dots, \text{cardinality}(\mathcal{U}^{\text{poly}, V}(p))\}$. The set of linear matrix inequalities in (16) is equivalent to the condition in (10) by setting $F = YG^{-1}$ and $P = \bar{Q}^{-1}$.

Proof. The theorem can be proven by using the Projection Lemma (see [26, Chapter 2]). The details are omitted in the interest of brevity. \square

Hence, a solution approach for MPC is to first solve the linear matrix inequality in Theorem VIII.1 and then, if a solution for P is found, apply dynamic programming (after state and action *discretization*, see, e.g., [27], [28]). Note that the discretization process might yield a large-scale dynamic programming problem, which motivates the convex programming approach presented next.

B. Convex programming approach

Consider the following parameterization of *history-dependent* control policies. Let $j_0, \dots, j_h \in \{1, \dots, L\}$ be the realized indices for the disturbances in the first $h+1$ stages of the MPC problem, where $h \in \{0, \dots, N-1\}$; for $h \geq 1$, we will refer to the control to be exerted at stage h as $\bar{U}_h(j_0, \dots, j_{h-1})$. Similarly, we will refer to the state at stage h as $\bar{X}_h(j_0, \dots, j_{h-1})$. The quantities $\bar{X}_h(j_0, \dots, j_{h-1})$ and $\bar{U}_h(j_0, \dots, j_{h-1})$ enable us to keep track of the (exponential) growth of the scenario tree. In terms of this new notation, the system dynamics (3) can be rewritten according to:

$$\begin{aligned} \bar{X}_0 &:= x_{k|k}, \bar{U}_0 \in \mathbb{R}^{N_u}, \\ \bar{X}_1(j_0) &= A_{j_0} \bar{X}_0 + B_{j_0} \bar{U}_0, \text{ for } h = 1, \\ \bar{X}_h(j_0, \dots, j_{h-1}) &= A_{j_{h-1}} \bar{X}_{h-1}(j_0, \dots, j_{h-2}) + \\ &\quad B_{j_{h-1}} \bar{U}_{h-1}(j_0, \dots, j_{h-2}), \text{ for } h \geq 2. \end{aligned} \quad (17)$$

Indeed, the MPC problem is defined as an optimization problem over *Markov* control policies. However, in the convex programming approach, we re-define the MPC problem as an optimization problem over *history-dependent* policies. One can show (with a virtually identical proof) that the stability Theorem VIII.1 still holds when history-dependent policies are considered. Furthermore, since Markov policies are optimal in our setup (see [15, Theorem 2]), the value of the optimal cost stays the same. The key advantage of history-dependent policies is that their additional flexibility leads to a convex optimization problem for the determination of the model predictive control law. Specifically, the model predictive control law can be determined according to the following algorithm.

Solution for MPC — Given an initial state $x_{k|k} = x_k$ and a prediction horizon $N \geq 1$, solve:

$$\begin{aligned} \min & \gamma_1 \\ & \gamma_1, W, G_{j_N}, Y_{j_N}, \bar{Q}, \gamma_2(j_0, \dots, j_{N-1}) \\ & \bar{U}_0, \bar{U}_h(j_0, \dots, j_{h-1}), h \in \{1, \dots, N\} \\ & \bar{X}_h(j_0, \dots, j_{h-1}), h \in \{1, \dots, N\} \\ & j_0, \dots, j_{N-1} \in \{1, \dots, L\} \end{aligned}$$

subject to

- The linear matrix inequality in equation (16).
- The linear matrix inequality

$$\begin{bmatrix} \gamma_2(j_0, \dots, j_{N-1})I & \bar{X}_N(j_0, \dots, j_{N-1})^T \\ * & \bar{Q} \end{bmatrix} \succeq 0.$$

- The system dynamics in equation (17).
- The objective epigraph constraint:

$$\begin{aligned} \rho_{k, k+N}(c_0(\bar{X}_0, \bar{U}_0), \dots, c_{N-1}(\bar{X}_{N-1}, \bar{U}_{N-1}), \\ \gamma_2(j_0, \dots, j_{N-1})) \leq \gamma_1. \end{aligned}$$

Then, set $\pi^{MPC}(x_k) = \bar{U}_0$.

The proof of the correctness of this algorithm is omitted due to lack of space and will be presented in the journal version of this paper. This algorithm is clearly suitable only for ‘‘moderate’’ values of N , given the combinatorial explosion of the scenario tree. Indeed, one could devise a *hybrid* algorithm where the above computation is split into two steps: (1) the terminal cost function is computed offline, and (2) the MPC control law is computed online. We tested this approach and found that the computation time is very short for reasonably large problems. The details are deferred to a future publication.

TABLE I
STATISTICS FOR RISK-AVERSE MPC.

Level of risk	Sample Mean	Sample Dispersion	Sample Standard Deviation	Mean Time Per Iter. (Sec)
c=0	2.9998	0.2889	0.4245	4.8861
c=0.25	3.3012	0.2643	0.3520	5.2003
c=0.5	3.4178	0.2004	0.2977	4.4007
c=0.75	3.5898	0.1601	0.2231	4.4577
c=1	3.6072	0.0903	0.1335	4.6498

IX. NUMERICAL EXPERIMENTS

In this section we present numerical experiments run on a 2.3 GHz Intel Core i5, MacBook Pro laptop, using the MATLAB YALMIP Toolbox, version 2.6.3 [29]. Consider the stochastic system: $x_{k+1} = A(w_k)x_k + B(w_k)u_k$, where $w_k \in \{1, 2, 3\}$ and

$$A_1 = \begin{bmatrix} 2 & 0.5 \\ -0.5 & 2 \end{bmatrix}, A_2 = \begin{bmatrix} 0.01 & 0.1 \\ 0.05 & 0.01 \end{bmatrix}, A_3 = \begin{bmatrix} 1.5 & -0.3 \\ 0.2 & 1.5 \end{bmatrix},$$

$$B_1 = \begin{bmatrix} 3 & 0.1 \\ 0.1 & 3 \end{bmatrix}, B_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, B_3 = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 2 \end{bmatrix}.$$

The probability mass function for the process disturbance is uniformly distributed, i.e., $\mathbb{P}(w_k = i) = \frac{1}{3}$, for $i \in \{1, 2, 3\}$. In this example, we want to explore the risk aversion capability of the risk-averse MPC algorithm presented in Section VII (the solution relies on the convex programming approach). We consider as risk-aversion metric the mean upper semi-deviation metric (see Section II), where c ranges in the set $\{0, 0.25, 0.5, 0.75, 1\}$. The initial condition is $x_0(1) = x_0(2) = 1$ and the number of lookahead steps is 3. We set $Q = 1 \times I_{2 \times 2}$ and $R = 10^{-4}I_{2 \times 2}$. We perform 100 Monte Carlo simulations for each value of c . When $c = 0$, the problem is reduced to a risk-neutral minimization problem. On the other hand, one enforces maximum emphasis on regulating semi-deviation (dispersion) by setting $c = 1$. Table I summarizes our results (computation times are given in seconds). When $c = 0$ (risk neutral formulation), the average cost is the lowest (with respect to the different choices for c), but the dispersion is the largest. Conversely, when $c = 1$, the dispersion is the lowest, but the average cost is the largest.

X. CONCLUSION AND FUTURE WORK

In this paper we presented a framework for risk-averse MPC. Advantages of this framework include: (1) it is axiomatically justified; (2) it is amenable to dynamic and convex optimization; and (3) it is general, in the sense that it captures a full range of risk assessments from risk-neutral to worst case (given the generality of Markov polytopic risk metrics).

This paper leaves numerous important extensions open for further research. First, we plan to study the case with state and control constraints (preliminary work suggests that such constraints can be readily included in our framework). Second, it is of interest to consider the inclusion of multi-period risk constraints. Third, we plan to characterize the sub-optimal performance of risk-averse MPC by designing lower and upper bounds for the optimal cost. Fourth, we plan to study in detail the offline and hybrid counterparts of the online MPC algorithm presented in this paper. Finally, it is of interest to study the robustness of our approach with respect to perturbations in the system's parameters.

REFERENCES

- [1] J. Qin and T. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7):733–764, 2003.
- [2] Y. Wang and S. Boyd. Fast model predictive control using online optimization. *IEEE Transactions on Control Systems Technology*, 18(2):267–278, 2010.
- [3] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: stability and optimality. *Automatica*, 36(6):789–814, 2000.
- [4] D. Bernardini and A. Bemporad. Stabilizing model predictive control of stochastic constrained linear systems. *IEEE Transactions on Automatic Control*, 57(6):1468–1480, 2012.
- [5] M. Kothare, V. Balakrishnan, and M. Morari. Robust constrained model predictive control using Linear Matrix Inequalities. *Automatica*, 32(10):1361–1379, 1996.
- [6] C. de Souza. Robust stability and stabilization of uncertain discrete-time Markovian jump linear systems. *IEEE Transactions on Automatic Control*, 51(5):836–841, 2006.
- [7] B. Park and W. Kwon. Robust one-step receding horizon control of discrete-time Markovian jump uncertain systems. *Automatica*, 38(7):1229–1235, 2002.
- [8] J. Primbs A and C. Sung. Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise. *IEEE Transactions on Automatic Control*, 54(2):221–230, 2009.
- [9] B. Defourny, D. Ernst, and L. Wehenkel. Risk-aware decision making and dynamic programming. In *NIPS Workshop on Model Uncertainty and Risk in Reinforcement Learning*, 2008.
- [10] T. M. Moldovan and P. Abbeel. Risk aversion in Markov decision processes via near optimal Chernoff bounds. In *NIPS*, pages 3140–3148, 2012.
- [11] D. A. Iancu, M. Petrik, and D. Subramanian. Tight approximations of dynamic risk measures. 2013. Submitted, available at <http://arxiv.org/abs/1106.6102>.
- [12] A. Zafra-Cabeza, J.M. Maestre, M. A. Ridao, E. F. Camacho, and L. Sánchez. A hierarchical distributed model predictive control approach to irrigation canals: a risk mitigation perspective. *Journal of Process Control*, 21(5):787 – 799, 2011.
- [13] P.-J. van Overloop, S. Weijts, and S. Dijkstra. Multiple model predictive control on a drainage canal system. *Control Engineering Practice*, 16(5):531 – 540, 2008.
- [14] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of operations research*, 21(3):544–561, 2006.
- [15] A. Ruszczyński. Risk averse dynamic programming for Markov decision process. *Journal of Mathematical Programming*, 125(2):235–261, 2010.
- [16] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1998.
- [17] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.
- [18] P. Cheridito and M. Stadje. Time inconsistency of VaR and time-consistent alternatives. *Finance Research Letters*, 6(1):40–46, 2009.
- [19] A. Shapiro. On a time consistency concept in risk averse multi-stage stochastic programming. *Operations Research Letters*, 37(3):143–147, 2009.
- [20] B. Roorda, J. M. Schumacher, and J. Engwerda. Coherent acceptability measures in multi-period models. *Mathematical Finance*, 15(4):589–612, 2005.
- [21] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16(1):69–95, 2005.
- [22] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: semi-deviations as risk measures. *European Journal of Operational Research*, 116(1):33–50, 1999.
- [23] R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
- [24] D. Bertsimas and D. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.
- [25] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: the optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.
- [26] R. Skelton, T. Iwasaki, and K. Grigoriadis. *A unified algebraic approach to linear control design*. CRC Press, 1998.
- [27] Y. Chow and M. Pavone. A uniform-grid discretization algorithm for dynamic programming with risk constraints. In *Conference on Decision and Control*, 2013.
- [28] C. Chow and J. Tsitsiklis. An optimal one-way multigrid algorithm for discrete-time stochastic control. *Automatic Control, IEEE Transactions on*, 36(8):898–914, 1991.
- [29] J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *CCA/ISIC/CACSD*, September 2004.