Stochastic Optimal Control via Local Occupation Measures

Flemming Holtorf\textsuperscript{1,2}, Alan Edelman\textsuperscript{1,2,3}, and Christopher Rackauckas\textsuperscript{1,2,4,5}

\textsuperscript{1}Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, 02139 Cambridge, MA
\textsuperscript{2}Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Ave, 02139 Cambridge, MA
\textsuperscript{3}Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Ave, 02139 Cambridge, MA
\textsuperscript{4}JuliaHub, Alewife, MA, USA
\textsuperscript{5}Pumas-AI, Centreville, VA, USA

February 29, 2024

Abstract

Viewing stochastic processes through the lens of occupation measures has proved to be a powerful angle of attack for the theoretical and computational analysis of a wide range of stochastic optimal control problems. We present a simple modification of the traditional occupation measure framework derived from resolving the occupation measures locally on a partition of the control problem’s space-time domain. This notion of local occupation measures provides fine-grained control over the construction of structured semidefinite programming relaxations for a rich class of stochastic optimal control problems with embedded diffusion and jump processes via the moment-sum-of-squares hierarchy. As such, it bridges the gap between discretization-based approximations to the solution of the Hamilton-Jacobi-Bellmann equations and approaches based on convex optimization and the moment-sum-of-squares hierarchy. We demonstrate with examples that this approach enables the computation of high quality bounds on the optimal value for a large class of stochastic optimal control problems with notable performance gains relative to the traditional occupation measure framework.

1 Introduction

The optimal control of stochastic processes is one of the archetypical problems of decision-making under uncertainty with a myriad of applications in science and engineering. Despite their ubiquity, however, only a small subset of such stochastic optimal control problems admits the computation of a globally optimal control policy in a tractable and certifiable manner. As a consequence, engineers are often forced to resort to one of many available heuristics for the design of control policies in practice. And although such heuristics often perform remarkably well, they seldom come with a simple mechanism to quantify rigorously the degree

\textsuperscript{*}E-mail: holtorf@mit.edu
\textsuperscript{†}E-mail: edelman@mit.edu
\textsuperscript{‡}E-mail: crackauc@mit.edu
of suboptimality they introduce, ultimately leaving it to the engineer’s intuition when the controller design process shall be terminated.

In response to this undesirable situation, the task of computing theoretically guaranteed yet informative bounds gauging the best attainable performance for various classes of stochastic optimal control and related problems has received considerable attention in the recent past; contributions range from bounding schemes for the optimal control of systems described by deterministic nonlinear ordinary 

[Las+08; HLS08; GQ09; Hen+23b] and partial differential equations 

[KHL18; KHL22; Hen+23a] over discrete-time Markov control problems 

[Her+99; SLD09] to the control of diffusion and other continuous-time stochastic processes 

[HS00; CS02; LGS18; HJV21; Hol+23]. In particular the framework of occupation measures has proved to be a versatile and effective approach to this task. The notion of occupation measures allows for the translation of a rich class of stochastic optimal control problems into infinite-dimensional linear programs over Borel measure spaces 

[FV89; Vin93; BB96; KS98], for which a sequence of increasingly tight, tractable semidefinite programming (SDP) relaxations is readily constructed via the moment-sum-of-squares (MSOS) hierarchy 

[Las01; Par00]. A key limitation of this framework, however, remains in its poor scalability. Specifically, the problem size of the SDP relaxations grows combinatorially with the hierarchy level and often high levels are necessary to establish informative bounds in practice. The notorious numerical ill-conditioning of moment problems involving high-order moments further exacerbates this limitation.

In this paper, we set out to improve the practicality of the occupation measure approach to stochastic optimal control by proposing a simple modification of the traditional framework. To that end, we introduce a localized notion of occupation measures based on partitioning of the state space of the controlled process and the control horizon. Analogous to its traditional counterpart, the resultant local occupation measure framework enables the construction of SDP relaxations for a large class of stochastic optimal control problems via the moment-sum-of-squares hierarchy. In contrast to the traditional approach, however, the resultant relaxations can be tightened without increasing the hierarchy level, but instead by simply refining the spatio-temporal partition of the problem domain. Such a “tightening-by-refinement” provides two major practical advantages:

1. It avoids numerical ill-conditioning originating from high-order moments which in practice often prohibits the accurate solution of SDP relaxations furnished by high levels of the MSOS hierarchy.

2. It provides more fine-grained and easily interpretable control over tightening of the SDP relaxations when compared to increasing the level in the MSOS hierarchy.

As we demonstrate with examples, these advantages hold the potential to construct equally or even tighter relaxations that can be solved notably faster than those derived with the traditional approach. Another potential advantage worth mentioning yet beyond the scope of this work is that the proposed approach is similar in spirit to a wide range of numerical approximation techniques for the solution of partial differential equations (PDEs); as such, the resultant moment-sum-of-squares relaxations exhibit a benign, weakly-coupled block structure akin that of discretized PDEs which may be exploited, for example by distributed optimization techniques 

[SZA20; Boy+10].

A partitioning approach closely related to the here proposed local occupation measure framework has recently been studied by Cibulka et al. [CKH21] in the context of approximating the region of attraction for deterministic control systems via sum-of-squares programming. In another related work, Holtorf and Barton [HB24] have used temporal partitioning in order to improve MSOS bounding schemes for trajectories of stochastic chemical systems modeled by jump processes. Both works report significant computational merits of the respective modifications. Here, we unify and extend both works by introducing the notion of local occupation measures which applies beyond deterministic control problems to jump and diffusion control problems alike. The resulting framework is independent from and can be complemented by other approaches aimed at improving the tractability and practicality of the MSOS hierarchy, such as symmetry
reduction [Rie+13; Aug+23], sparsity exploitation [SK20; Wan+21; ZFP19], and linear/second-order cone programming hierarchies [AM14; ADH17; AH19; AM19].

The remainder of this article is structured as follows: In Section 2, we review the concept of occupation measures and show how it enables the construction of tractable convex relaxations for a large class of stochastic optimal control problems with embedded diffusion processes. In Sections 3 and 4 we introduce the notion of local occupation measures and study its interpretation in the context of stochastic optimal control from the dual (polynomial) and primal (moment) perspective, respectively. Section 5 is dedicated to highlight the advantages of the proposed framework for the construction of high quality relaxations with regard to the scaling properties and structure of the resultant optimization problems. In Section 6 we showcase the potential of the proposed approach with an example problem from population control. In Section 7 we discuss the extension of the described local occupation measure framework to discounted infinite horizon control problems as well as the control of jump processes, supported with an example from systems biology. We conclude with some final remarks in Section 8.

2 Problem description & preliminaries

We consider a continuous-time diffusion process \( x_t \) in \( \mathbb{R}^n \) driven by a standard \( \mathbb{R}^m \)-Brownian motion \( B_t \) and controlled by a non-anticipative control process \( u_t \) in \( \mathbb{R}^n \),

\[
dx_t = f(x_t, u_t) \, dt + g(x_t, u_t) \, dB_t, \tag{1}
\]

and study the associated finite horizon optimal control problem

\[
J := \inf_{u_t} \mathbb{E}_{\nu_0} \left[ \int_{[0,T]} \ell(x_t, u_t) \, dt + \phi(x_T) \right] \quad \text{(OCP)}
\]

s.t. \( x_t \) satisfies (1) on \([0, T]\) with \( x_0 \sim \nu_0 \),

\( (x_t, u_t) \in X \times U \) on \([0, T]\),

\( u_t \) is non-anticipative.

Here, \( \mathbb{E}_{\nu_0} \) denotes the expectation with respect to the probability measure \( \mathbb{P}_{\nu_0} \) over the paths of the diffusion process (1). The subscript \( \nu_0 \) indicates the dependence on the distribution of the initial state, which we assume to be known. Throughout, we further assume that all problem data is described in terms of polynomials in the following sense.

**Assumption 1.** The drift coefficient \( f : X \times U \to \mathbb{R}^n \), diffusion matrix \( g g^\top : X \times U \to \mathbb{R}^{n \times n} \), stage cost \( l : X \times U \to \mathbb{R} \) and terminal cost \( \phi : X \times U \to \mathbb{R} \) are componentwise polynomial functions jointly in both arguments. The state space \( X \) and the set of admissible control inputs \( U \) are basic closed semialgebraic sets.

We say a control process \( u_t \) is admissible if the the controlled process \( (x_t, u_t) \) satisfies the constraints in Problem (OCP). Furthermore, we make the following well-posedness assumption that ensures that the optimal value of (OCP) is finite.

**Assumption 2.** The controlled diffusion process (1) has finite moments for any admissible control process, i.e., \( \mathbb{E}_{\nu_0}[p(x_t, u_t)] \) is finite for all polynomials \( p \) and \( t \in [0, T] \).

Note that this assumption does not impose strong practical restrictions as it is for instance implied if the distribution of the controlled process has exponentially decaying tails or if \( X \) and \( U \) are compact.

The key insight enabling the construction of convex relaxations of (OCP) is that the controlled process described by (1) admits a weak-form characterization in terms of a pair of occupations measures: the
instantaneous and expected state-action occupation measure \([FV89; KS98; BB96]\). This characterization endows the control problem with a convex, albeit infinite-dimensional, geometry, sidestepping the nonlinear dependence of the paths of the diffusion process \([1]\) on the control process.

The instantaneous occupation measure \(\nu\) is given by the probability to observe \(x_T\) in any Borel set \(B \subset X\). Formally, we define
\[
\nu(B) := P_{\nu_0}[x_T \in B].
\]
or equivalently,
\[
\langle w, \nu \rangle := E_{\nu_0}[w(T, x_T)]
\]
for every continuous test function \(w \in C([0, T] \times X)\), where
\[
\langle w, \nu \rangle := \int_X w(T, x) \, d\nu(x)
\]
denotes the standard duality bracket between continuous functions and finite measures.

The expected state-action occupation measure \(\xi\) is defined as the average time the controlled process \((t, x_t, u_t)\) remains in a Borel subset of \([0, T] \times X \times U\); formally, we define
\[
\xi(B_T \times B_X \times B_U) := E_{\nu_0} \left[ \int_{[0, T] \cap B_T} \mathbb{1}_{B_X \times B_U}((x_t, u_t)) \, dt \right]
\]
for any Borel subsets \(B_T \subset [0, T]\), \(B_X \subset X\), \(B_U \subset U\); or equivalently,
\[
\langle w, \xi \rangle := E_{\nu_0} \left[ \int_{[0, T]} w(t, x_t, u_t) \, dt \right]
\]
for any continuous test function \(w \in C([0, T] \times X \times U)\). The instantaneous and expected state-action occupation measures are finite, non-negative measures by construction.

The occupation measure pair \((\nu, \xi)\) characterizes the expected time evolution of sufficiently smooth observables \(w \in C^{1,2}([0, T] \times X)\) of the process \([1]\) by Dynkin’s formula \([OS07, Theorem 1.24]\),
\[
E_{\nu_0}[w(T, x_T)] = E_{\nu_0}[w(0, x_0)] + E_{\nu_0} \left[ \int_{[0, T]} A w(s, x_s, u_s) \, ds \right],
\]
or equivalently,
\[
\langle w, \nu \rangle = \langle w, \nu_0 \rangle + \langle Aw, \xi \rangle,
\]
where \(A : C^{1,2}([0, T] \times X) \to C([0, T] \times X \times U)\) denotes the (extended) infinitesimal generator of the diffusion process \([1]\) \([OS07]\), i.e.,
\[
A : w(t, x) \mapsto \frac{\partial w}{\partial t}(t, x) + f(x, u)^\top \nabla_x w(t, x) + \frac{1}{2} \text{Tr} \left( g g^\top(x, u) \nabla^2_x w(t, x) \right).
\]

Conversely, we say that a measure pair \((\nu, \xi)\) is a weak solution to \([1]\) on the interval \([0, T]\) if it satisfies Equation \((2)\) for all test functions \(w \in C^{1,2}([0, T] \times X)\). This notion of weak solutions to \([1]\) motivates the
\footnote{\noindent That is functions on the domain \([0, T] \times X\) with continuous first and second derivatives (in the sense of Whitney \([Whi92]\)) in the first and second argument, respectively.}
following weak form of (OCP) [FV89]:
\[ J^* := \inf_{\nu, \xi} \langle \ell, \xi \rangle + \langle \phi, \nu \rangle \]  
(weak OCP)
\[ \text{s.t. } \langle w, \nu \rangle - \langle Aw, \xi \rangle = \langle w, \nu_0 \rangle, \quad w \in C^{1,2}([0, T] \times X), \]
\[ \nu \in \mathcal{M}_+(X), \]
\[ \xi \in \mathcal{M}_+([0, T] \times X \times U). \]

where \( \mathcal{M}_+(Y) \) denotes the cone of finite, positive Borel measures supported on the set \( Y \). Problem (weak OCP) is an infinite-dimensional linear program [NA87] and generally a relaxation of (OCP), albeit conditions for their equivalence can be established (see for example [BB96, Section 4]).

From a practical perspective, (weak OCP) remains intractable as an infinite dimensional linear program; however, Assumption 1 enables the construction of a sequence of increasingly tight SDP relaxations via the MSOS hierarchy [Par00; Las01]. To that end, (weak OCP) is relaxed to the optimization over moment sequences of the measures \( \nu \) and \( \xi \) truncated at finite order \( d \). For polynomial test functions, constraints of the form (2) reduce to affine constraints on the moment sequence as \( A \) maps polynomials to polynomials under Assumption 1. Similarly, the conic constraints \( \nu \in \mathcal{M}_+(X) \) and \( \xi \in \mathcal{M}_+([0, T] \times X \times U) \) can be relaxed to positive semidefiniteness constraints of certain moment and localizing matrices, which under Assumption 1 reduce to linear matrix inequalities [Las01].

The infinite-dimensional linear programming dual [NA87] to (weak OCP) has an informative interpretation that serves as motivation for the partitioning strategy presented in the next section. The dual reads
\[ \sup_w \int_X w(0, x) d\nu_0(x) \]  
(subHJB)
\[ \text{s.t. } Aw + \ell \leq 0, \text{ on } [0, T] \times X \times U, \]
\[ w(T, \cdot) \leq \phi, \text{ on } X, \]
\[ w \in C^{1,2}([0, T] \times X), \]
where the decision variable \( w \) can be interpreted as a smooth underestimator of the value function associated with the control problem (OCP):

**Corollary 1.** Let \( w \) be feasible for (subHJB) and let \( \delta_z \) denote the Dirac measure centered at \( z \). Then, \( w \) underestimates the value function
\[ V(t, z) := \inf_{u_s} \mathbb{E}_{\delta_z} \left[ \int_t^T \ell(x_s, u_s) ds + \phi(x_T) \right] \]  
(5)
\[ \text{s.t. } x_s \text{ satisfies (1) on } [t, T] \text{ with } x_t \sim \delta_z, \]
\[ (x_s, u_s) \in X \times U \text{ on } [t, T], \]
\[ u_s \text{ is non-anticipative.} \]

for any \((t, z) \in [0, T] \times X\).

**Proof.** Let \( z \in X \) and \( 0 \leq t \leq T \) and fix any admissible control policy \( u_s \), i.e., a control policy such that the path of the stochastic process \((x_s, u_s)\) remains in \( X \times U \) on \([t, T]\). Then, Constraints (3) and (4) imply that
\[ \mathbb{E}_{\delta_z} \left[ -\int_t^T A w(s, x_s, u_s) ds + w(T, x_T) \right] \leq \mathbb{E}_{\delta_z} \left[ \int_t^T \ell(x_s, u_s) ds + \phi(x_T) \right]. \]

The left-hand-side coincides with \( w(t, z) \) by Dynkin’s formula. The result follows by minimizing over all admissible control policies. \( \square \)
Analogous to its primal counterpart, the MSOS hierarchy gives rise to a sequence of increasingly tight SDP restrictions of \((\text{subHJB})\) by restricting \(w\) to be a polynomial of degree at most \(d\) and imposing the non-negativity constraints by means of sufficient sum-of-squares conditions \([\text{Las01; Par00}]\). The restriction is weakened by increasing the degree \(d\) yielding a monotonically increasing sequence of lower bounds for the optimal value \(J^*\) of \((\text{weak OCP})\). The following theorem establishes a set of easily verifiable conditions under which this sequence converges from below to \(J^*\) (implying also strong duality between \((\text{subHJB})\) and \((\text{weak OCP})\)).

**Theorem 1.** Let \(J_d\) be the optimal value of the \(d^{th}\) level MSOS restriction of \((\text{subHJB})\) (resp. relaxation of \((\text{weak OCP})\)). If Assumption 1 holds and moreover \(X\) and \(U\) are represented as

\[
X = \{ x : p_i(x) \geq 0, \ i = 1, \ldots, v, \ R_X - \|x\|^2 \geq 0 \},
\]

\[
U = \{ u : q_i(x) \geq 0, \ i = 1, \ldots, w, \ R_U - \|u\|^2 \geq 0 \},
\]

with suitable polynomials \(p_i\) and \(q_i\), and sufficiently large \(R_X\) and \(R_U\), then \(J_d \uparrow J^*\).

**Proof.** First note that under the given assumptions, the set \([0, T] \times X \times U\) is compact. Thus, it suffices to impose condition (2) for all polynomial test functions in \((\text{weak OCP})\) as it is a dense subset of \(C^1([0, T] \times X)\). Further observe that constraint (2) implies that every feasible pair \((\nu, \xi)\) has constant mass. Specifically, for test functions \(w(t, x) \equiv 1\) and \(w(t, x) = t\), constraint (2) reduces to \(\langle 1, \nu \rangle = 1\) and \(\langle 1, \xi \rangle = T\), respectively. The result thus follows from \([\text{Tac22, Corollary 8}]\). \(\square\)

**Remark 1.** The condition imposed by Theorem 1 on the representation of \(X\) and \(U\) is only marginally stronger than imposing their compactness in addition to Assumption 1. If \(X\) and \(U\) are compact basic closed semialgebraic sets, one can always add redundant ball constraints \(R_X - \|x\|^2 \geq 0\) and \(R_U - \|u\|^2 \geq 0\) to their description to satisfy the condition in Theorem 1.

3 The dual perspective revisited: piecewise polynomial approximation

In order to construct improved approximations to the value function in the spirit of \((\text{subHJB})\), we consider a generalization of problem \((\text{subHJB})\) that seeks a piecewise smooth underapproximation of the value function over the problem’s space-time domain \([0, T] \times X\). To that end, we consider a discretization \(0 = t_0 < t_1 < \cdots < t_n = T\) of the control horizon and a collection of state space restrictions \(X_1, \ldots, X_n \subset X\) which satisfy the following assumption and hence form a partition of \(X\).

**Assumption 3.** The collection \(X_1, \ldots, X_n \subset \mathbb{R}^{n_X}\) satisfies

1. \(X = \bigcup_{k=1}^{n_X} X_k\),
2. \(X_i \cap X_j = \emptyset\) for all \(1 \leq i \neq j \leq n_X\).
3. the closure \(\overline{X}_i\) and boundary \(\partial X_i\) are basic closed semialgebraic for all \(i = 1, \ldots, n_X\)

**Remark 2.** A partition that satisfies Assumption 3 is in practice easily obtained by choosing \(X_i = X \cap I_i\), where \(I_1, \ldots, I_{n_X}\) is a collection of disjoint boxes whose union covers \(X\).

The elements \([t_{i-1}, t_i] \times X_k\) then form a partition of the problem’s entire space-time domain and we can
formulate the following natural generalization of \text{(subHJB)}:

$$
\begin{align*}
\sup_w & \sum_{k=1}^{n_X} \int_{X_k} w_{i,k}(0,x) \, d\nu_0(x) \\
\text{s.t.} & \quad A w_{i,k} + \ell \geq 0 \text{ on } [t_{i-1}, t_i] \times X_k \times U, \\
& \quad \forall (i,k) \in P, \\
& \quad w_{i,k}(t_{i-1}, \cdot) \geq w_{i-1,k}(t_{i-1}, \cdot) \text{ on } X_k, \\
& \quad \forall (i,k) \in P^o, \\
& \quad w_{i,k} = w_{i,j} \text{ on } [t_{i-1}, t_i] \times (\partial X_j \cap \partial X_k), \\
& \quad \forall (i,j,k) \in \partial P, \\
& \quad w_{nT,k}(T, \cdot) \leq \phi \text{ on } X_k, \\
& \quad \forall k \in \{1, \ldots, n_X\}, \\
& \quad w_{i,k} \in C^{1,2}([0,T] \times X_k), \quad \forall (i,k) \in P,
\end{align*}
$$

\text{(pw-subHJB)}

with the index sets

$$
\begin{align*}
P & := \{(i,k) : 1 \leq i \leq n_T, 1 \leq k \leq n_X\}, \\
P^o & := \{(i,k) : 2 \leq i \leq n_T, 1 \leq k \leq n_X\}, \\
\partial P & := \{(i,j,k) : 1 \leq i \leq n_T, 1 \leq k \neq j \leq n_X\}.
\end{align*}
$$

The constraints in Problem \text{(pw-subHJB)} ensure that a valid underestimator of the value function can be constructed from the function pieces \{w_{i,k} : (i,k) \in P\} for all elements of the partition. As such, Problem \text{(pw-subHJB)} yields a lower bound for the optimal value of \text{(OCP)}. This is formalized in the following Corollary.

\textbf{Corollary 2.} Let \{w_{i,k} : (i,k) \in P\} be feasible for \text{(pw-subHJB)} and define

$$
w(t,x) = w_{i,k}(t,x) \text{ with } i = \max\{j : t \in [t_{j-1}, t_j]\} \text{ and } k \text{ such that } x \in X_k.
$$

(11)

Then, \(w\) underestimates the value function \(V\) as defined in Equation (10).

\textbf{Sketch.} The idea is to split the paths of the process \((t,x_i,u_i)\) up into pieces during which it is confined to a single subdomain \([t_{i-1}, t_i] \times X_k \times U\). For each of those pieces an analogous argument as in Corollary 1 applies to show that \(w_{i,k}\) underestimates the value function for a process confined to the partition element \([t_{i-1}, t_i] \times X_k \times U\). Additionally, Constraints (7) and (8) ensure conservatism when the process crosses between different time intervals and subdomains of the state space, respectively. Specifically, Constraint (7) enforces that \(w(t,x_i)\) can at most decrease when traced backward in time across the boundary between the intervals \([t_i, t_{i+1}]\) and \([t_{i-1}, t_i]\), ensuring that \(w\) cannot cross \(V\) at such time points. Similarly, Constraint (8) imposes spatial continuity thus enforces that \(w\) cannot cross \(V\) when the process crosses spatial boundaries between partition elements. The formal argument is given in Appendix A.

\textbf{Remark 3.} Contrasting the monotonicity condition (7) enforced between subsequent time intervals, the stronger continuity requirement (8) at the boundary between spatial subdomains is necessary as the process may cross the boundary in any direction due to stochastic vibrations. In case of a deterministic process \((g = 0)\) this condition may be further relaxed as we only require that \(w(t,x_i)\) must at most increase for all trajectories of the system when crossing the boundary between two subdomains. Cibulka et al. [CKH21] show in a similar argument that in this case it suffices to impose that

$$
(w_{i,k}(t,x) - w_{i,j}(t,x))n_{j,k}^\top f(x,u) \geq 0, \forall (t,x,u) \in [t_{i-1}, t_i] \times (\partial X_j \cap \partial X_k) \times U,
$$

where \(n_{j,k}\) denotes the normal vector of the boundary between \(X_j\) and \(X_k\) pointing from \(X_j\) to \(X_k\).
Remark 4. Valid MSOS restrictions of \((\text{pw-subHJB})\) are readily obtained simply by restricting \(w\) to be a degree \(d\) polynomial and imposing non-negativity constraints through sufficient sum-of-squares constraints on the closure of the respective sets. Note that such sufficient sum-of-squares constraints are indeed well-posed due to Condition (3) in Assumption 3.

4 The primal perspective revisited: local occupation measures

In this section, we discuss the primal counterpart of the construction presented in the previous section. The primal counterpart of \((\text{pw-subHJB})\) reads

\[
\inf_{\nu,\xi} \sum_{(i,k) \in P} \langle \ell, \xi_{i,k} \rangle + \sum_{k=1}^{n_X} \langle \phi, \nu_{n_T,k} \rangle \\
\text{s.t.} \quad \langle w, \nu_{i-1,k} \rangle - \langle w, \nu_{i,k} \rangle = \langle Aw, \xi_{i,k} \rangle + \sum_{j \neq k} \langle w, \pi_{i,j,k} \rangle,
\]

\[
\forall w \in C^{1,2}([t_{i-1}, t_i] \times X_k), \quad \forall (i,k) \in P,
\]

\[
\nu_{i,k} \in \mathcal{M}^+(X_k), \quad \forall (i,k) \in P,
\]

\[
\xi_{i,k} \in \mathcal{M}^+([t_{i-1}, t_i] \times X_k \times U), \quad \forall (i,k) \in P,
\]

\[
\pi_{i,j,k} = -\pi_{i,k,j} \in \mathcal{M}([t_{i-1}, t_i] \times (\partial X_j \cap \partial X_k)),
\]

\[\forall (i,j,k) \in \partial P,\]

where \(\mathcal{M}(Y)\) refers to the space of signed measures supported on \(Y\). The decision variables in \((\text{pw-weak OCP})\) can be interpreted as localized generalization of the occupation measure pair introduced in Section 2. Specifically, the restriction of the expected state-action occupation measures \(\xi\) to a subdomain \([t_{i-1}, t_i] \times X_k \times U\) from the partition generates the local state-action occupation measure \(\xi_{i,k}\):

\[
\xi_{i,k}(B_T \times B_X \times B_U) = \xi((B_T \cap [t_{i-1}, t_i]) \times (B_X \cap X_k) \times B_U).
\]

Likewise, the local instantaneous occupation measures with respect to different time points \(t_i\) and subdomains \(X_k\) are given by the restriction of the instantaneous occupation measure at time \(t_i\) to \(X_k\), i.e.,

\[
\nu_{i,k}(B) = \mathbb{P}_{v_0}(x_{t_i} \in B \cap X_k).
\]

The measure \(\pi_{i,j,k}\) in \((\text{pw-weak OCP})\) takes the role of a slack variable and accounts for transitions of the process between the spatial subdomains \(X_j\) and \(X_k\) in the time interval \([t_{i-1}, t_i]\). Formally, \(\pi_{i,j,k}\) can be defined by

\[
\langle w, \pi_{i,j,k} \rangle := \mathbb{E}_{v_0} \left[ \sum_{n=1}^{N_{+}^{j,k}} w \left( \tau_{n+}^{j,k}, x_{\tau_{n+}^{j,k}} \right) - \sum_{n=1}^{N_{-}^{j,k}} w \left( \tau_{n-}^{j,k}, x_{\tau_{n-}^{j,k}} \right) \right],
\]

where \(\tau_{n+}^{j,k}\) and \(\tau_{n-}^{j,k}\) denote the \(n\)th time points in \([t_{i-1}, t_i]\) at which the process transitions from subdomain \(X_j\) into \(X_k\) and vice versa, respectively. With these interpretations, we can observe that the equality constraints in \((\text{pw-weak OCP})\) reduce to Dynkin’s formula applied between the stopping times of leaving and entering a given subdomain \(X_k\) in the time interval \([t_{i-1}, t_i]\) (see Appendix A for a more detailed derivation).

Finally, it is important to emphasize here that the above interpretation of the decision variables in \((\text{pw-weak OCP})\) as local occupation measures shows immediately that every feasible point for \((\text{pw-weak OCP})\) generates a
feasible point for \(\text{weak OCP}\) via the assignment \(\xi = \sum_{(i,k) \in P} \xi_{i,k}\) and \(\nu = \sum_{k=1}^{n_X} \nu_{n_T,k}\) with equal objective value. Analogously, any smooth function \(w\) that is feasible for \(\text{subHJB}\) generates upon restriction to the individual subdomains of the partition a feasible point for \(\text{pw-subHJB}\) with equal objective value. This property carries over directly to the MSOS restrictions and relaxations of \(\text{pw-subHJB}\) and \(\text{pw-weak OCP}\), respectively, as long as every closed subdomain \(X_k\) is represented in terms of a strictly greater set of polynomial inequalities than \(X\) is. This condition, which is easily obeyed in practice (see Remark 2), therefore guarantees that MSOS restrictions and relaxations of \(\text{pw-subHJB}\) and \(\text{pw-weak OCP}\), respectively, furnish bounds that are at least as tight as those obtained from their traditional counterparts.

5 Moment-sum-of-squares approximations: structure & scaling

The construction of tractable relaxations of the problems \(\text{subHJB}\) or \(\text{weak OCP}\) relies on the restriction to optimization over polynomials of fixed degree \(d\) or the relaxation to optimization over moment sequences truncated at order \(d\), respectively. Increasing this approximation order \(d\) has traditionally been the only mechanism used to weaken the restriction, respectively strengthen the relaxation, to improve the resultant bounds to a desired level. The main motivation behind the proposed partitioning approach lies in circumventing the limited practicality and interpretability of this tightening mechanism. With the proposed notion of local occupation measures, refinement of the space-time domain partition serves as an additional bound tightening mechanism. Table 1 summarizes how the MSOS SDP restrictions and relaxations of \(\text{pw-subHJB}\) and \(\text{pw-weak OCP}\) scale in size with respect to the different tightening mechanisms of increasing \(n_X, n_T\) (refining the partition), or \(d\) (increasing the approximation order). The linear scaling of the SDP sizes with respect to \(n_X\) and \(n_T\) underlines the fine-grained control over the tightening process via refinement of the partition. In particular, it opens the door to exploit problem specific insights such as the knowledge of critical parts of the (extended) state space \([0, T] \times X\) to be resolved more finely than others, to construct tighter relaxations without incurring a combinatorial increase in the number of partition elements. This flexibility and interpretability is in stark contrast to tightening the bounds by increasing the approximation order \(d\) as translating such insights into specific moments to be constrained or polynomial basis elements to be considered for the value function approximator is significantly less straightforward. It is further worth emphasizing that not only the linear scaling with respect to \(n_T\) and \(n_X\) is desirable but in particular that the invariance of the linear matrix inequality (LMI) dimension promotes practicality due to the unfavorable scaling of interior point algorithms whose running time scales worse than cubically with respect to this quantity \([JNW23]\).

Table 1: Scaling of MSOS SDP approximations with respect to different tightening mechanisms

|   | #variables | # LMIs | dimension of LMIs |
|---|------------|--------|-------------------|
| \(d\) | \(O\left(\frac{n+2}{d}\right)\) | \(O(1)\) | \(O\left(\frac{n+2}{d/2}\right)\) |
| \(n_T\) | \(O(n_T)\) | \(O(n_T)\) | \(O(1)\) |
| \(n_X\) | \(O(n_X)\) | \(O(n_X)\) | \(O(1)\) |

Additionally, the problems \(\text{pw-subHJB}\) and \(\text{pw-weak OCP}\) give rise to highly structured SDPs. Specifically, all constraints involve only variables corresponding to adjacent subdomains. As a consequence, the structure of the constraints is analogous to those arising from discretized PDEs may be exploited with suitable distributed optimization algorithms and computing architectures.

\(^2\)here, \(n = 1 + n_x + \max\{\text{deg}_xf - 1, \text{deg}_xg - 2\}\)
6 Example: population control

6.1 Control Problem

We demonstrate the computational merits of the proposed local occupation measure framework with an example problem from the field of population control. The problem is adjusted from Savorgnan et al. [SLD09] where it has been studied in a discrete time, infinite horizon setting. The objective is to control the population size of a primary predator and its prey in an ecosystem featuring the prey species, primary predator species as well as a secondary predator species. The interactions between the primary predator and prey population are described by a standard Lotka-Volterra model, while the effect of the secondary predator species is modeled by a Brownian motion. The population sizes are assumed to be controlled via hunting of the primary predator species. This model gives rise to the diffusion process

\[
\begin{align*}
dx_{t,1} &= (\gamma_1 x_{t,1} - \gamma_2 x_{t,1} x_{t,2}) \, dt + \gamma_3 x_{t,1} \, dB_t, \\
dx_{t,2} &= (\gamma_4 x_{t,1} x_{t,2} - \gamma_3 x_{t,2} - x_{t,2} u_t) \, dt,
\end{align*}
\]

where \(x_1, x_2, u\) refer to the prey species, predator species and hunting effort, respectively. The model parameters \(\gamma = (1, 2, 1, 2, 0.025)\) and initial condition \(x_0 \sim \delta_{(1,0.25)}\) are assumed to be known deterministically. Moreover, we assume that the admissible hunting effort is confined to \(U = [0, 1]\). Under these assumptions, it is easily verified that the process state \(x_t\) evolves by construction within the non-negative orthant \(X = \{x_1, x_2 \geq 0\}\) for any admissible control policy. For the control problem we further choose a time horizon of \(T = 10\) and stage cost

\[
\ell(x, u) = (x_1 - 0.75)^2 + \frac{(x_2 - 0.5)^2}{10} + \frac{(u - 0.5)^2}{10}
\]

penalizing variations from the target population sizes.

6.2 Partition of Problem Domain

In order to investigate the effect of different discretizations on bound quality and computational cost, we utilize a simple grid partition of the state space \(X\) as parameterized by the number of grid cells \(n_1\) and \(n_2\) in the \(x_1\) and \(x_2\) direction, respectively. As \(X\) is the non-negative orthant in our example, and hence semi-infinite, we choose to discretize the compact interval box \([0, 1.5] \times [0, 1.5]\) with a uniform grid of \((n_1 - 1) \times (n_2 - 1)\) cells and cover the remainder of \(X\) with appropriately chosen semi-infinite interval boxes. This choice is motivated by the insight that the uncontrolled system resides with high probability in \([0, 1.5] \times [0,1.5]\).

The temporal domain is partitioned uniformly into \(n_T\) subintervals, i.e., \(t_i = i \Delta t\) with \(\Delta t = T/n_T\). Throughout, we refer to a specific partition with the associated triple \((n_1, n_2, n_T)\). The computational experiments are conducted for all partitions corresponding to the triples \(\{(n_1, n_2, n_T) : 1 \leq n_1, n_2 \leq 5, 1 \leq n_T \leq 10\}\).

6.3 Evaluation of Bound Quality

In order to assess the tightness of the bounds obtained with different approximation orders and discretizations, we compare the relative optimality gap \((\bar{J} - J)/\bar{J}\), where \(\bar{J}\) and \(J\) refer to the lower bound furnished by an instance of the sum-of-squares restriction of (pw-subHJB) and to the control cost associated with the best known admissible control policy, respectively. The best known control policy was constructed from the approximate value function \(w^*\) obtained as the solution of the sum-of-squares restriction of (pw-subHJB).
with approximation order $d = 4$ on the grid described by $n_1 = n_2 = 4$ and $n_T = 10$. To that end, we employed the following control law mimicking a one-step model-predictive controller

$$u^*_t \in \arg \min_{u \in U} A w^*(t, x_t, u) + \ell(x_t, u)$$

and estimated the associated control cost

$$\bar{J} = \mathbb{E}_{x_0} \left[ \int_{[0,T]} \ell(x_t, u^*_t) \, dt \right]$$

by the ensemble average of 50,000 sample trajectories.

### 6.4 Computational Aspects

All computational experiments presented in this section were conducted on a MacBook M1 Pro with 16GB unified memory. All sum-of-squares programs and the associated SDPs were constructed using our custom developed and publicly available package MarkovBounds.jl built on top of SumOfSquares.jl [Wei+19] and the MathOptInterface [Leg+22]. All resultant SDPs were solved using Mosek v10.

### 6.5 Results

We put special emphasis on investigating the effect of refining the discretization of the problem domain on bound quality and computational cost. Focusing on the effect on computational cost in isolation first, Figure 1 indicates that the computational cost for the solution of MSOS programs generated by the restriction of (pw-subHJB) to polynomials of degree at most $d$ scales approximately linearly with the number of cells $n_1 \times n_2 \times n_T$ of the spatiotemporal partition. On the other hand, Figure 1 also shows that increasing the approximation order $d$ results in a much more rapid increase in computational cost. These results are in line with the discussion in Section 5.

![Figure 1: Linear scaling with respect to the number of grid cells for fixed approximation order](image)

Figure 2 shows the trade-off between bound quality and computational cost for different approximation orders and partitions. First, it is worth noting that the proposed partitioning strategy enables the computation of overall tighter bounds with an approximation order of only up to $d = 6$ when compared to the traditional formulation with an approximation order of up to $d = 18$. It is further worth emphasizing that beyond $d = 18$, numerical issues prohibited an accurate solution of the SDPs arising from the traditional formulation such

3see [https://github.com/FHoltorf/MarkovBounds.jl](https://github.com/FHoltorf/MarkovBounds.jl)
that no tighter bounds could be obtained this way. Furthermore, almost across almost the entire accuracy range a significant speed-up could be achieved by using the proposed partitioning strategy instead of only increasing the approximation order. Lastly, the results indicate that a careful choice of partitioning is crucial to achieve good performance. Figure 2 suggests that for this example particularly good performance is achieved when only the time domain is partitioned; additionally partitioning the spatial domain becomes an effective means of bound tightening only after the time domain has been resolved sufficiently finely.

![Figure 2](image-url)

(a) spatial & temporal partitioning (b) exclusively temporal partitions highlighted ($n_1 = n_2 = 1$)

Figure 2: Trade-off between computational cost and bound quality for different approximation orders $d$ and domain discretizations ($n_1, n_2, n_T$). The red markers correspond to MSOS restrictions of the labeled approximation order for the traditional formulation $\text{subHJB}$.

7 Extensions

Before we close, we briefly discuss two direct extensions to the described local occupation measure framework showcasing its versatility.

7.1 Discounted infinite horizon problems

Consider the following discounted infinite horizon stochastic control problem with discount factor $\rho > 0$:

$$\inf_{u_t} \mathbb{E}_{\nu_0} \left[ \int_{[0,\infty)} e^{-\rho t} \ell(x_t, u_t) \, dt \right]$$

s.t. $x_t$ satisfies (1) on $[0, \infty)$ with $x_0 \sim \nu_0$, $(x_t, u_t) \in X \times U$, on $[0, \infty)$,

$u_t$ is non-anticipative.

The construction of a weak formulation of this problem akin to weak OCP can be done in full analogy to Section 2. To that end, note that the infinitesimal generator $\tilde{A}$ maps functions of the form $\tilde{w}(t, x) = e^{-\rho t} w(t, x)$ to functions of the same form, i.e.,

$$\tilde{A} \tilde{w}(t, x, u) = e^{-\rho t} (A w(t, x, u) - \rho w(t, x, u)).$$

By analogous arguments as in Section 2 it therefore follows that any function $w \in C^{1,2}([0,\infty) \times X)$ that satisfies

$$A w(t, x, u) - \rho w(t, x, u) + \ell(x, u) \geq 0, \forall (t, x, u) \in [0, \infty) \times X \times U$$
generates a valid subsolution \( \hat{w}(t, x) = e^{-\rho t} w(t, x) \) of the value function associated with the infinite horizon problem. Since the proposed partitioning approach does neither rely on boundedness of the state space nor control horizon in order to establish valid bounds, it follows that it readily extends to the infinite horizon setting.

### 7.2 Jump processes with discrete state space

Many application areas ranging from chemical physics to queuing theory call for models that describe stochastic transitions between discrete states. In those cases, jump processes are a common modeling choice \[\text{Gil92, Bre03}\]. In the following, we will show that the proposed local occupation measure framework extends with only minor modifications to stochastic optimal control of a large class of such jump processes. To that end, we will consider controlled, continuous-time jump processes driven by \( m \) independent Poisson counters \( n_i(t) \) with associated propensities \( a_i(x_t, u_t) \):

\[
dx_t = \sum_{i=1}^{m} [h_i(x_t, u_t) - x_t] \, dn_{i,t}.
\]

We will again assume that the process can be fully characterized by polynomials, but we now additionally impose the assumption that the state space of the process is discrete.

**Assumption 4.** The jumps \( h_i : X \times U \to X \), propensities \( a_i : X \times U \to \mathbb{R}_+ \), stage cost \( l : X \times U \to \mathbb{R} \) and terminal cost \( \phi : X \times U \to \mathbb{R} \) are polynomial functions jointly in both arguments. The state space is a discrete, countable set and the set of admissible control inputs \( U \) is a basic closed semialgebraic set.

The local occupation measure framework outlined previously for diffusion processes can be extended for computing lower bounds on stochastic optimal control problems with such jump processes embedded:

\[
\inf_{u_t} \mathbb{E}_{\nu_0} \left[ \int_{[0,T]} l(x_t, u_t) \, dt + \phi(x_T) \right] \quad \text{(jump OCP)}
\]

s.t. \( x_t \) satisfies (12) on \([0, T]\) with \( x_0 \sim \nu_0 \),

\[
(x_t, u_t) \in X \times U, \quad \text{on } [0, T],
\]

\( u_t \) is not anticipative.

Given the extended infinitesimal generator \( \mathcal{A} : C^{1,0}([0,T] \times X) \to C([0,T] \times X \times U) \) associated with the process (12),

\[
\mathcal{A}w \mapsto \frac{\partial w}{\partial t}(t, x) + \sum_{i=1}^{m} a_i(x, u)(w(t, h_i(x, u)) - w(t, x)),
\]

the weak form of (jump OCP) and its dual are analogous to (weak OCP) and (subHJB), respectively. Further, given a partition of the problem’s space-time domain as introduced in Section 3, the analog of
Problem \( (\text{pw-subHJB}) \) seeking a piecewise smooth subsolution of the value function takes the form

\[
\sup_{w_{i,k}(i,k) \in P} \sum_{k=1}^{n_X} \int_{X_k} w_{1,k}(0, \cdot) \, d\nu_0 \\
\text{s.t.} \quad A w_{i,k} + \ell \geq 0 \text{ on } [t_{i-1}, t_i] \times X_k \times U, \\
\forall (i, k) \in P, \\
w_{i,k}(t_{i-1}, \cdot) \geq w_{i-1,k}(t_{i-1}, \cdot) \text{ on } X_k, \\
\forall (i, k) \in P^o, \\
w_{i,k} = w_{i,j} \text{ on } [t_{i-1}, t_i] \times N_{X_k}(X_j), \\
\forall (i, j, k) \in \partial P, \\
w_{nT,k}(T, \cdot) \leq \phi \text{ on } X_k, \\
\forall k \in \{1, \ldots, n_X \} \\
w_{i,k} \in C^{1,0}([0, T] \times X_k), \forall (i, k) \in P,
\]

where \( N_{X_k}(X_j) \) denotes the “neighborhood” of \( X_k \) in \( X_j \) defined as all states in \( X_j \) which have a non-zero transition probability into \( X_k \); formally,

\[
N_{X_k}(X_j) = \{ x \in X_j : \exists u \in U \text{ such that } h_i(x, u) \in X_k \text{ for some } i \text{ and } a_i(x, u) > 0 \}.
\]

Note that under Assumption 4, \( A \) again maps polynomials to polynomials laying the basis for the application of the MSOS hierarchy to construct tractable relaxations of \( (\text{jump OCP}) \). In contrast to the discussion in Section 2 however, the state space \( X \) of a jump process is closed basic semialgebraic if and only if it is finite. Thus, the MSOS hierarchy provides finite SDP relaxations of the weak form of \( (\text{jump OCP}) \) only in the case of a finite state space \( X \). Moreover, even if \( X \) is finite but of large cardinality, these relaxations may not be practically tractable due to the large number or high degree of the polynomial inequalities needed to describe such a set. If \( X \) is infinite (or of sufficiently large cardinality), tractable MSOS relaxations can only be constructed at the price of introducing additional conservatism. From the dual perspective, this additional conservatism is introduced by imposing the non-negativity conditions in \( (\text{subHJB}) \) on a basic semialgebraic overapproximation of \( X \); in particular polyhedral overapproximations are a common choice \([DB18; Ghu+17; Kun+19; SH17; HB24]\). The framework of local occupation measures provides a way to reduce this conservatism. In order to construct tractable relaxations for \( (\text{jump pw-subHJB}) \) via the MSOS hierarchy, it of course remains still necessary to replace any infinite (or very large) partition element by a closed basic semialgebraic overapproximation; however, the union of suitably chosen overapproximations of the individual partition elements will generally be less conservative than a global overapproximation.

### 7.3 Example: optimal gene regulation for protein expression

We demonstrate the efficacy of the local occupation measure framework for the control of jump processes with an example from cellular biology. Specifically, we consider the problem of optimal regulation of protein expression through actuation of the promoter kinetics in the biocircuit illustrated in Figure 3. The biocircuit is modeled as a jump process reflecting the stochastic nature of chemical reactions in cellular environments with low molecular copy numbers \([Gil92]\). The associated jump process has three states encoding the molecular counts of protein \( (x_1) \), active promoter \( (x_2) \), and inactive promoter \( (x_3) \) undergoing jump transitions...
in response to the following chemical reactions with associated rates:

\[ h_1 : (x_1, x_2, x_3) \mapsto (x_1 + 1, x_2, x_3), \]
\[ a_1(x, u) = 10x_2 \]  
expression

\[ h_2 : (x_1, x_2, x_3) \mapsto (x_1 - 1, x_2, x_3), \]
\[ a_2(x, u) = 0.1x_1 \]  
(degradation)

\[ h_3 : (x_1, x_2, x_3) \mapsto (x_1, x_2 - 1, x_3 + 1), \]
\[ a_3(x, u) = 0.1x_1x_2 \]  
(repression)

\[ h_4 : (x_1, x_2, x_3) \mapsto (x_1, x_2 + 1, x_3 - 1), \]
\[ a_4(x, u) = 10(1 - u)x_3 \]  
(activation)

\[ h_5 : (x_1, x_2, x_3) \mapsto (x_1, x_2 - 1, x_3 + 1), \]
\[ a_5(x, u) = 10ux_2 \]  
(inactivation)

The expression of protein can be controlled indirectly via the activation and inactivation rates of the promoter. Admissible control actions \( u \) are constrained to lie within the interval \( U = [0, 1] \). Moreover, we assume a deterministic initial condition \( x_0 \sim \delta(0, 1, 0) \) and exploit that due to the reaction invariant \( x_{t,2} + x_{t,3} = x_{0,2} + x_{0,3} \) the state space \( X \) is effectively two-dimensional, i.e., we eliminate \( x_{t,3} = 1 - x_{t,2} \). It can be easily verified that, after elimination of the reaction invariant, the state space of the jump process is given by

\[ X = \{x \in \mathbb{Z}_+^2 : x_2 \in \{0, 1\}\} \]
such that Assumption 4 is satisfied.

The goal of the control problem is to stabilize the protein level in the cell at a desired value of 10 molecules. To that end, we choose to minimize the stage cost

\[ \ell(x, u) = (x_1 - 10)^2 + 10(u - 0.5)^2 \]
over the horizon \([0, 10]\).

In order to investigate the effect of different partitions of the problem domain on bound quality and computational cost, we discretize the time horizon uniformly into \( n_T \) intervals and partition the state space into \( 2n_X \) singletons

\[ X_i = \begin{cases} 
(i - 1, 0), & i \leq n_X \\
(i - n_X - 1, 1), & i > n_X \text{ for } i = 1, \ldots, 2n_X 
\end{cases} \]

and lump the remaining part of the state space in the last partition element

\[ X_{2n_X + 1} = \{x \in \mathbb{Z}_+^2 : x_1 \geq n_X, x_2 \in \{0, 1\}\}. \]
We explore the partitions corresponding to all combinations of $n_T \in \{2, 4, \ldots, 18, 20\}$ and $n_X \in \{0, 8, \ldots, 32, 40\}$.

Note that the partition elements $X_1, \ldots, X_{2n_X}$ are already basic closed semialgebraic such that no overapproximation is required for the construction of valid MSOS restriction of the non-negativity constraints in (jump pw-subHJB). In contrast, $X_{2n_X+1}$ is infinite, hence not basic closed semialgebraic. We therefore strengthen the formulation of the MSOS restriction of (jump pw-subHJB) by imposing the non-negativity conditions on the polyhedral convex hull of $X_{2n_X+1}$, thereby recovering tractability.

Figure 4 shows the trade-off between computational cost and bound quality achieved by different choices for the partition of the problem domain and approximation order. The bound quality is again measured by the relative optimality gap, estimated as described in Section 6.3. Analogous to the diffusion control example considered in Section 6, the results demonstrate that an adequate partitioning of the problem domain substantially reduces the cost of computing bounds of a given quality when compared to the traditional approach. Moreover, notably tighter bounds could be computed overall due to a less conservative overapproximation of the process’ infinite state space in the formulation of the bounding problems.

8 Conclusion

We have proposed a simple partitioning strategy for improving the practicality of MSOS relaxations for stochastic optimal control problems with polynomial data. From the primal perspective, this strategy can be interpreted as constructing the MSOS relaxation for a linear program over finitely many occupation measures “localized” on elements of a partition of the control problem’s space-time domain. From the dual perspective, the bounding problems seek a maximal piecewise-polynomial underestimator to the value function via sum-of-squares programming.

The key advantage of this framework over application of the MSOS hierarchy to the traditional occupation measure formulation for stochastic optimal control is that it offers a flexible and interpretable mechanism to tighten the obtained semidefinite bounding problems without degree augmentation – simple refinement of the problem domain partition. On the one hand, this enables tightening of the bounding problems at as benign as linearly increasing cost, contrasting the combinatorial scaling incurred by naive degree augmentation. On the other hand, it promotes practicality by providing a way to avoid high degree sum-of-squares constraints and their notorious implications for poor numerical conditioning. As demonstrated with two examples, these advantages can lead to notable improvements in practical utility of the occupation measure approach to stochastic optimal control.

In future work, we will investigate the use of distributed optimization techniques to further improve efficacy.
of the proposed framework by exploiting the weakly-coupled block structure of the bounding problems.

Acknowledgements

FH gratefully acknowledges helpful discussions with Paul I. Barton and Frank Schäfer.

This material is based upon work supported by the National Science Foundation under grant no. OAC-1835443, grant no. SII-2029670, grant no. ECCS-2029670, grant no. OAC-2103804, and grant no. PHY-2021825. We also gratefully acknowledge financial support from the MIT-Switzerland Lockheed Martin Seed Fund and MIT International Science and Technology Initiatives (MISTI).

References

[ADH17] A. A. Ahmadi, S. Dash, and G. Hall. “Optimization over structured subsets of positive semidefinite matrices via column generation”. In: Discrete Optimization 24 (2017), pp. 129–151.

[AH19] A. A. Ahmadi and G. Hall. “On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity”. In: Mathematics of Operations Research 44.4 (2019), pp. 1192–1207.

[AM14] A. A. Ahmadi and A. Majumdar. “DSOS and SDSOS optimization: LP and SOCP-based alternatives to sum of squares optimization”. In: 2014 48th Annual Conference on Information Sciences and Systems (CISS). IEEE. 2014, pp. 1–5.

[AM19] A. A. Ahmadi and A. Majumdar. “DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization”. In: SIAM Journal on Applied Algebra and Geometry 3.2 (2019), pp. 193–230.

[Aug+23] N. Augier, D. Henrion, M. Korda, and V. Magron. “Symmetry reduction and recovery of trajectories of optimal control problems via measure relaxations”. In: arXiv preprint arXiv:2307.03787 (2023).

[BB96] A. G. Bhatt and V. S. Borkar. “Occupation Measures for Controlled Markov Processes: Characterization and Optimality”. In: The Annals of Probability 24.3 (1996), pp. 1531–1562.

[Boy+10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. 2010.

[Bre03] L. Breuer. From Markov jump processes to spatial queues. Springer Science & Business Media, 2003.

[CS02] M. J. Cho and R. H. Stockbridge. “Linear programming formulation for optimal stopping problems”. In: SIAM Journal on Control and Optimization 40.6 (2002), pp. 1965–1982.

[CKH21] V. Cibulka, M. Korda, and T. Haniš. “Spatio-temporal decomposition of sum-of-squares programs for the region of attraction and reachability”. In: IEEE Control Systems Letters 6 (2021), pp. 812–817.

[DB18] G. R. Dowdy and P. I. Barton. “Bounds on stochastic chemical kinetic systems at steady state”. In: The Journal of Chemical Physics 148.8 (2018), p. 84106.

[FV89] W. H. Fleming and D. Vermes. “Convex duality approach to the optimal control of diffusions”. In: SIAM Journal on Control and Optimization 27.5 (1989), pp. 1136–1155.
[GQ09] V. Gaitsgory and M. Quincampoix. “Linear programming approach to deterministic infinite horizon optimal control problems with discounting”. In: *SIAM Journal on Control and Optimization* 48.4 (2009), pp. 2480–2512.

[Ghu+17] K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh. “Exact lower and upper bounds on stationary moments in stochastic biochemical systems”. In: *Physical Biology* 14.4 (2017), 04LT01.

[Gil92] D. T. Gillespie. “A rigorous derivation of the chemical master equation”. In: *Physica A: Statistical Mechanics and its Applications* 188.1-3 (1992), pp. 404–425.

[HS00] K. Helmes and R. Stockbridge. “Numerical comparison of controls and verification of optimality for stochastic control problems”. In: *Journal of Optimization Theory and Applications* 106 (2000), pp. 107–127.

[Hen+23a] D. Henrion, M. Infusino, S. Kuhlmann, and V. Vinikov. “Infinite-dimensional moment-SOS hierarchy for nonlinear partial differential equations”. In: *arXiv preprint arXiv:2305.18768* (2023).

[HJV21] D. Henrion, M. Junca, and M. Velasco. “Moment-SOS hierarchy and exit time of stochastic processes”. In: *arXiv preprint arXiv:2101.06009* (2021).

[Hen+23b] D. Henrion, M. Korda, M. Kružík, and R. Rios-Zertuche. “Occupation measure relaxations in variational problems: the role of convexity”. In: *arXiv preprint arXiv:2303.02434* (2023).

[HLS08] D. Henrion, J. B. Lasserre, and C. Savorgnan. “Nonlinear optimal control synthesis via occupation measures”. In: *2008 47th IEEE Conference on Decision and Control*. IEEE. 2008, pp. 4749–4754.

[Her+99] O. Hernández-Lerma, J. B. Lasserre, O. Hernández-Lerma, and J. B. Lasserre. “The linear programming approach”. In: *Further Topics on Discrete-Time Markov Control Processes* (1999), pp. 203–249.

[HB24] F. Holtorf and P. I. Barton. “Tighter bounds on transient moments of stochastic chemical systems”. In: *Journal of Optimization Theory and Applications* 200.1 (2024), pp. 104–149.

[Hol+23] F. Holtorf, F. Schäfer, J. Arnold, C. Rackauckas, and A. Edelman. “Performance Bounds for Quantum Control”. In: *arXiv preprint arXiv:2304.03366* (2023).

[IW14] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. Elsevier, 2014.

[JNW23] S. Jiang, B. Natura, and O. Weinstein. “A faster interior-point method for sum-of-squares optimization”. In: *Algorithmica* (2023), pp. 1–42.

[KHL18] M. Korda, D. Henrion, and J. B. Lasserre. “Moments and convex optimization for analysis and control of nonlinear partial differential equations”. In: *arXiv preprint arXiv:1804.07565* (2018).

[KHL22] M. Korda, D. Henrion, and J. B. Lasserre. “Moments and convex optimization for analysis and control of nonlinear PDEs”. In: *Handbook of Numerical Analysis*. Vol. 23. Elsevier, 2022, pp. 339–366.

[Kun+19] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona. “Bounding the stationary distributions of the chemical master equation via mathematical programming”. In: *The Journal of Chemical Physics* 151.3 (2019), p. 34109.

[KS98] T. G. Kurtz and R. H. Stockbridge. “Existence of Markov Controls and Characterization of Optimal Markov Controls”. In: *SIAM Journal on Control and Optimization* 36.2 (1998), pp. 609–653.

[LGS18] A. Lamperski, K. R. Ghusinga, and A. Singh. “Analysis and control of stochastic systems using semidefinite programming over moments”. In: *IEEE Transactions on Automatic Control* 64.4 (2018), pp. 1726–1731.

[Las01] J. B. Lasserre. “Global optimization with polynomials and the problem of moments”. In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817.
A Proof of Corollary 1

Proof. Fix $z \in X$ and $t \in [t_{n-1}, T]$. Now consider an admissible control process $u_s$ such that all paths of the controlled process $(s, x_s, u_s)$ lie in $[t, T] \times X \times U$ with $x_t \sim \delta_z$. Further define $\tau_0 = t$ and $\tau_i$ for $i \geq 1$ to be the minimum between $T$ and the time point at which the process crosses for the $i$th time from
one subdomain of the partition \(X_1, \ldots, X_{n_X}\) to another. By construction, the process is confined to some (random) subdomain \(X_k\) in the interval \([\tau_i, \tau_{i+1}]\). Since \(w_{nT,k}\) is sufficiently smooth on \([\tau_i, \tau_{i+1}] \times X_k\), Ito’s lemma applies and yields that

\[
w_{nT,k}(\tau_{i+1}, x_{\tau_{i+1}}) = w_{nT,k}(\tau_i, x_{\tau_i}) + \int_{\tau_i}^{\tau_{i+1}} A w_{nT,k}(s, x_s, u_s) ds + \int_{\tau_i}^{\tau_{i+1}} \nabla_x w_{nT,k}(s, x_s)^\top g(x_s, u_s) dB_s.
\]

Now note that by Constraint (6),

\[
\int_{\tau_i}^{\tau_{i+1}} A w_{nT,k}(s, x_s, u_s) ds \geq - \int_{\tau_i}^{\tau_{i+1}} \ell(x_s, u_s) ds.
\]

Further note that

\[
\mathbb{E}_{\delta_s} \left[ \int_{\tau_i}^{\tau_{i+1}} \nabla_x w_{nT,k}(s, x_s)^\top g(x_s, u_s) dB_s \right] = 0
\]

as the integrand is square-integrable by Assumption (2) and \(\tau_i \leq \tau_{i+1}\) are stopping times with respect to the natural filtration \([W_t]_{t \geq 0}\). Thus, after taking expectations, we obtain

\[
\mathbb{E}_{\delta_s} \left[ w_{nT,k}(\tau_i, x_{\tau_i}) \right] \leq \mathbb{E}_{\delta_s} \left[ \int_{\tau_i}^{\tau_{i+1}} \ell(x_s, u_s) ds + w_{nT,k}(\tau_{i+1}, x_{\tau_{i+1}}) \right].
\]

Moreover, continuity holds at any crossing between any distinct subdomains \(X_k\) and \(X_j\) due to Constraint (8) such that

\[
\mathbb{E}_{\delta_s} \left[ w(\tau_i, x_{\tau_i}) \right] = \mathbb{E}_{\delta_s} \left[ w_{nT,k}(\tau_i, x_{\tau_i}) \right] = \mathbb{E}_{\delta_s} \left[ w_{nT,j}(\tau_i, x_{\tau_i}) \right],
\]

when the process crosses from \(X_k\) to \(X_j\) at \(\tau_i\). Now using that \(\mathbb{E}_{\delta_s} [w(\tau_0, x_{\tau_0})] = w(t, z)\), we obtain by summing over the time intervals \([\tau_0, \tau_1], \ldots, [\tau_N, \tau_{N+1}]\) that

\[
w(t, z) \leq \mathbb{E}_{\delta_s} \left[ \int_{\tau_i}^{\tau_{i+1}} \ell(x_s, u_s) ds + w(\tau_{N+1}, x_{\tau_{N+1}}) ds \right].
\]

Letting \(N \to \infty\), it follows that

\[
w(t, z) \leq \mathbb{E}_{\delta_s} \left[ \int_{t}^{T} \ell(x_s, u_s) ds + w(T, x_T) \right]
\]

as \(\tau_N \to T\) almost surely. Finally using that \(w(T, x) \leq \phi(x)\) on \(X\) due to Constraint (9) and the fact that all results hold for any admissible control policy, we obtain the desired result \(w(t, z) \leq V(t, z)\).

It remains to show that \(w\) preserves the lower bounding property across the boundaries introduced by discretization of the time domain. To that end, note that by an analogous argument as before, we have for any \(t \in [t_{i-1}, t_i]\) that

\[
w(t, z) \leq \mathbb{E}_{\delta_s} \left[ \int_{t}^{t_i} \ell(x_s, u_s) ds + \lim_{s \nearrow t_i} w(s, x_s) \right].
\]

Since Constraint (7) implies that \(\lim_{s \nearrow t_i} w(s, x) \leq w(t_i, x)\) on \(X\), it finally follows by induction that \(w(t, z) \leq V(t, z)\) for any \(t \in [0, T]\) and \(z \in X\). \(\square\)