A Dynamic Programing Aproach to Evaluating Multivariate Gaussian Probabilities

Morgan Jones, Matthew M. Peet

Abstract— **We propose a method of approximating multivariate Gaussian probabilities using dynamic programming. We show that solving the optimization problem associated with a class of discrete-time finite horizon Markov decision processes with non-Lipschitz cost functions is equivalent to integrating a Gaussian functions over polytopes. An approximation scheme for this class of MDP's is proposed and explicit error bounds under the supremum norm for the optimal cost to go functions are derived.**

I. INTRODUCTION

Integration of a Gaussian function over a polytope is a central computational bottleneck in several control and optimization problems, including machine learning [1], chance constrained optimization [2], [3], [4], and statistical modeling [5]. While this problem is known to be a computationally challenging problem [6], in this paper, we show that it can be reformulated as an optimization problem associated with a Markov Decision Process (MPD). Next, we show that MDPs of this form can be uniformly approximated by MDPs with countable state space. Finally, we show that this sequence of approximated MDPs can be efficiently solved using a variation of Belman's equation. The solution is then demonstrated in several numerical examples.

Many methods for integrating a Gaussian function over a polytope have emerged in the literature. Genz [7] represents the state of the art, where the algorithm makes a series of transformations to reduce integration over a hyper-rectangle to integration over a unit cube. Here lattice point numerical integration can be used and explicit error bounds can be achieved. However Genz only looks at the specific case where the integration is over a rectangle. Several other algorithms of Gaussian integration over rectangles can be found in [8]. Another approach is to use bounding methods where the polytope is inner approximated by closed and bounded sectors such as in [9], but no error bound can be found here. A common approach is to use expectation propagation but as seen in [6] this method performs badly on anything that is more complicated than a rectangular integration region. An alternative method is to use probabilistic methods where confidence intervals can be provided instead of error bounds [10]. In this paper we propose an integration algorithm over a possibly non-compact general polytope with explicit error bounds.

M. Jones is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: `morgan.c.jones@asu.edu`

M. Peet is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: `mpeet@asu.edu`

In Section III we show the equivalence of solving the optimization problem associated with a class of MDP's and evaluating Gaussian probabilities. MDP's describe the mathematical framework for modeling discrete time evolving processes involving a decision making situation coupled with partly random outcomes. Each MDP has an associated optimization problem of picking the sequence of decisions that minimizes the total expected cost of the process. MDP's appear in a vast number of fields such as economics, computer science, engineering etc; an in depth list of application of MDP's can be found in the survey [11].

MDP's are commonly solved using dynamic programing [12]. Unfortunately in practice it is rare to be able find an analytical solution to Bellman's equation and thus the problem must be solved numerically, see [13] as an example. In this paper we are interested in MDP's where the state and control spaces can be uncountable (for example $[0,1]$). In these cases for an algorithm to solve the problem it becomes necessary to approximate the MDP by discretization; that is we replace the state and control spaces with a countable set. One hopes there is sufficient continuity in the original MDP such that as the discretization sharpness increases a solution can be found arbitrarily close to the true solution.

In the literature there has been much work done on deriving error bounds for discretization approximations of MDP's with compact control and state spaces and Lipschitz cost functions [14], [15]. However in many practical problems the state dynamics are of the form $x(t+1) = Ax(t) + B\varepsilon(t)$ where $\varepsilon \sim \mathcal{N}(0,1)$, inducing the non-compact state space of $x \in \mathbb{R}^n$. A major contribution was made in [16] where a discretization scheme was proposed and error bounds were proved for a general class of MDP's with locally compact state and control spaces. In this paper we modify and extend the work of [16] to the case when the terminal cost function of the MDP is non-Lipschitz. The discretization scheme we propose is to first approximate the cost function by a Lipschitz continuous function and then to use the discretization scheme from [16].

The rest of this paper is organized as follows. In section III we show the relation of MDP's and integrating Gaussian random variables over polytopes. In section IV we introduce the class of MDP's we are interested in approximating. In Section V we show how to approximate this class of MDP's. In Section VI we present our numerical results and in VII we finish with our conclusion.

II. NOTATION

For a matrix $A \in \mathbb{R}^{m,n}$ we denote the j'th column of A by $A_{.,j} = [a_{1,j}, \ldots, a_{m,j}]^T$.

For $A \in \mathbb{R}^{m \times n}$ we define $||A||_{\max} = \max\{|a_{i,j}|\}$.

We define the power set of a set S to be the set of all subsets denoted by $S^{\cup} := \{U : U \subseteq S\}$.

For $T \in \mathbb{N}$ we denote the set $[T] = \{0, \ldots, T\}$.

We define the positive scalars as $\mathbb{R}^{++} := \{x \in \mathbb{R} : x > 0\}$.

For functions $f_1 : X \to \mathbb{R}$ and $f_2 : X \to \mathbb{R}$ we denote $f_1(x) \vee f_2(x) := \max\{f_1(x), f_2(x)\}$.

We denote the Hausdorff metric space in $\mathbb{R}^n$ as D^n with metric d_H, which is the set of non-empty subsets of $\mathbb{R}^n$ where if $X, Y \in D^n$, then $d_H(X,Y) = \max\{\sup_{\{x \in X\}} \inf_{\{y \in Y\}} ||x - y||_2, \sup_{\{y \in Y\}} \inf_{\{x \in X\}} ||x - y||_2\}$.

The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be Lipschitz continuous if there exists $L > 0$ such that :

$$||f(x_1) - f(x_2)||_2 \le L ||x_1 - x_2||_2 \qquad \text{for all } x_1, x_2 \in X \quad (1)$$

For a Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$, we denote by L_f the smallest constant L such that Equation (1) holds.

For bounded function on X, we denote the infinity norm as $||h||_\infty := \sup_{x \in X} |h(x)|$.

For a given weighting function $w : \mathbb{R}^n \to \mathbb{R}^{++}$, we also define the weighted infinity norm $||v||_w := \sup_{x \in \mathbb{R}^n} \{\frac{|v(x)|}{w(x)}\}$ and $v \in \mathbb{L}_w(X)$ to be the space of Lipschitz continuous functions with finite $||v||_w$.

We denote $\mathscr{B}(X)$ to be the Borel sigma algebra of some set X.

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We say $Z : \Omega \to \mathbb{R}$ is a real valued random variable if it is a $\mathscr{F}$-measurable function. For any $B \in \mathscr{F}$ we denote the law of Z by $\mathbb{P}_Z(B) := \mathbb{P}(\{w : Z(w) \in B\})$. For a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ we define the expectation as $\mathbb{E}_Z[g(Z)] := \int_{\mathbb{R}} g d\mathbb{P}_Z$. Furthermore we say $Z \sim \mathcal{N}(\mu, \Sigma)$, $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ if $\mathbb{P}_Z(B) = \int_B \phi(x) dx$ where $\phi : \mathbb{R}^n \to [0,1]$ is given by $\phi(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$.

For any subset $X \subset Z$, we define the indicator function $\mathbb{1}_X : Z \to \{0,1\}$ as

$$\mathbb{1}_X(x) = \begin{cases} 1, & \text{if } x \in X \\ 0, & \text{otherwise.} \end{cases}$$

In Section V, we will make use of a parameterized smoothed indicator function $g_{\lambda,b}(x) : \mathbb{R}^m \to [0,1]$ which is defined for any $b \in \mathbb{R}^m$ and $\lambda > 0$ as $g_{\lambda,b}(x) := \Pi_{i=1}^m g_i(x)$ where

$$g_i(x) := \begin{cases} 1, & \text{if } x_i < b_i - \frac{1}{\lambda} \\ -\lambda(x_i - b_i), & \text{if } b_i - \frac{1}{\lambda} < x_i < b_i \\ 0, & \text{if } b_i < x_i. \end{cases} \quad (2)$$

Associated with $g_{\lambda,b}$, we define the region of smoothing $\mathscr{R} \subset \mathbb{R}^n$ as

$$\mathscr{R}_{\lambda,b} = \{x \in \mathbb{R}^n : g_{\lambda,b}(x) \ne \mathbb{1}_{\{x \in \mathbb{R}^n : x < b\}}(x)\}$$

Suppose (X, D) is a compact metric space. We say the set $\Gamma_\beta = \{x_1, \ldots, x_n\}$ is an β-partition of X if:

- There exists disjoint subsets, $X_1, \ldots, X_n$, of X such that $\cup_{i=1}^n X_i = X$ and $x_i \in X_i$ for $i \in \{1, \ldots, n\}$.
- $D(x, x_i) \le \beta$ for all $x \in X_i$.

Furthermore given a partition $\Gamma_\beta = \{x_1, \ldots, x_n\}$ of some space X, we define $p_{X,\Gamma_\beta} : X \to \Gamma_\beta$ as $p_{X,\Gamma_\beta}(x) = x_i$ for every $x \in X_i$.

III. MULTI-VARIABLE GAUSSIAN INTEGRATION OVER POLYTOPES CAN BE WRITTEN AS A DYNAMIC PROGRAMING PROBLEM

Our aim is to compute:

$$\mathbb{P}_Z(Z \in \mathscr{P}) = \int_{x \in \mathscr{P}} \phi(x) dx \quad (3)$$

Where $Z \sim \mathcal{N}(\mu, \Sigma)$, $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n \times n}$ $\mathscr{P} = \{x \in \mathbb{R}^n : Ax \le b\}$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Remark 1: For any $\Sigma > 0$, there exists an invertible $C \in \mathbb{R}^{n \times n}$ such that $\Sigma = CC^T$ and under the transformation $N = C^{-1}(Z - \mu)$ we see $N \sim \mathcal{N}(0, I)$. Thus $\mathbb{P}_Z(Z \in \mathscr{P}) = \mathbb{P}_N(N \in \mathscr{P}')$ where $\mathscr{P}' = \{x \in \mathbb{R}^n : ACx \le b - A\mu\} = \{x \in \mathbb{R}^n : A'x \le b'\}$. Therefore without loss of generality we can assume $\mu = 0$ and $\Sigma = I$ for the rest of this paper.

Lemma 1: For every polytope $\mathscr{P} \subset \mathbb{R}^n$ there exists $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ such that $\mathscr{P} = \{x \in \mathbb{R}^n : Ax \le b\}$ with $a_{i,n} \ne 0 \ \forall i \in \{1, \ldots, m\}$.

Proof: Since $\mathscr{P}$ is a polytope there exists some $A' \in \mathbb{R}^{m \times n}$ and $b' \in \mathbb{R}^m$ such that $\mathscr{P} = \{x \in \mathbb{R}^n : A'x \le b'\}$. Now $A'x \le b' \iff TA'x \le Tb'$ for any elementwise-nonnegative invertible $T \in \mathbb{R}^{m \times m}$. WLOG we assume there is at least one nonnegative element in the last column of A' (otherwise we can restrict the space to $\mathbb{R}^{n-1}$) and by relabeling coordinates we assume $a'_{m,n} \ne 0$. Consider the

matrix $T = \begin{bmatrix} \frac{1}{2} & 0 & \ldots & 0 & ||A'||_{\max}/|a'_{m,n}| \\ 0 & \frac{1}{2} & \ldots & 0 & ||A'||_{\max}/|a'_{m,n}| \\ \vdots & \ldots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{1}{2} & ||A'||_{\max}/|a'_{m,n}| \\ 0 & 0 & \ldots & 0 & ||A'||_{\max}/|a'_{m,n}| \end{bmatrix}$. Clearly T

has all nonnegative elements and is invertible as all of its columns are independent. It follows $\mathscr{P} = \{x \in \mathbb{R}^n : TA'x < Tb'\}$ where $TA'_{.,n} = [\frac{a_{1,n}}{2} + sign(a_{m,n})||A'||_{\max}, \ldots, \frac{a_{m-1,n}}{2} + sign(a_{m,n})||A'||_{\max}, sign(a_{m,n})||A'||_{\max}]^T$ which clearly has no nonzero elements since $||A'||_{\max} > a_{i,n}$ for $1 \le i \le m$. ∎

We consider the Dynamic Programing (DP) problem:

$$J = \mathbb{E}_{x_T}\{\mathbb{1}_{\{x \le b\}}(x_T)\} \quad \text{Subject to:} \quad (4)$$
$$x_{i,t+1} = x_{i,t} + a_{i,t+1}\varepsilon_t \quad t \in [T-1], \quad i \in \{1, \ldots, m\}$$
$$x_{i,0} = 0 \quad i \in \{1, \ldots, m\}$$
$$\varepsilon_t \sim \mathcal{N}(0,1) \quad t \in [T-1]$$

Proposition 1: The objective function, J, defined in (4) is equal to $\int_{x \in \mathscr{P}} \phi(x) dx$, where $\mathscr{P} = \{x \in \mathbb{R}^T : Ax \le b\}$ and $A = \{a_{i,t}\} \in \mathbb{R}^{m \times T}$.

Proof: Let us denote $x_t = (x_{1,t}, \ldots, x_{m,t})^T$. From the second line in (4) we see $x_{t+1} = x_t + A_{.,t+1}\varepsilon_t$. Thus:

$$x_T = x_{T-1} + A_{.,T}\varepsilon_{T-1} = x_{T-2} + A_{.,T-1}\varepsilon_{T-2} + A_{.,T}\varepsilon_{T-1}$$
$$\cdots \cdots = A_{.,1}\varepsilon_0 + \cdots + A_{.,T}\varepsilon_{T-1} = Az$$

Where $z = (\varepsilon_0, \ldots, \varepsilon_{T-1})$ and thus $z \sim \mathcal{N}(0, I_{T \times T})$. Now considering the objective function in (2):

$$J = \mathbb{E}_{x_T} \mathbb{1}_{\{x \leq b\}}(x_T) = \mathbb{E}_z \mathbb{1}_{\{Az \leq b\}}(z) = \mathbb{P}_z(Az \leq b) = \int_{x \in \mathscr{P}} \phi(x) dx$$

■

Proposition 1 shows that computing integrals of Gaussian functions over polytopes is equivalent to solving a DP problem. Later we will discuss how to find an approximate solution to DP problems of the form (4).

IV. MARKOV DECISION PROCESSES

In this section state the properties of the class of MDP's we are interested in.

A. Markov Decision Processes

In this chapter we follow closely the notation and definitions of [17].

Definition 1: We say $\mathcal{M}$ is a finite time horizon Markov Decision Process (MDP) if it is a six tuple $\mathcal{M} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c, h), T)$ such that the following hold,

- X is a locally compact Borel space, with metric d_X, representing the state space. $\{X_t\}_{t \in \mathbb{N}}$ is a family of locally compact Borel subsets of X representing the state space at time t.
- $\mathscr{A}$ is a locally compact Borel space with metric d_A representing the set of admissible inputs.
- ψ is a map $X \to \mathscr{A}^\cup$ such that for each $x \in X$, $\psi(x)$ is a measurable subset of $\mathscr{A}$ representing the set of feasible controls that can be used at state $x \in X$. We suppose $\mathbb{K}_t = \{(x, a) : x \in X_t, a \in \psi(x)\}$ and $\mathbb{K} = \{(x, a) : x \in X, a \in \psi(x)\}$ are measurable subsets of $X \times \mathscr{A}^\cup$. (Note if $\mathscr{A} = \emptyset$ then we define $\mathbb{K} = X$ and $\mathbb{K}_t = X_t$).
- $\{Q_t\}_{t \in \mathbb{N}}$ is a family of stochastic kernels. That is, for $B \in \mathscr{B}(X_{t+1})$ the map $B \to Q(B|x, a)$ is a probability measure on $(X_{t+1}, \mathscr{B}(X_{t+1}))$ for all $(x, a) \in \mathbb{K}_t$, and $(x, a) \to Q(B|x, a)$ is a measurable function on $\mathbb{K}$ for every $B \in \mathscr{B}(X_{t+1})$. When $\mathscr{A} = \emptyset$, we simplify our notation by $Q_t(B|x) = Q_t(B|x, a)$. We denote the Lebesgue integral $\int_B f(y) Q_t(dy|x, a) := \int_B f(y) d\mu$ where μ is the induced measure created by the stochastic kernel, $\mu(B) = Q_t(B|x, a)$.
- $c : \mathbb{K} \to \mathbb{R}$ is a measurable function representing the cost per stage and $h : X \to \mathbb{R}$ is a measurable function representing the terminal cost.
- $T \in \mathbb{N}$ with $T < \infty$, representing the terminal time step.

Furthermore we denote $\mathbb{M}$ to be the set of all finite time horizon MDP's.

Definition 2: Consider a MDP $\mathcal{M} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c, h), T) \in \mathbb{M}$. We define a policy to be a sequence of maps $\pi = \{\pi_t\}_{t \in [T-1]}$ such that $\pi_t : X_t \to \mathscr{A}$ and for all $t \in [T-1]$ $\pi_t(x) \in \psi(x) \; \forall x \in X_t$. We denote the space of policies for the MDP $\mathcal{M}$ by $\Pi_{\mathcal{M}}$.

Definition 3: For every MDP $\mathcal{M} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c, h), T) \in \mathbb{M}$ we can define its associated optimization problem, $\mathscr{L}_{\mathcal{M}}(x_0)$.

$$\min_{\pi \in \Pi_{\mathcal{M}}} G_{\mathcal{M}}(x_0, \pi) := \mathbb{E}_x \Big[\sum_{t=0}^{T-1} c(x_t, \pi_t(x_t)) + h(x_T) \Big] \text{ Given,}$$

$$\mathbb{P}_{x_{t+1}}(x_{t+1} \in B | x_t = x, a = \pi_t(x_t)) = Q_t(B|x, a), \; B \in \mathscr{B}(X_{t+1}),$$

$$x(0) = x_0,$$

where $G_{\mathcal{M}}(x_0, \pi)$ denotes the expected cost for the policy $\pi \in \Pi_{\mathcal{M}}$ and initial condition $x_0 \in X_0$ associated with $\mathcal{M}$.

Definition 4: Consider a MDP, $\mathcal{M} \in \mathbb{M}$. The optimal total expected cost, $G^* : X_0 \to \mathbb{R}$ is defined by $G^*_{\mathcal{M}}(x) = \inf_{\pi \in \Pi_{\mathcal{M}}} G_{\mathcal{M}}(x, \pi)$ for $x \in X_0$. We define $\pi^* \in \Pi_{\mathcal{M}}$ to be the optimal policy if $G_{\mathcal{M}}(x, \pi^*) = G^*_{\mathcal{M}}(x)$ for any $x \in X_0$.

Commonly the associated optimization problem for an MDP is solved using a method called dynamic programing where Bellman's equation, which we will define in the next definition, is recursively solved backwards in time.

Definition 5: For a MDP $\mathcal{M} \in \mathbb{M}$ we define the optimal cost to go function (OCTGF) $J_{\mathcal{M}, t} : X \to \mathbb{R}$ recursively as:

$$J_{\mathcal{M}, T}(x) = h(x) \quad x \in X_T \tag{5}$$

$$J_{\mathcal{M}, t}(x) = \inf_{a \in \psi(x)} \Big\{ c(x, a) + \int_{X_{t+1}} J_{\mathcal{M}, t+1}(y) Q_t(dy|x, a) \Big\}$$

$$x \in X_t, \quad t \in [T-1]$$

Proposition 2: For any MDP $\mathcal{M} \in \mathbb{M}$, if $J_{\mathcal{M}, t}(x)$ is the associated OCTGF and $G^*_{\mathcal{M}}(x)$ is the optimal expected cost, then $J_{\mathcal{M}, 0}(x) = G^*_{\mathcal{M}}(x)$ for all $x \in X_0$. Moreover, for every t, there exists $f_t : X \to \mathscr{A}$ such that

$$J_{\mathcal{M}, t}(x) = c(x, f_t(x)) + \int_{X_{t+1}} J_{\mathcal{M}, t+1}(y) Q_t(dy|x, f_t(x)) \}.$$

f_t then defines the optimal policy as $\pi^* = \{f_t\}_{t \in [T-1]}$.

Proof: See [17]. ■

Corollary 1: There exists $\mathcal{M} \in \mathbb{M}$ such that the associated optimization problem, $\mathscr{L}_{\mathcal{M}}(0)$, is equivalent to (4).

Proof: We propose an MDP $\mathcal{M} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c, h), T) \in \mathbb{M}$ with an associated optimization problem equivalent to (4). We define the elements of $\mathcal{M}$ as follows,

- $X = \mathbb{R}^m$ and $X_t = X$ for all $t \in \mathbb{N}$. (6)
- $\mathscr{A} = \emptyset$. (7)
- $\psi(x) = \emptyset$ for all $x \in X$. (8)
- We can define the family of stochastic kernels for $B \in \mathscr{B}(\mathbb{R}^m)$ and $x \in \mathbb{R}^m$,

$$Q_t(B|x) = \int_{-\infty}^{\infty} \mathbb{1}_B(x + A_{\cdot, t+1} y) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-y^2}{2}\right) dy \tag{9}$$

Where $A = [A_{\cdot, 1}, \cdots, A_{\cdot, T}] \in \mathbb{R}^{m \times T}$ as in (4).
- The cost per stage and the terminal cost is,

$$c(x) = 0, \quad h(x) = \mathbb{1}_{\{x < b\}}(x). \tag{10}$$

Where $b \in \mathbb{R}^m$ is from (4).
- The finite time horizon is given to be,

$$T < \infty \quad \text{is as defined in Eq. (4).} \tag{11}$$

■

B. Readily-aproximable MDP's

Next we introduce similar properties of MDP's that [16] approximates, however we allow for discontinuity in the terminal cost function and require Property 8.

Definition 6: We say a six tuple $\mathcal{M} = (((\{X_t\}_{t\in\mathbb{N}},X),\mathcal{A},\psi,\{Q_t\}_{t\in\mathbb{N}},(c,h),T)$ is an **approximable MDP** or $\mathcal{M} \in \mathbb{A} \subset \mathbb{M}$ if $\mathcal{M}$ satisfies the following Properties 1-8.

Property 1: $X = \mathbb{R}^m$.

Property 2: $\psi(x)$ is compact for all $x \in X$.

Property 3: The map $\psi: X \to \mathcal{A}^{\cup}$ is Lipschitz continuous with respect to the Hausdorff norm. So $d_H(\psi(x),\psi(y)) \leq L_\psi d_X(x,y)$ for some constant $L_\psi > 0$.

Property 4: The cost function, $c : \mathbb{K} \to \mathbb{R}$, is Lipschitz continuous on $\mathbb{K}$. The terminal cost function, $h : X \to \mathbb{R}$, can be written in the form $h(x) = h_1(x) + h_2(x)$ where h_1 is a Lipschitz continuous and h_2 is of the form $h_2(x) = \begin{cases} f_1(x) & x \leq b \\ f_2(x) & x > b \end{cases}$. Where $b \in X$ and f_1 and f_2 are bounded and Lipschitz continuous (we note h_1 is not necessarily bounded). Furthermore there exists a positive lower semicontinuous function $w : X \to \mathbb{R}$ and a positive constant $\bar{c} > 0$ such that

$$|h_1(x)| + \|h_2\|_\infty + \sup_{a\in\psi(x)}|c(x,a)| < \bar{c}w(x). \quad (12)$$

Before we proceed to Property 5 we will introduce some additional notation. Given a function $v : X \to \mathbb{R}$ and an MDP $\mathcal{M} = (((\{X_t\}_{t\in\mathbb{N}},X),\mathcal{A},\psi,\{Q_t\}_{t\in\mathbb{N}},(c,h),T) \in \mathbb{M}$, we define $\zeta_{v,t}^{\mathcal{M}} : \mathbb{K}_t \to \mathbb{R}$ by,

$$\zeta_{v,t}^{\mathcal{M}}(x,a) = \int_{X_{t+1}} v(y)Q_t(dy|x,a). \quad (13)$$

We note for the MDP with tuple elements defined (6) to (11) we have $\zeta_{v,t}^{\mathcal{M}}(x) = \int_{\mathbb{R}} v(x+A_{\cdot,t+1}w)\phi(w)dw$ for $x \in \mathbb{R}^m$.

Property 5: There exists w satisfying (12) such that $\zeta_{w,t}^{\mathcal{M}}(x,a)$ is upper continuous on $\mathbb{K}_t$. In addition there exists $\bar{d} > 0$ such that $\zeta_{w,t}^{\mathcal{M}}(x,a) \leq \bar{d}w(x)$ for all $(x,a) \in \mathbb{K}_t$.

Property 6: For every bounded and continuous function v on X, the MDP $\mathcal{M}$ has the property that the induced function $\zeta_{v,t}^{\mathcal{M}}$ is continuous on $\mathbb{K}$ for each $t \in \mathbb{N}$.

Property 7: There exists a constant $L_q > 0$ such that $\forall t \in \mathbb{N}$, (x,a) and (y,b) in $\mathbb{K}_t$ and for any Lipschitz continuous function $v : X \to \mathbb{R}$ with Lipschitz constant $L_v > 0$:

$$|\zeta_{v,t}^{\mathcal{M}}(x,a) - \zeta_{v,t}^{\mathcal{M}}(y,b)| < L_q L_v |d_X(x,y) + d_A(a,b)|$$

Property 8: Consider $w : X \to \mathbb{R}$ and $b \in \mathbb{R}^m$ as in Property 4, then for all $\theta > 0$ there exists an $\Lambda \in \mathbb{R}$ such that, $|sup_{x\in X, u\in\psi(x)}\int_{y\in\mathcal{R}_{\lambda,b}} w(y)Q_{T-1}(dy|x,u)| < \theta$ for all $\lambda > \Lambda$. Next we will prove a Lemma showing the MDP associated with tuple elements defined (6) to (11) has Property 8. Then in the next proposition we will show the MDP is in $\mathbb{A}$.

Lemma 2: Consider the stochastic kernel, of the MDP with tuple elements 6 to 11,

$$Q_t(B|x) = \int_{-\infty}^{\infty} \mathbb{1}_B(x+A_{\cdot,t+1}y)\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-y^2}{2}\right)dy$$

Where $B \in \mathcal{B}(\mathbb{R}^m)$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times T}$, then for all $\lambda > 0$ we have,

$$\left|\sup_{x\in\mathbb{R}^m}\int_{y\in\mathcal{R}_{\lambda,b}} Q_{T-1}(dy|x)\right| < \frac{m}{\min_{1\leq i\leq m}\{|a_{i,T}|\}\lambda}. \quad (14)$$

Proof: By Lemma 1 we have $a_{i,T} \neq 0$ $\forall i \in \{1,..,m\}$. For some $x \in \mathbb{R}^m$ and $\lambda > 0$ we have,

$$\int_{y\in\mathcal{R}_{\lambda,b}} Q_{T-1}(dy|x) = \mathbb{P}_{\varepsilon_{T-1}}(x+A_{\cdot,T}\varepsilon_{T-1} \in \mathcal{R}_{\lambda,b}) \quad (15)$$

$$= \mathbb{P}_{\varepsilon_{T-1}}\left(\cup_{1\leq i\leq m}\left\{x_i + a_{i,T}\varepsilon_{T-1} \in (b_i - \frac{1}{\lambda},b_i)\right\}\right)$$

$$\leq \sum_{i=1}^m \mathbb{P}_{\varepsilon_{T-1}}\left(\varepsilon_{T-1} \in \frac{1}{a_{i,T}}(b_i - \frac{1}{\lambda} - x_i,b_i - x_i)\right).$$

Where $\varepsilon_{T-1} \sim \mathcal{N}(0,1)$. For $i \in \{1,...,m\}$ let us consider the function $f_i : \mathbb{R} \to [0,1]$ defined by,

$$f_i(x) = \mathbb{P}_{\varepsilon_{T-1}}\left(\varepsilon_{T-1} \in \frac{1}{a_{i,T}}(b_i - \frac{1}{\lambda} - x,b_i - x)\right)$$

$$= \int_{\frac{b_i-\frac{1}{\lambda}-x}{a_{i,T}}}^{\frac{b_i-x}{a_{i,T}}} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-w^2}{2}\right)dw.$$

It can be shown $x^* = b_i - \frac{1}{2\lambda}$ is the point at which f_i attains its maximum. Now,

$$f_i(x^*) = \int_{-\frac{1}{2a_{i,T}\lambda}}^{\frac{1}{2a_{i,T}\lambda}} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-w^2}{2}\right)dw \leq \int_{-\frac{1}{2a_{i,T}\lambda}}^{\frac{1}{2a_{i,T}\lambda}} dw$$

$$= \frac{1}{|a_{i,T}|\lambda} \leq \frac{1}{\min_{1\leq i\leq m}|a_{i,T}|\lambda}.$$

Now by substituting this into (15) we derive (14). ∎

Proposition 3: Let us denote the MDP with tuple elements defined (6) to (11) by $\mathcal{M}$, then $\mathcal{M} \in \mathbb{A}$.

Proof: To show $\mathcal{M} \in \mathbb{A}$ we will show $\mathcal{M}$ satisfies Properties 1-8.

Property 1: True since $X = \mathbb{R}^m$.

Properties 2 and 3: $\emptyset$ is compact and $\psi(x) = \emptyset$ $\forall x \in X$, moreover it follows $d_H(\psi(x),\psi(y)) = 0$ for all $x,y \in X$.

Property 4: $c(x,a) \equiv 0$, $h_1(x) \equiv 0$ and $h_2(x) = \begin{cases} 1 & x \leq b \\ 0 & x > b \end{cases}$.

We can trivially select $w(x) \equiv 1$ in this case.

Property 5: The probability measure of the entire state space is 1. $\zeta_{w,t}^{\mathcal{M}}(x,a) = \int_X Q_t(dy|x,a) = 1 = w(x)$.

Property 6: Consider continuous and bounded function $v : \mathbb{R}^m \to \mathbb{R}$ and let $C = \|v\|_\infty$. Let us denote $\phi(u) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right)$. We can use Dominated Convergence Theorem (DCT) to show $\zeta_{v,t}^{\mathcal{M}}(x) = \int_{y\in\mathbb{R}^m} v(y)Q_t(dy|x) = \int_{-\infty}^{\infty} v(x+A_{t+1}u)\phi(u)du$ is continuous with respect to x. Suppose $\lim_{n\to\infty}x_n = x$ and let $g_n(u) = v(x_n+A_{t+1}u)\phi(u)$. Since v is continuous clearly $\lim_{n\to\infty}g_n = g = v(x+A_{t+1}u)\phi(u)$. Now $g_n(u) \leq \sup_x|v(x+A_{t+1}u)|\phi(u) \leq C\phi(u)$. Thus $g_n(u)$ is dominated by some integrable function $C\phi(u)$ ($\int|C\phi(u)|du = C < \infty$) and tends point-wise to $g(u)$. It follows by DCT $\lim_{n\to\infty}\int g_n(u)du = \int g(u)du$, showing $\zeta_{v,t}^{\mathcal{M}}(x)$ is continuous.

Property 7: We will show $L_q = 1$. Suppose v is a Lipschitz

continuous function.

$$\left|\zeta_{v,t}^{\mathscr{M}}(x,a) - \zeta_{v,t}^{\mathscr{M}}(y,b)\right|$$

$$= \left|\int_{-\infty}^{\infty}[v(x+A_{\cdot,t+1}w) - v(y+A_{\cdot,t+1}w)]\phi(w)dw\right|$$

$$\leq L_v\int_{-\infty}^{\infty}|(x+A_{\cdot,t+1}w) - (y+A_{\cdot,t+1}w)|\phi(w)dw$$

$$= L_v|x-y|$$

<u>Property 8</u>: The result follows from Lemma 2. ∎

V. Approximating MDP's

Given $\mathscr{M} \in \mathbb{A}$ our approximation scheme has two stages; smoothing and discretization. During the smoothing stage the terminal cost function of the MDP is approximated with a Lipschitz continuous function. During the discretization stage the state and control spaces are approximated with compact spaces and then further approximated to countable sets.

A. Smoothing

For any MDP $\mathscr{M} \in \mathbb{A}$ we will show how to use the function $g_{\lambda,b}(x)$ to construct a sequence of MDP's with smooth terminal cost function and OCTGF's that converge to the OCTGF of $\mathscr{M}$ under the supremum norm.

Definition 7: Consider an approximable MDP $\mathscr{M} = (({\{X_t\}_{t\in\mathbb{N}}},X),\mathscr{A},\psi,{\{Q_t\}_{t\in\mathbb{N}}},(c,h),T) \in \mathbb{A}$. By Property 4 we can write $h(x) = h_1(x) + h_2(x)$ where h_1 is Lipschitz continuous and $h_2(x) = \begin{cases} f_1(x) & x \leq b \\ f_2(x) & x > b \end{cases}$.
Let us define the smooth function $\tilde{h}_\lambda(x;b;\mathscr{M}) = f_1(x)g_{\lambda,b}(x) + f_2(x)(1 - g_{\lambda,b}(x))$. We call the MDP $\tilde{\mathscr{M}}_\lambda = (({\{X_t\}_{t\in\mathbb{N}}},X),\mathscr{A},\psi,{\{Q_t\}_{t\in\mathbb{N}}},(c,h_1 + \tilde{h}_\lambda),T)$ the λ-smoothed MDP of $\mathscr{M}$. Furthermore we define the map $\Phi_1 : \mathbb{A} \times \mathbb{R}^+ \to \mathbb{A}$ by $\Phi_1(\mathscr{M},\lambda) = \tilde{\mathscr{M}}_\lambda$.
Next we will show that the terminal cost function of the λ-smoothed MDP is Lipschitz continuous.

Corollary 2: The function $\tilde{h}_\lambda : X \to \mathbb{R}$ defined by $\tilde{h}_\lambda(x;b,\mathscr{M}) = f_1(x)g_{\lambda,b}(x) + f_2(x)(1 - g_{\lambda,b}(x))$, where f_1 and f_2 are any bounded Lipschitz functions, is Lipschitz continuous with Lipschitz constant $L_{\tilde{h}_\lambda} = [L_{f_1} + L_{f_2} + 2\lambda m \max\{||f_1||_\infty,||f_2||_\infty\}]$. Where $m=\dim(X)$.

Remark 2: The image of the map Φ_1 is a subset of $\mathbb{A}$. Furthermore for any $\mathscr{M} \in \mathbb{A}$ and $\lambda > 0$ there exists a function $w : X \to \mathbb{R}$ such that both $\mathscr{M}$ and $\phi_1(\mathscr{M},\lambda)$ satisfy Property 4 using w.

In the next lemma we will give the Lipschitz properties of the OCTGF of a λ-smoothed MDP.

Lemma 3: For some $\lambda > 0$ consider the OCTGF's J_t and $\tilde{J}_t$ of the MDP's $\mathscr{M} \in \mathbb{A}$ and $\tilde{\mathscr{M}} = \Phi_1(\mathscr{M},\lambda)$ respectively. Then $||J_t||_w < \infty$ and $\tilde{J}_t \in \mathbb{L}_w(X)$, where $w : X \to \mathbb{R}$ is as in Property 4 for $\mathscr{M}$. Furthermore,

$$L_{\tilde{J}_t} = [L_c + L_q L_{\tilde{J}_{t+1}}][1 + L_\psi] \tag{16}$$

$$L_{\tilde{J}_T} = L_{\tilde{h}_\lambda} + L_h.$$

Proof: See Lemma 2.5 in [16]. ∎

Corollary 3: Consider the OCTGF, $\tilde{J}_t$, of a MDP $\tilde{\mathscr{M}} = \Phi_1(\mathscr{M},\lambda)$ for some $\mathscr{M} \in \mathbb{A}$. Then its Lipschitz constant,

$L_{\tilde{J}_t} > 0$ satisfies,
$L_{\tilde{J}_t} = (L_q[1+L_\psi])^{T-t}[L_{f_1} + L_{f_2} + 2\lambda m \max\{||f_1||_\infty,||f_2||_\infty\} + L_h] + L_c[1+L_\psi]\sum_{i=1}^{T-t}(L_q[1+L_\psi])^{i-1} \quad \forall t \in [T]$.

The next Proposition proves that the OCTGF for a λ-smoothed MDP converges to the OCTGF of its corresponding approximable MDP under the supremum norm as $\lambda \to \infty$.

Proposition 4: Consider an MDP $\mathscr{M} = (({\{X_t\}_{t\in\mathbb{N}}},X),\mathscr{A},\psi,{\{Q_t\}_{t\in\mathbb{N}}},(c,h),T) \in \mathbb{A}$ and its corresponding λ-smoothed MDP $\tilde{\mathscr{M}} = (({\{X_t\}_{t\in\mathbb{N}}},X),\mathscr{A},\psi,{\{Q_t\}_{t\in\mathbb{N}}},(c,h_1 + \tilde{h}_\lambda),T) = \Phi_1(\mathscr{M},\lambda)$ with OCTGF's denoted by $J_t(x)$ and $\tilde{J}_t(x)$ respectively. Then for $\theta > 0$ there exists $\Lambda \in \mathbb{R}$ and $w : X \to \mathbb{R}$ such that $||J_t||_w < \infty$, $\tilde{J}_t \in \mathbb{L}_w(X)$ and for all $\lambda > \Lambda$ we have $\sup_{x\in X}\left|\tilde{J}_t(x) - J_t(x)\right| < (||J_T||_w + ||\tilde{J}_T||_w)\theta$ for any $t \in [T-1]$.

Proof: Consider $w : X \to \mathbb{R}$ as in Property 4 of $\mathscr{M}$ then by Lemma 3 $||J_t||_w < \infty$, $\tilde{J}_t \in \mathbb{L}_w(X)$. For $t \in [T-1]$ using Bellman's equation (5) we have,

$$\left|\tilde{J}_{t-1}(x) - J_{t-1}(x)\right| \tag{17}$$

$$\leq \left(\inf_{u\in\psi(x)}\sup_{a\in\psi(x)}\left|\int_{X_T}\tilde{J}_t(y)Q_{t-1}(dy|x,u) - \int_{X_T}J_t(y)Q_{t-1}(dy|x,a)\right|\right.$$

$$\left.\vee \inf_{a\in\psi(x)}\sup_{u\in\psi(x)}\left|\int_{X_T}\tilde{J}_t(y)Q_{t-1}(dy|x,u) - \int_{X_T}J_t(y)Q_{t-1}(dy|x,a)\right|\right)$$

$$+ \left(\inf_{u\in\psi(x)}\sup_{a\in\psi(x)}|c(x,u) - c(x,a)| \vee \inf_{a\in\psi(x)}\sup_{u\in\psi(x)}|c(x,u) - c(x,a)|\right)$$

We now proceed by downward induction starting at $t = T-1$. Let $\theta > 0$, by Property 8 of $\mathscr{M}$ $\exists\Lambda > 0$ such that $\forall\lambda > \Lambda$

$$\left|sup_{x\in X,u\in\psi(x)}\int_{y\in\mathscr{R}_{\lambda,B}}w(y)Q_{T-1}(dy|x,u)\right| < \theta$$

For $\lambda > \Lambda$ we see,

$$\left|\int_{X_T}\tilde{J}_T(y)Q_{T-1}(dy|x,u) - \int_{X_T}J_T(y)Q_{T-1}(dy|x,a)\right| \tag{18}$$

$$\leq \left|\zeta_{h_1+\tilde{h}_\lambda,T}^{\mathscr{M}}(x,u) - \zeta_{h_1+\tilde{h}_\lambda,T}^{\mathscr{M}}(x,a)\right|$$

$$+ \left|\int_{X_T}\tilde{J}_T(y)Q_{T-1}(dy|x,a) - \int_{X_T}J_T(y)Q_{T-1}(dy|x,a)\right|$$

$$\leq L_q(L_{h_1} + L_{\tilde{h}_\lambda})d_A(u,a)$$

$$+ \left|\int_{X_T}\tilde{J}_T(y)Q_{T-1}(dy|x,a) - \int_{X_T}J_T(y)Q_{T-1}(dy|x,a)\right|$$

$$\leq L_q(L_{h_1} + L_{\tilde{h}_\lambda})d_A(u,a) + \left|\int_{y\in\mathscr{R}_{\lambda,b}}\tilde{J}_T(y)Q_{T-1}(dy|x,a)\right|$$

$$+ \left|\int_{y\in\mathscr{R}_{\lambda,b}}J_T(y)Q_{T-1}(dy|x,a)\right|$$

$$+ \left|\int_{y\in X_T/\mathscr{R}_{\lambda,b}}\tilde{J}_T(y)Q_{T-1}(dy|x,a) - \int_{y\in X_T/\mathscr{R}_{\lambda,b}}J_T(y)Q_{T-1}(dy|x,a)\right|$$

$$\leq L_q(L_{h_1} + L_{\tilde{h}_\lambda})d_A(u,a)$$

$$+ (||J_T||_w + ||\tilde{J}_T||_w)\left|\int_{y\in\mathscr{R}_{\lambda,b}}w(y)Q_{T-1}(dy|x,a)\right|$$

$$+ \left|\int_{y\in X_T/\mathscr{R}_{\lambda,b}}[(h_1(y) + \tilde{h}_\lambda(y)) - h(y)]Q_{T-1}(dy|x,a)\right|$$

$$\leq L_q(L_{h_1} + L_{\tilde{h}_\lambda})d_A(u,a) + (||J_T||_w + ||\tilde{J}_T||_w)\theta.$$

Where the triangle inequality is used in the first and third inequality, Property 7 is used in the second inequality,

Property 4 is used in the fourth inequality and in the fifth inequality Property 8 and the fact $h(y) = h_1(y) + \tilde{h}_\lambda(y)$ $\forall y \in X/\mathscr{R}_{\lambda,b}$ is used.

Also by Property 4 of $\mathscr{M}$,

$$|c(x,u) - c(x,a)| \le L_c d_A(u,a). \tag{19}$$

Moreover,

$$\inf_{u \in \psi(x)} \sup_{a \in \psi(x)} d_A(u,a) \vee \inf_{a \in \psi(x)} \sup_{u \in \psi(x)} d_A(u,a) \tag{20}$$
$$= d_H(\psi(x), \psi(x)) = 0.$$

Thus it follows by substituting $t = T$ into (17) and further using (18), (19) and (20),

$$\sup_{x \in X} |\tilde{J}_{\lambda,T-1}(x) - J_{T-1}(x)| \le (||J_T||_w + ||\tilde{J}_T||_w)\theta.$$

Now we proceed by downward induction. Assuming the result to be true for $s+1$, $\exists \Lambda$ such that $\forall \lambda > \Lambda$ we have $\sup_{x \in X} |\tilde{J}_{s+1}(x) - J_{s+1}(x)| < (||J_T||_w + ||\tilde{J}_T||_w)\theta$. Now for $\lambda > \Lambda$,

$$\left| \int_{X_{s+1}} \tilde{J}_{s+1}(y) Q_s(dy|x,u) - \int_{X_{s+1}} J_{s+1}(y) Q_s(dy|x,a) \right| \tag{21}$$
$$\le \left| \zeta_{\tilde{J}_{s+1},s}^{\mathscr{M}}(x,u) - \zeta_{\tilde{J}_{s+1},s}^{\mathscr{M}}(x,a) \right| + \int_{X_{s+1}} |\tilde{J}_{s+1}(y) - J_{s+1}(y)| Q_s(dy|x,a)$$
$$\le L_P L_{\tilde{J}_{s+1}} d_A(u,a) + \sup_{y \in X} |\tilde{J}_{s+1}(y) - J_{s+1}(y)| \int_{X_{s+1}} Q_s(dy|x,a)$$
$$\le L_P L_{\tilde{J}_{s+1}} d_A(u,a) + (||J_T||_w + ||\tilde{J}_T||_w)\theta.$$

Where the first inequality uses the triangle rule, the second inequality uses Property 7 and the third inequality uses the induction hypothesis.

Thus it follows by substituting $t = s+1$ into (17) and further using (21), (19) and (20),

$$|\tilde{J}_s(x) - J_s(x)| \le (||J_T||_w + ||\tilde{J}_T||_w)\theta. \qquad \blacksquare$$

B. Discretization

In this section we show how to mathematically discretize a λ-smoothed readily-aporximable MDP. We will show the discretized MDP can be made arbitrarily close to the λ-smoothed MDP.

Definition 8: For any MDP $\tilde{\mathscr{M}} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c,h), T) \in \Phi_1(\mathbb{A}, \mathbb{R}^+)$ and $\varepsilon > 0$ we define $\mathbb{H}_{\tilde{\mathscr{M}},\alpha}$ as a set of compact subsets of X where we say $\{H_t\}_{t \in [T]} \in \mathbb{H}_{\tilde{\mathscr{M}},\alpha}$ if
- H_t is a compact subset of X_t for all $t \in [T]$ and
- $\sup_{(x,a) \in \mathbb{J}_{t-1}} \int_{X/H_t} w(y) Q_{t-1}(dy|x,a) \le \alpha$ for all $t \in [T]$, where $\mathbb{J}_{t-1} = \{(x,a) : x \in H_{t-1}, a \in \psi(x)\}$ and $w : X \to \mathbb{R}$ is as in Property 4 for some $\mathscr{M}$ such that $\tilde{\mathscr{M}} = \Phi_1(\mathscr{M},\lambda)$ for some $\lambda > 0$.

Lemma 4: $\forall \tilde{\mathscr{M}} \in \Phi_1(\mathbb{A}, \mathbb{R}^+)$ and $\alpha > 0$ $\mathbb{H}_{\tilde{\mathscr{M}},\alpha} \ne \emptyset$.

Proof: We note $\Phi_1(\mathbb{A}, \mathbb{R}^+)$ is a subset of $\mathbb{A}$ and thus the result of the Lemma follows from Lemma 2.9 [16]. $\blacksquare$

Consider $\alpha > 0$. For some $\{H_t\}_{t \in [T]} \in \mathbb{H}_{\tilde{\mathscr{M}},\alpha}$ let us denote the map $\Phi_2 : X^\cup \times \Phi_1(\mathbb{A}, \mathbb{R}^+) \to \mathbb{M}$ such

that for $\tilde{\mathscr{M}} = ((\{X_t\}_{t \in \mathbb{N}}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c,h), T) \in \Phi_1(\mathbb{A}, \mathbb{R}^+)$ we have $\Phi_2(\{H_t\}_{t \in [T]}, \tilde{\mathscr{M}}) = ((\{H_t\}_{t \in [T]}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c,h), T)$.

Proposition 5: Denote the MDP with tuple elements (6) to (11) by $\mathscr{M}$. For any $\lambda > 1$ let $\alpha = \frac{\sqrt{2}m \max|a_{i,t}|}{\sqrt{\log(\lambda)}} \left(\frac{1}{\lambda}\right)^{\frac{1}{\max a_{i,t}^2}}$ and consider the family of sets $H_t := \gamma_{t,\lambda}^m \subset \mathbb{R}^m$ where $\gamma_{t,\lambda} = [-t\sqrt{2\log(\lambda)}, t\sqrt{2\log(\lambda)}]$, then $\{H_t\}_{t \in [T]} \in \mathbb{H}_{\phi_1(\mathscr{M},\lambda),\alpha}$.

Proof: Clearly H_t is a compact subset of $\mathbb{R}^m$ $\forall t \in [T]$. Next using $w(x) \equiv 1$ that can be used in Property 4 of $\mathscr{M}$ we show,

$$\sup_{x \in H_{t-1}} \int_{X/H_t} Q_{t-1}(dy|x) < \alpha. \tag{22}$$

Recalling $\varepsilon \sim \mathcal{N}(0,1)$,

$$\sup_{x \in H_{t-1}} \int_{X/H_t} Q_{t-1}(dy|x) = \sup_{x \in H_{t-1}} \mathbb{P}_\varepsilon(x + A_{\cdot,t}\varepsilon \notin H_t) \tag{23}$$
$$\le \sup_{x \in H_{t-1}} \left\{ \mathbb{P}_\varepsilon \left(\cup_{i \in \{1,..,m\}} \left\{ x_i + a_{i,t}\varepsilon > t\sqrt{2\log(\lambda)} \right\} \right) \right.$$
$$\left. + \mathbb{P}_\varepsilon \left(\cup_{i \in \{1,..,m\}} \left\{ x_i + a_{i,t}\varepsilon < -t\sqrt{2\log(\lambda)} \right\} \right) \right\}$$
$$\le \sum_{i=1}^m \left\{ \sup_{x_i \in \gamma_{t-1,\lambda}} \mathbb{P}_\varepsilon \left(x_i + a_{i,t}\varepsilon > t\sqrt{2\log(\lambda)} \right) \right.$$
$$\left. + \sup_{x_i \in \gamma_{t-1,\lambda}} \mathbb{P}_\varepsilon \left(x_i + a_{i,t}\varepsilon < -t\sqrt{2\log(\lambda)} \right) \right\}$$

We will show,

$$\sup_{x \in \gamma_{t-1,\lambda}} \mathbb{P}_\varepsilon \left(x + a\varepsilon > t\sqrt{2\log(\lambda)} \right) \le \frac{|a|}{\sqrt{2\log(\lambda)}} \left(\frac{1}{\lambda}\right)^{\frac{1}{a^2}} \tag{24}$$

by considering the cases $a > 0$, $a < 0$ and $a = 0$ separately. For $a > 0$,

$$\sup_{x \in \gamma_{t-1,\lambda}} \mathbb{P}_\varepsilon \left(x + a\varepsilon > t\sqrt{2\log(\lambda)} \right) = \sup_{x \in \gamma_{t-1,\lambda}} \mathbb{P}_\varepsilon \left(\varepsilon > \frac{t\sqrt{2\log(\lambda)} - x}{a} \right)$$
$$\le \mathbb{P}_\varepsilon \left(\varepsilon > \frac{\sqrt{2\log(\lambda)}}{a} \right) < \frac{a}{\sqrt{2\log(\lambda)}} \exp\left(-\frac{\log(\lambda)}{a^2} \right)$$
$$\le \frac{a}{\sqrt{2\log(\lambda)}} \left(\frac{1}{\lambda}\right)^{\frac{1}{a^2}}.$$

Where the second inequality uses Lemma 5 and the last inequality follows since we can assume $\lambda > 1$. The case $a < 0$ follows by a similar proof. The case $a = 0$ is trivial, $\sup_{x \in \gamma_{t-1,\lambda}} \mathbb{P}(x + a\varepsilon > t\sqrt{2\log(\lambda)}) = 0$. A similar argument of considering the different cases of a can show

$$\sup_{x \in \gamma_{t-1,\lambda}} \mathbb{P}(x + a\varepsilon < -t\sqrt{2\log(\lambda)}) \le \frac{|a|}{\sqrt{2\log(\lambda)}} \left(\frac{1}{\lambda}\right)^{\frac{1}{a^2}}. \tag{25}$$

Now, substituting (24) and (25) into (23), we get (22). $\blacksquare$

Next we return to a general λ-smoothed MDP and approximate its state and control space's with a countable set.

Definition 9: Given an approximable MDP $\mathscr{M} \in \mathbb{A}$ we can define a corresponding MDP $\tilde{\mathscr{M}} = ((\{H_t\}_{t \in [T]}, X), \mathscr{A}, \psi, \{Q_t\}_{t \in \mathbb{N}}, (c,\tilde{h}), T) = \Phi_2(\{H_t\}_{t \in [T]}, \Phi_1(\mathscr{M},\lambda))$ for some compact family $\{H_t\}_{t \in [T]}$ and $\lambda > 0$. Furthermore given

- Γ_t is a β-partition of H_t.
- $\Theta_t(x)$ is a η-partition of $\psi(x)$.
- $\hat{Q}$ defined for $x \in \Gamma_t$, $y \in \Gamma_{t+1}$ and $a \in \Theta_t(x)$ as $\hat{Q}_t(y|x,a) := Q_t(p_{H_{t+1},\Gamma_{t+1}}^{-1}(y)|x,a)$ for $t \in [T-1]$ where $p_{H_{t+1},\Gamma_{t+1}}^{-1}(y)$ is the pre-image of y,

we define the map $\Phi_3 : \mathbb{R}^+ \times \mathbb{R}^+ \times Im\{\Phi_2\} \to \mathbb{M}$ by $\hat{\mathscr{M}} = \Phi_3(\beta,\eta,\tilde{\mathscr{M}})$ if $\hat{\mathscr{M}} = ((\{\Gamma_t\}_{t\in[T]},X),\mathscr{A},\Theta_t,\hat{Q},(c,\tilde{h}),T)$.

Note if $\psi(x) = \emptyset$ $\forall x \in X$ we simplify our notation and set $\eta = 0$.

Definition 10: We define the total approximation map $\Psi : \mathbb{A} \times (\mathbb{R}^+)^3 \times (X^\cup)^T \to \mathbb{M}$ by $\Psi(\mathscr{M},\lambda,\beta,\eta,\{H_t\}_{t\in[T]}) = \Phi_3(\beta,\eta,\Phi_2(\{H_t\}_{t\in[T]},\Phi_1(\mathscr{M},\lambda)))$.

Definition 11: Let $\hat{J}_t$ be the OCTGF of some MDP in $Im\{\Psi\}$ with state space $(\{\Gamma_t\}_{t\in[T]},X)$. For any compact set H_t such that $\Gamma_t \subset H_t$ we define the extended OCTGF of $\hat{J}_t(x)$ as follows:

$$\hat{F}_t(x) = \hat{J}_t(p_{H_t,\Gamma_t}(x)) \quad x \in H_t$$

Next we will state a theorem that gives a bound for error of the OCTGF's of the MDP's in $Im\{\Phi_1\}$ and associated MDP's mapped under Φ_3.

Theorem 1: Consider some $\mathscr{M} \in \mathbb{A}$. For some $\lambda > 0$ suppose $\tilde{J}_t$ is the OCTGF of $\tilde{\mathscr{M}} = \Phi_1(\mathscr{M},\lambda)$. For any $\alpha,\beta,\eta > 0$ there exists $\{H_t\}_{t\in[T]} \in \mathbb{H}_{\tilde{\mathscr{M}},\alpha}$ such that if we denote $\hat{J}_t$ as the extended OCTGF of $\Psi(\mathscr{M},\lambda,\beta,\eta,\{H_t\}_{t\in[T]})$ then for $t \in [T-1]$,

$$\sup_{x\in H_T} |\tilde{J}_T(x) - \hat{J}_T(x)| \le (L_h + L_{\tilde{h}_\lambda})\beta \qquad (26)$$

$$\sup_{x\in H_t} |\tilde{J}_t(x) - \hat{J}_t(x)| \le ||\tilde{J}_{t+1}||_w \alpha + (L_{\tilde{J}_{t+1}}L_q + L_c)\eta \qquad (27)$$

$$+ \sup_{y\in H_{t+1}} |\tilde{J}_{t+1}(y) - \hat{J}_{t+1}(y)| + L_{\tilde{J}_t}\beta$$

Where the function $w : X \to \mathbb{R}$ and Lipschitz constant L_q are as in Property 4 and Property 7 of $\mathscr{M}$ respectively.

Moreover in the case where the control space of $\mathscr{M}$ is empty we set $\eta = 0$.

Proof: See Theorem 3.4 in [16]. ∎

C. Error Bounds

In this section we will show how Theorem 1 can be combined with Proposition 4 to show that the OCTGF's of an MDP $\mathscr{M} \in \mathbb{A}$, and the approximated MDP $\hat{\mathscr{M}} = \Psi(\mathscr{M},\lambda,\beta,\eta,\{H_t\}_{t\in[T]})$ are arbitrary close together.

Theorem 2: Consider some MDP $\mathscr{M} \in \mathbb{A}$ with OCTGF denoted by J_t. For any $\alpha,\beta,\eta,\theta > 0$ there exists $\{H_t\}_{t\in[T]} \in \mathbb{H}_{\tilde{M},\alpha}$ and $\Lambda > 0$ such that for all $\lambda > \Lambda$ and any $t \in [T-1]$,

$$\sup_{x\in H_t} |J_t(x) - \hat{J}_t(x)| \le \theta\left\{||J_T||_w + ||\tilde{J}_T||_w\right\} + \alpha\left\{\sum_{i=1}^{T-t}||\tilde{J}_{t+i}||_w\right\} \qquad (28)$$

$$+ \eta\left\{L_q\left(\sum_{i=1}^{T-t}L_{\tilde{J}_{t+i}} + (T-t)L_c\right)\right\} + \beta\left\{L_h + L_{\tilde{h}_\lambda} + \sum_{i=1}^{T-t}L_{\tilde{J}_{t+i-1}}\right\}$$

Where $\tilde{J}_t$ and $h + \tilde{h}_\lambda$ is the OCTGF and terminal cost function of $\Phi_1(\mathscr{M},\lambda)$ respectively. $\hat{J}_t$ is the extended OCTGF of $\Psi(\mathscr{M},\lambda,\beta,\eta,\{H_t\}_{t\in[T]})$. The function $w : X \to \mathbb{R}$ and constant L_q are as in Properties 4 and 7 of $\mathscr{M}$ respectively.

Moreover in the case where the control space of $\mathscr{M}$ is empty we set $\eta = 0$.

Proof: By the triangle inequality,

$$|J_t(x) - \hat{J}_t(x)| \le |J_t(x) - \tilde{J}_t(x)| + |\tilde{J}_t(x) - \hat{J}_t(x)| \qquad (29)$$

We then use Proposition 4 to bound $|J_t(x) - \tilde{J}_t(x)|$. Then we recursively solve (26) and (27) in Theorem 4 to bound $|\tilde{J}_t(x) - \hat{J}_t(x)|$. Substituting these bounds into (29) the result (28) follows. ∎

We now specialize Theorem 2 to the MDP with tuple elements defined (6) to (11). Since in this specific case the control space, $\mathscr{A}$, is empty we can set $\eta = 0$ in Theorem 2.

Corollary 4: Consider the MDP with tuple elements defined (6) to (11) by $\mathscr{M}$. For any $\lambda > 1$ and $\beta > 0$ let $H_t = [-t\sqrt{2\log(\lambda)}, t\sqrt{2\log(\lambda)}]^m$ then the extended OCTGF, $\hat{J}_t$, of $\Psi(\mathscr{M},\lambda,\beta,0,\{H_t\}_{t\in[T]})$ satisfies,

$$\left|\int_{x\in\mathscr{P}}\phi(x)dx - \hat{J}_0(x_0)\right| \le \frac{2m}{\min_{i\in[m]}\{|a_{i,T}|\}\lambda} \qquad (30)$$

$$+ \frac{\sqrt{2}mT\max|a_{i,t}|}{\sqrt{\log(\lambda)}}\left(\frac{1}{\lambda}\right)^{\frac{1}{\max a_{i,t}^2}} + 2m\lambda\beta(T+1).$$

where $x_0 = (0,..,0)$, $\mathscr{P} = \{x \in \mathbb{R}^T : Ax < b\}$ and $A \in \mathbb{R}^{m\times T}$ and $b \in \mathbb{R}^m$.

Proof: By Proposition 3 $\mathscr{M} \in \mathbb{A}$, thus $\hat{\mathscr{M}}$ is well defined. Let us denote the OCTGF of the MDP's $\mathscr{M}$ and $\tilde{\mathscr{M}} = \Phi_1(\mathscr{M},\lambda)$ by J_t and $\tilde{J}_t$. By Proposition 1 $J_0(x_0) = \int_{\mathbf{x}\in\mathscr{P}}\phi(\mathbf{x})d\mathbf{x}$. Using Corollary 3 and Corollary 2 $L_{\tilde{J}_t}$ and $L_{\tilde{h}}$ can be calculated. Proposition 5 shows $H_t \in \mathbb{H}_{\tilde{\mathscr{M}},\alpha}$ where $\alpha = \frac{\sqrt{2}m\max|a_{i,t}|}{\sqrt{\log(\lambda)}}\left(\frac{1}{\lambda}\right)^{\frac{1}{\max a_{i,t}^2}}$. Now Theorem 2 can be applied to the specific MDP $\mathscr{M}$, where Lemma 2 is used to select an appropriate θ; and using induction and (5) it can be shown $||\tilde{J}_t||_w \le 1$ $\forall t \in [T]$. ∎

VI. NUMERICAL RESULTS

Using the smoothing and discretization procedure laid out in section V we find approximate solutions to the optimization problem associated with the MDP with tuple elements defined (6) to (11). The algorithm recursively solves (5) for different discretization parameters, (λ,β). In all simulations (λ,β) were parametrized by $n \in \mathbb{N}$; we selected $\lambda = \sqrt{n}$ and $\beta = \frac{1}{n}$. Figure 1 shows the results of computing the probability that a two dimensional Gaussian variable is in the positive orthant; this can be written as an integral of the form (3) where $A = [1,1]^T$ and $b = [0,0]^T$. Using (30) it can be shown the error bound for integration over the positive orthant is $E = \frac{28}{\sqrt{n}} + \frac{4}{\sqrt{n}\log n}$, which is of order $\mathscr{O}\left(\frac{1}{\sqrt{n}}\right)$. The order of the actual error of the algorithm, when compared to the true value of 0.5, seems to also be $\mathscr{O}\left(\frac{1}{\sqrt{n}}\right)$; indicating our error bounds are tight in some cases.

In Figure 2 we evaluate an integral of the Form (3) where $A = \begin{bmatrix} 0.5 & 0.7 & 1 & 0.9 \\ 0.2 & 0.7 & 0.5 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 0.5 \end{bmatrix}$. The horizontal line represents the Monte Carlo approximation of 10^8 samples. The curved line represents the OCTGF, $\hat{J}_0(0)$, of the approximated MDP with tuple elements (6) to (11).
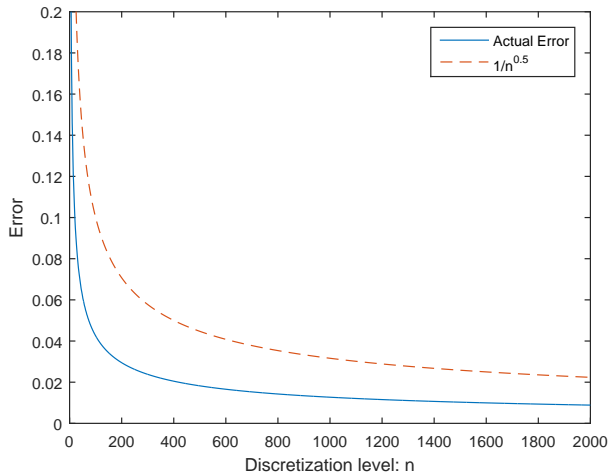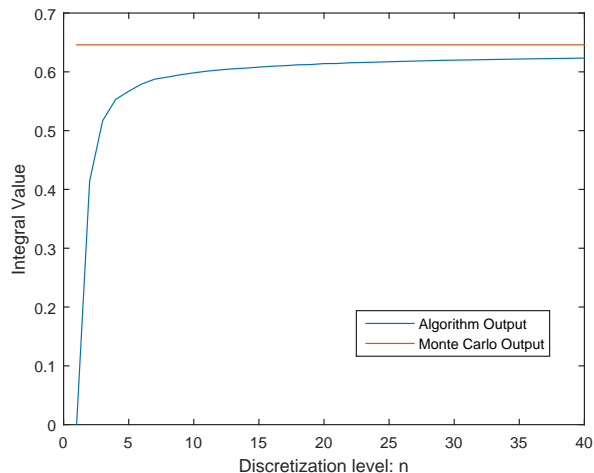
Fig. 1.

Fig. 2.

VII. Conclusion

In this paper we showed that given a multivariate Gaussian integral over a polytope it is possible to construct an MDP such that the solution of the MDP's associated optimization problem is equal to the integral. In general this class of MDP's have non-compact uncountable state spaces and discontinuous terminal cost functions. In this paper we use Bellman's equation to solve the associated optimization problem. However in general there is no analytical solution to Bellman's equation for MDP's of this class and thus an approximation is required. We proposed an approximation scheme that maps our class of MDP's to a much simpler class of MDP's with countable state and control spaces. Moreover we derived bounds on the supremum norm error of the optimal cost to go functions of the MDP and the mapped MDP. The main contribution of this paper is thus a dynamic programing based algorithm for evaluating multivariate Gaussian integration over polytopes with a priori error bounds.

Our numerical results presented in section VI are consistent our error bounds in section V. There are substantial computational costs to this dynamic programing approach but using this approach we are able to compute the integral to any degree of accuracy. This paper links computing multivariate Gaussian integration over polytopes to dynamic programing; a well developed computational technique.

VIII. Appendix

Lemma 5: For $t > 0$ $\mathbb{P}_\varepsilon(\varepsilon > t) \leq \frac{1}{t}\exp(\frac{-t^2}{2})$ where $\varepsilon \sim \mathcal{N}(0,1)$.

$Proof:$ $\mathbb{P}_\varepsilon(\varepsilon > t) = \int_t^\infty \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)dx$

$$< \int_t^\infty \frac{x}{t}\exp\left(\frac{-x^2}{2}\right)dx$$

$$= \frac{1}{t}\exp\left(\frac{-t^2}{2}\right)$$

Where the second inequality uses the fact that inside the integral domain $x \geq t$ and $\frac{1}{\sqrt{2\pi}} < 1$. ∎

References

[1] H. L. Xuejun Lio and L. Carin, "Quadratically gated mixture of experts for incomplete data classification," *International conference of machine learning*.

[2] D. van Hessem and O.H.Bosgra, "Closed-loop stocastic dynamic process optimization under input and state constraints," *ACC*, 2002.

[3] C. S. D.H. van Hessem and O.H.Bosgra, "Lmi-based closed-loop economic optimizatiion of stochastic process operation under state and input constraints," *Selected topics in Signals, Systems and Control*, vol. 12, 2001.

[4] A. Prekopa and T. Szantai, "Flood control reswevoir system design using stochastic programing," *Mathematical Programing Study*, vol. 9, pp. 138–151, 1978.

[5] R. D. Bock and R. D.Gibbons, "High-dimensional multivariate probit analysis," *International Biometric Society*, vol. 52, pp. 1183–1194, 1996.

[6] J. P. Cunningham, P. Hennig, and S. Lacoste Julien, "Gaussian probabilities and expectation propagation," *Journal of Machine Learning Research*, vol. 1, p. 1, 2011.

[7] A. Genz, "Numerical computation of multivariate normal probabilities," *J. Comp. Graph Stat.*, vol. 1, pp. 141–149, 1992.

[8] I. d. H.I. Gassmann and T. Szantai, "Computing multivariate normal probabilities: A new look," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 920–949, 2002.

[9] U. D. Hanebeck and M. Dolgov, "Adaptive lower bounds for gaussian measures of polytopes," *18th International Conference on Information Fusion*, 2015.

[10] W. P. Vijverberg, "Monte carlo evaluation of multivariate normal probabilities," *Journal of Econometrics*, vol. 76, pp. 281–307, 1997.

[11] A. Arapostathis, V. S. Borkar, and E. Fernandez-Gaucherand, "Discrete-time controlled markov processes with average cost criterion: A survey," *Siam J. Control and Optimization*, vol. 31, pp. 282–344, 1993.

[12] R. E. Bellman, R. E. Kalaba, and T. Teichmann, "Dynamic programing," *Physics Today*, vol. 19, p. 98, 1966.

[13] M. Jones and M. Peet, "Solving dynamic programming with supremum terms in the objective and application to optimal battery scheduling for electricity consumers subject to demand charges," *CDC*, 2017.

[14] D. P. Bertsekas, "Convergence if discretization procedures in dynamic programing," *IEEE S-CS*, p. 415, 1974.

[15] O. Hernandez-Lerma and J. Lasserre, "Adaptive markov control processes," *Appl. Math. Sci*, vol. 79, 1989.

[16] F. Dufour and T. Prieto-Rumeau, "Approximation of markov decision processes with general state space," *Journal of Mathematical Analysis and Applications*, vol. 388, pp. 1254–1267, 2012.

[17] O. Hernandez-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, vol. 42. 1999.

[18] J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica*, vol. 65, pp. 487–516, 1997.