Numerical Approximations for Average Cost Markov Decision Processes

François Dufour [1] Tomás Prieto-Rumeau [2]

[1]INRIA, Bordeaux, France

[2]UNED, Madrid, Spain

INRIA Team TAO Seminar
February 9, 2016

Numerical Approximations for Average Cost MDPs

Statement of the problem

- We are interested in approximating the optimal average cost and an optimal policy of a discrete-time Markov control process.
- We consider a control model with general state and action spaces.
- Most of the approximation results in the literature are concerned with MDPs with discrete state and action spaces.

Our approach

- We propose procedures to discretize the state and action spaces.
- Discretization of the state space is based on sampling an underlying probability measure.
- Discretization of the action space is made by selecting actions that are "dense" in the Hausdorff metric.
- We show that our approximation error converges in probability to zero at an exponential speed.

Dynamics of the control model

It is a stochastic controlled dynamic system.

- The system is in state x_0.
- The controller takes an action a_0 and incurs a cost $c(x_0, a_0)$.
- The system makes a transition $x_1 \sim Q(\cdot|x_0, a_0)$.
- The system is in state x_1. Etc.

On an infinite horizon we have:

- a state process: $\{x_t\}_{t\geq 0}$;
- an action process: $\{a_t\}_{t\geq 0}$;
- a cost process: $\{c(x_t, a_t)\}_{t\geq 0}$.

Definition of the control model

The control model $\mathcal{M}$

Consider a control model $(X, A, \{A(x) : x \in X\}, Q, c)$ where

- The state space X is a Borel space, with metric ρ_X.
- The action space A is a Borel space, with metric ρ_A.
- $A(x)$ is the measurable set of available actions in state $x \in X$.
- $Q \equiv Q(B|x, a)$ is a stochastic kernel on X given $\mathbb{K}$, where

$$\mathbb{K} = \{(x, a) \in X \times A : a \in A(x)\}.$$

- $c : \mathbb{K} \to \mathbb{R}$ is a measurable cost function.

Definition of the control model

- A control policy is a sequence $\pi = \{\pi_t\}_{t\in\mathbb{N}}$ of stochastic kernels π_t on A given H_t ($H_0 := X$ and $H_t := \mathbb{K} \times H_{t-1}$) satisfying $\pi_t(A(x_t)|h_t) = 1$ for all $h_t \in H_t$ and $t \in \mathbb{N}$, where $h_t := (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$.
  Let Π be the family of **randomized** history-dependent policies.

- Let $\mathbb{F}$ be the family of **deterministic stationary** policies, i.e., the class of $f : X \to A$ such that $f(x) \in A(x)$ for $x \in X$.

Optimality criteria

Given $\pi \in \Pi$ and an initial state $x \in X$, the total expected α-discounted cost $(0 < \alpha < 1)$ and the long-run average cost are

$$
\begin{aligned}
V_\alpha(x,\pi) &= E^{\pi,x}\Big[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\Big] \\
J(x,\pi) &= \limsup_{t\to\infty} E^{\pi,x}\Big[\frac{1}{t}\sum_{k=0}^{t-1} c(x_t,a_t)\Big].
\end{aligned}
$$

Definition of the control model

Optimality criteria

- The optimal discounted cost is

$$V_\alpha^*(x) = \inf_{\pi\in\Pi} V_\alpha(x,\pi).$$

- The optimal average cost is

$$J^*(x) = \inf_{\pi\in\Pi} J(x,\pi).$$

- A policy $\pi^* \in \Pi$ is average optimal if

$$J(x,\pi^*) = J^*(x) \quad \text{for all } x \in X.$$

Discretizing the state space

Main idea

- We suppose that there exists a probability measure μ on X and a nonnegative measurable function $q(\cdot|\cdot,\cdot)$ on $X \times \mathbb{K}$ such that

$$Q(B|x,a) = \int_B q(y|x,a)\mu(dy)$$

for all measurable $B \subseteq X$ and every $(x,a) \in \mathbb{K}$.

- On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we take a sample of n i.i.d. random observations $\{Y_k\}_{1 \leq k \leq n}$ with distribution μ and we consider the empirical probability measure

$$\mu_n(B) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{I}\{Y_k \in B\}.$$

Discretizing the state space

Main idea

- In the transition kernel, we replace μ with μ_n

$$Q(B|x,a) = \int_B q(y|x,a)\mu(dy) \rightsquigarrow \int_B q(y|x,a)\mu_n(dy)$$

- We have "discretized" the state space: from X to $\{Y_k\}_{1\leq k\leq n}$. Integration is discretized: from μ to μ_n.
- We must be able to compute the estimation error

$$\left| \int_X g(y)\mu(dy) - \int_X g(y)\mu_n(dy) \right|.$$

- We need a **convergence** $\mu_n \to \mu$ allowing to control such estimation errors for a **certain class** of functions g.

Convergence of probability measures on Polish spaces

Metrics

- 1-*Wasserstein metric.* For probability measures in $\mathcal{P}_1(X)$ with finite first moment: $\int_X \rho_X(x, x_0)\mu(dx) < \infty$:

$$W_1(\lambda, \mu) = \inf_{\{\nu:\nu_1=\lambda,\nu_2=\mu\}} \int_{X\times X} \rho_X(x_1, x_2)\nu(dx_1, dx_2).$$

- N.B.: The p-Wasserstein metric uses $\big(\rho_X(x_1, x_2)\big)^p$.
- The dual Kantorovich-Rubinstein characterization gives

$$W_1(\lambda, \mu) = \sup_{f\in\mathbb{L}_1(X)} \Big| \int f d\mu - \int f d\lambda \Big|$$

for all 1-Lipschitz continuous functions.

Convergence of empirical probability measures

Theorem (Boissard, 2011)

If $\mu \in \mathcal{P}_1(X)$ *satisfies the modified transport inequality:*

$$W_1(\mu,\lambda) \le C\Big(H(\lambda|\mu) + \sqrt{H(\lambda|\mu)}\Big)$$

for some $C > 0$ *and all* $\lambda \in \mathcal{P}_1(X)$ *then there exists* γ_0 *such that for all* $0 < \gamma \le \gamma_0$ *there exist* $C_1, C_2 > 0$ *with*

$$\mathbb{P}\{W_1(\mu_n,\mu) > \gamma\} \le C_1 \exp\{-C_2 n\} \quad \text{for all } n \ge 1.$$

Here, $H(\lambda|\mu)$ is the entropy $H(\lambda|\mu) = \int \log \frac{d\lambda}{d\mu} d\lambda$. A sufficient condition is the existence of $a > 0$ and $x_0 \in X$ such that

$$\int_X \exp\{a \cdot \rho_X(x,x_0)\}\mu(dx) < \infty.$$

Our setting

If f is L_f-Lipschitz-continuous

$$\left| \int f(y)\mu_n(dy) - \int f(y)\mu(dy) \right| \leq L_f W_1(\mu_n, \mu)$$

and the probability that

$$\left| \int f(y)\mu_n(dy) - \int f(y)\mu(dy) \right| > \gamma$$

goes to zero at an exponential rate. So, we will place ourselves in the "Lipschitz continuity" setting.

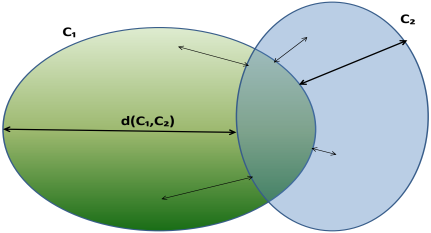
- The elements of the control model will be supposed to be Lipschitz-continuous.
- The action space will be approximated in a "Lipschitz-continuous" way.

Hypotheses

For each $x \in X$, the set $A(x)$ is compact, and $x \mapsto A(x)$ is Lipschitz continuous with respect to the Hausdorff metric, i.e.,

$$d_H(A(x), A(y)) \leq L\rho_X(x, y) \quad \text{for all } x, y \in X,$$

with $d_H(C_1, C_2) = \max\{\sup_{x_1 \in C_1} \rho_X(x_1, C_2), \sup_{x_2 \in C_2} \rho_X(x_2, C_1)\}$.



Hypotheses

There exists a Lipschitz-continuous function $w : X \to [1, \infty)$ such that for all $(x, a) \in \mathbb{K}$

- The cost function c is Lipschitz-continuous and

$$|c(x, a)| \leq \bar{c}w(x).$$

The density function $q(y|x, a)$ verifies

- $q(y|x, a) \leq \bar{q}w(x)$.
- It is Lipschitz-continuous in y (resp., (x, a)) uniformly in (x, a) (resp., y).
- $y \mapsto w(y)q(y|x, a)$ is $Lw(x)$-Lipschitz-continuous.

Hypotheses

- $Qw(x_0, a_0)$ is finite for some $(x_0, a_0) \in \mathbb{K}$ and there is some $0 < d < 1$ such that

$$\int_X w(y)|Q(dy|x,a) - Q(dy|x',a')| \leq 2d(w(x) + w(x')) \quad (1)$$

  for all (x, a) and (x', a') in $\mathbb{K}$.
- As a consequence of (1), there exists $b \geq 0$ such that

$$Qw(x,a) \leq dw(x) + b \quad \text{for all } (x,a) \in \mathbb{K}.$$

This is the usual "contracting" condition for average cost MDPs. We impose (1) because it implies a uniform geometric ergodicity condition under which we can use the vanishing discount approach to average optimality.

Dynamic programming equation

Notation

We say that $u : X \to \mathbb{R}$ is in $\mathbb{L}_w(X)$ if u is Lipschitz-continuous and there exists $M > 0$ with $|u(x)| \leq Mw(x)$ for all $x \in X$.

Theorem (Discounted cost)

Given a discount factor $0 < \alpha < 1$, the optimal discounted cost $V_\alpha^* \in \mathbb{L}_w(X)$ and it satisfies the α-DCOE

$$V_\alpha^*(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V_\alpha^*(y) Q(dy|x,a) \right\} \quad \text{for } x \in X.$$

$x \mapsto V_\alpha(x,\pi)$ *might not be continuous, but* $x \mapsto \inf_{\pi \in \Pi} V_\alpha(x,\pi)$ *is continuous!*

Dynamic programming equation

Theorem (Average cost)

- *There exist* $g \in \mathbb{R}$ *and* $h \in \mathbb{L}_w(X)$ *that are a solution to the ACOE*

$$g + h(x) = \min_{a \in A(x)} \left\{ c(x,a) + \int_X h(y) Q(dy|x,a) \right\} \quad \text{for } x \in X.$$

- *We have* $g = J^*(x) = \inf_{\pi \in \Pi} J(x,\pi)$ *for all* $x \in X$.
- *If* $f \in \mathbb{F}$ *attains the minimum in the ACOE, then it is average optimal.*

Sketch of the proof: Define $h_\alpha(x) = V_\alpha^*(x) - V_\alpha^*(x_0)$. Show that $\{h_\alpha\}$ is equicontinuous, and that its Lipschitz constant does not depend on α. Let $\alpha \to 1$.

Approximation of the control model

Approximation of the action space

For all $\mathfrak{d} > 0$ there exists a family $A_{\mathfrak{d}}(x)$, for $x \in X$, of subsets of A satisfying:

- $A_{\mathfrak{d}}(x)$ is a nonempty closed subset of $A(x)$, for $x \in X$.
- For every $x \in X$,

$$d_H(A(x), A_{\mathfrak{d}}(x)) \leq \mathfrak{d} w(x).$$

- The multifunction $x \mapsto A_{\mathfrak{d}}(x)$ is $L_{\mathfrak{d}}$-Lipschitz continuous with respect to the Hausdorff metric, with $\sup_{\mathfrak{d}>0} L_{\mathfrak{d}} < \infty$.

Approximation of the control model

Definition

Given $n \geq 1$ and $\mathfrak{d} > 0$, the control model $\mathcal{M}_{n,\mathfrak{d}}$ is defined by the elements

$$(X, A, \{A_{\mathfrak{d}}(x) : x \in X\}, Q_n, c),$$

Recall that $Q(B|x,a) = \int_B q(y|x,a)\mu(dy)$. Here,

$$Q_n(B|x,a) = \frac{\int_B q(y|x,a)\mu_n(dy)}{\int_X q(y|x,a)\mu_n(dy)} = \frac{\sum_{k:Y_k \in B} q(Y_k|x,a)}{\sum_{k=1}^n q(Y_k|x,a)}.$$

Note that $Q_n(\cdot|x,a)$ has finite support, and it assigns probability proportional to $q(Y_k|x,a)$ to Y_k.

Properties of $\mathcal{M}_{n,\mathfrak{d}}$

If $v \in \mathbb{L}_w(X)$ —w-bounded and Lipschitz-continuous— we can compare Qv and $Q_n v$:

$$|Qv(x,a) - Q_n v(x,a)| \leq \mathcal{C}_v w(x) W_1(\mu, \mu_n).$$

We will use the notation:

- $\mathbb{K}_{\mathfrak{d}} = \{(x,a) \in X \times A \ : \ a \in A_{\mathfrak{d}}(x)\}$.
- $\Pi_{\mathfrak{d}}$ and $\mathbb{F}_{\mathfrak{d}}$ are the families of all policies and deterministic stationary policies for the control model $\mathcal{M}_{n,\mathfrak{d}}$.
- The expectation operator is $E^{\pi,x}_{n,\mathfrak{d}}$.
- Let

$$J^*_{n,\mathfrak{d}}(x) = \inf_{\pi \in \Pi_{\mathfrak{d}}} \limsup_{t \to \infty} E^{\pi,x}_{n,\mathfrak{d}} \left[\frac{1}{t} \sum_{k=0}^{t-1} c(x_t, a_t) \right].$$

Properties of $\mathcal{M}_{n,\mathfrak{d}}$

Define

$$\mathfrak{c} = \frac{1-d}{4(L_{wq} + L_q(1+4(d+b)))}$$

and suppose that $\omega \in \Omega$ is such that $W_1(\mu, \mu_n(\omega)) \leq \mathfrak{c}$. Then we have:

- $Q_n(X|x,a) = 1$ for all $(x,a) \in \mathbb{K}_{\mathfrak{d}}$.
- For all $(x,a) \in \mathbb{K}_{\mathfrak{d}}$,

$$Q_n w(x,a) \leq \frac{1+d}{2} w(x) + 2b.$$

- For all (x,a) and (x',a') in $\mathbb{K}_{\mathfrak{d}}$

$$\int_X w(y)|Q_n(dy|x,a) - Q_n(dy|x',a')| \leq (1+d)\cdot(w(x)+w(x'))$$

Properties of $\mathcal{M}_{n,\mathfrak{d}}$

Theorem

If $\omega \in \Omega$ is such that $W_1(\mu, \mu_n(\omega)) \leq \mathfrak{c}$ then

- *The control model $\mathcal{M}_{n,\mathfrak{d}}$ is uniformly geometrically ergodic and it verifies the "same" properties as $\mathcal{M}$.*
- *The optimal average cost $J^*_{n,\mathfrak{d}}(x) \equiv g^*_{n,\mathfrak{d}}$ is constant and it satisfies the ACOE: for all $x \in X$*

$$g^*_{n,\mathfrak{d}} + h(x) = \min_{a \in A_{\mathfrak{d}}(x)} \left\{ c(x,a) + \int_X h(y) Q_n(dy|x,a) \right\}$$

  for some $h \in \mathbb{B}_w(X)$.
- *Besides, h is unique up to additive constants.*

Convergence of the optimal average cost

Theorem

There exists $\varepsilon_0 > 0$ *such that for any* $0 < \varepsilon \le \varepsilon_0$ *there exist* $\mathfrak{d} > 0$ *and constants* $\mathcal{S}, \mathcal{T} > 0$ *such that*

$$\mathbb{P}^*\{|g^*_{n,\mathfrak{d}} - g| > \varepsilon\} \le \mathcal{S} \exp\{-\mathcal{T} n\}.$$

for all $n \ge 1$.

Sketch of the proof

- From the ACOE for $\mathcal{M}$ we have

$$g + h(x) \leq c(x,a) + Qh(x,a).$$

- Replace Q with Q_n and obtain

$$g + h(x) \leq c(x,a) + Q_n h(x,a) + Cw(x)W_1(\mu,\mu_n).$$

- Iterate this inequality t times, divide by t, and take the limit as $t \to \infty$ to obtain $g \leq g^*_{n,\mathfrak{d}} + CW_1(\mu,\mu_n)$.
- For an $\mathcal{M}$-canonical policy $f \in \mathbb{F}$

$$g + h(x) = c(x,f) + Qh(x,f).$$

- Take the "projection" $\tilde{f}$ of f on $\mathbb{F}_{\mathfrak{d}}$ and obtain

$$g + h(x) \geq c(x,\tilde{f}) + Qh(x,\tilde{f}) - C\mathfrak{d}w(x).$$

- Replace Q with Q_n and proceed as before.

Approximation of an optimal policy

Main idea

- Starting from the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$

$$g^*_{n,\mathfrak{d}} + h(x) = \min_{a \in A_{\mathfrak{d}}(x)} \left\{ c(x,a) + \int_X h(y) Q_n(dy|x,a) \right\},$$

  let $\widetilde{f}_{n,\mathfrak{d}} \in \mathbb{F}_{\mathfrak{d}}$ be a canonical policy.
- Since $\widetilde{f}_{n,\mathfrak{d}} \in \mathbb{F}$, "use it" in the control model $\mathcal{M}$ to obtain the expected average cost $J(x, \widetilde{f}_{n,\mathfrak{d}})$
- Compare $J(x, \widetilde{f}_{n,\mathfrak{d}})$ and g.

Approximation of an optimal policy

Difficulties

- For a function v, we have that Qv is Lipschitz-continuous, but $Q_n v$ is locally Lipschitz-continuous.
- The function h in the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$ is locally Lipschitz-continuous.
- We cannot directly compare Qh with $Q_n h$.
- There exists a Lipschitz-continuous $\tilde{h}$ with

$$||h - \tilde{h}||_w \leq CW_1(\mu, \mu_n).$$

- Use this $\tilde{h}$ to compare $Q\tilde{h}$ and $Q_n\tilde{h}$.

Approximation of an optimal policy

Theorem

There exists $\varepsilon_0 > 0$ *such that for any* $0 < \varepsilon \leq \varepsilon_0$ *there exist* $\mathfrak{d} > 0$ *and constants* $\mathcal{S}, \mathcal{T} > 0$ *such that*

$$\mathbb{P}^*\{J(\widetilde{f}_{n,\mathfrak{d}}, x) - g > \varepsilon\} \leq \mathcal{S} \exp\{-\mathcal{T} n\}.$$

for all $n \geq 1$ *and* $x \in X$*.*

Finite state and action approximations

- For applications, suppose that the sets $A_{\mathfrak{d}}(x)$ are finite.
- Take a sample $\Gamma_n = \{Y_k(\omega)\}$ of the probability measure μ.
- The control model $\mathcal{M}_{n,\mathfrak{d}}$ has finite state and action spaces.
- We need to determine its optimal average cost $g^*_{n,\mathfrak{d}}$.
- We need to solve the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$ to find a canonical policy.

The linear programming approach

Primal linear programming problem P

$$\min \sum_{x\in\Gamma_n}\sum_{a\in A_{\mathfrak{d}}(x)} c(x,a)z(x,a) \quad \text{subject to}$$

$$\sum_{a\in A_{\mathfrak{d}}(x)} z(x,a) = \sum_{x'\in\Gamma_n}\sum_{a'\in A_{\mathfrak{d}}(x')} z(x',a')Q_n(\{x\}|x',a')$$

$$\sum_{x\in\Gamma_n}\sum_{a\in A_{\mathfrak{d}}(x)} z(x,a) = 1 \quad \text{and} \quad z(x,a)\geq 0$$

It is known that $\min P = g^*_{n,\mathfrak{d}}$, the optimal average cost of the control model $\mathcal{M}_{n,\mathfrak{d}}$.

The linear programming approach

Dual linear programming problem D

$$\max \quad g \quad \text{subject to}$$

$$g + h(x) \le c(x,a) + \sum_{y \in \Gamma_n} Q_n(\{y\}|x,a)h(y)$$

$$g \in \mathbb{R} \quad \text{and} \quad h(x) \in \mathbb{R}.$$

Its optimal value is $g^*_{n,\mathfrak{d}}$ and, at optimality, we obtain a solution of

$$g^*_{n,\mathfrak{d}} + h(x) \le \min_{a \in A_{\mathfrak{d}}(x)} \left\{ c(x,a) + \sum_{y \in \Gamma_n} Q_n(\{y\}|x,a)h(y) \right\} \qquad (2)$$

but not necessarily of the ACOE.

Solving the ACOE by linear programming

Our approach to approximate an optimal policy is based on a canonical policy for $\mathcal{M}_{n,\mathfrak{d}}$. We need to solve the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$.

Lemma (Maximal property)

Let $\{z^*(x,a)\}$ be an optimal solution of P, and fix arbitrary x^* with $\sum_{a\in A_{\mathfrak{d}}(x^*)} z^*(x^*,a) > 0$.

Let h^* be the unique solution of the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$ such that $h^*(x^*) = 0$, and let h, with $h(x^*) = 0$, verify the inequalities in (2).

Then we have $h \leq h^*$.

Solving the ACOE by linear programming

Modified dual linear programming problem D'

$$\max \quad \sum_{x\in\Gamma_n} h(x) \quad \text{subject to}$$

$$g^*_{n,\mathfrak{d}} + h(x) \leq c(x,a) + \sum_{y\in\Gamma_n} Q_n(\{y\}|x,a)h(y)$$

$$h(x^*) = 0 \quad \text{and} \quad h(x)\in\mathbb{R}.$$

Theorem

Solving P and then D' yields a solution of the ACOE for $\mathcal{M}_{n,\mathfrak{d}}$.

An inventory management system

Consider the dynamics

$$x_{t+1} = \max\{x_t + a_t - \xi_t, 0\} \quad \text{for } t \in \mathbb{N}$$

where

- x_t is the stock level at the beginning of period t;
- a_t is the amount ordered at the beginning of period t;
- ξ_t is the random demand at the end of period t.

The capacity of the warehouse is $M > 0$. Therefore,

$$X = A = [0, M] \quad \text{and} \quad A(x) = [0, M - x].$$

An inventory management system

The controller incurs:

- a buying cost of $b > 0$ for each unit;
- a holding cost $h > 0$ for each period and unit;
- and receives $p > 0$ for each unit that is sold.

The running cost function is

$$c(x, a) = ba + h(x + a) - pE[\min\{x + a, \xi\}].$$

Theorem

If the $\{\xi_t\}$ are i.i.d. with distribution function F, with $F(M) < 1$, and density function f, which is Lipschitz continuous on $[0, M]$ with $f(0) = 0$, then the inventory management system satisfies our assumptions.

An inventory management system

Fix $0 < \mathfrak{p} < 1$. The probability measure μ is

$$\mu\{0\} = \mathfrak{p} \quad \text{and} \quad \mu(B) = \frac{1-\mathfrak{p}}{M}\lambda(B) \quad \text{for measurable } B \subseteq (0, M],$$

The density function of the demand is

$$f(x) = \frac{1}{\lambda^2}\, x e^{-x/\lambda} \quad \text{for } x \geq 0.$$

The approximating action sets are

$$A_{\mathfrak{d}}(x) = \left\{ \frac{(M-x)j}{q_{\mathfrak{d}}-1} \;:\; j = 0, 1, \ldots, q_{\mathfrak{d}} - 1 \right\}.$$

An inventory management system

We take 500 samples of size n for the parameters

$$M = 10,\ b = 7,\ h = 3,\ p = 17,\ \mathfrak{p} = 1/10,\ \lambda = 5/2,\ q_{\mathfrak{d}} = 20.$$

	$n = 50$	$n = 150$	$n = 300$
Mean	-26.8755	-26.4380	-26.2817
Std. Dev.	2.2119	1.4578	1.0145
	$n = 500$	$n = 700$	$n = 1000$
Mean	-26.1717	-26.1553	-26.1659
Std. Dev.	0.8104	0.6662	0.5734

Table: Estimation of the optimal average cost g.

An inventory management system

We determine the canonical policy $\widetilde{f}_{n,\mathfrak{d}}$ for $\mathcal{M}_{n,\mathfrak{d}}$ and we evaluate it for $\mathcal{M}$.

	$n = 50$	$n = 150$	$n = 300$
Mean	-25.6312	-25.8387	-25.9724
Std. Dev.	0.7648	0.5394	0.3954
	$n = 500$	$n = 700$	$n = 1000$
Mean	-26.0406	-26.0497	-26.0833
Std. Dev.	0.3387	0.3276	0.3133

Table: Estimation of the average cost of the policy $\widetilde{f}_{n,\mathfrak{d}}$.

An inventory management system

We compute the relative error of $J(x, \widetilde{f}_{n,\mathfrak{d}})$ with respect to g.

$n = 50$	$n = 150$	$n = 300$	$n = 500$	$n = 700$	$n = 1000$
4.63%	2.27%	1.18%	0.50%	0.40%	0.32%

Table: Relative error.

An inventory management system

We display the approximation of an optimal policy for the control model $\mathcal{M}$.

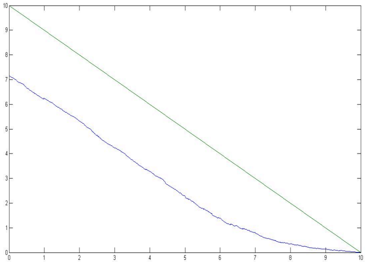


Figure: Estimation of an optimal policy

Conclusions

- We have proposed a general procedure to approximate a continuous state and action MDP.
- We can do this for a "Lipschitz-continuous" control model.
- We prove exponential rates of convergence (in probability).
- For applications, our method provides very good approximations.