Linear Programming Approximations of Constrained Markov Decision Processes [1]

Tomás Prieto-Rumeau

Statistics Department, UNED
Madrid, Spain

Coloquio de Sistemas Estocásticos
50 Aniversario del Departamento de Matemáticas del CINVESTAV-IPN
México DF, October 2011

[1] Joint work with François Dufour (INRIA, Bordeaux, France)

Introduction

- We are concerned with the numerical approximation of the solution of a constrained discrete-time discounted MDP.
- We are interested in obtaining explicit bounds for our approximation errors (and not just "convergence").

Introduction

- We are concerned with the numerical approximation of the solution of a constrained discrete-time discounted MDP.
- We are interested in obtaining explicit bounds for our approximation errors (and not just "convergence").
- We want to use discretization techniques suitable for the case of an MDP with noncompact state space.

Introduction

- We are concerned with the numerical approximation of the solution of a constrained discrete-time discounted MDP.
- We are interested in obtaining explicit bounds for our approximation errors (and not just "convergence").
- We want to use discretization techniques suitable for the case of an MDP with noncompact state space.
- We are going to approximate an infinite dimensional LP problem by a finite LP problem.

Constrained discrete-time MDPs

- Suppose that $\mathcal{M}$ is a (constrained) discrete-time MDP:

$$\mathcal{M} := \{X, A, (A(x), x \in X), P(dy|x,a), c(x,a), r(x,a)\}.$$

- The state space X is a locally compact Borel space (not necessarily compact).
- The action space A is a locally compact Borel space, and the action sets $A(x)$, for $x \in X$, are compact.
- The feasible state-actions set is $\mathbb{K} := \{(x,a) \in X \times A : a \in A(x)\}$.
- $P(B|x,a)$ is a stochastic kernel on X given $\mathbb{K}$.
- $c : \mathbb{K} \to \mathbb{R}$ and $r : \mathbb{K} \to \mathbb{R}^q$ are measurable cost-per-stage functions.

Constrained discrete-time MDPs

- The total expected discounted cost of a policy $\pi \in \Pi$ is

$$V(x, \pi, c) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right],$$

where $x \in X$ is the initial state, and $0 < \alpha < 1$ is a discount factor.

Constrained discrete-time MDPs

- The total expected discounted cost of a policy $\pi \in \Pi$ is

$$V(x,\pi,c) := E_x^\pi\Big[\sum_{t=0}^{\infty}\alpha^t c(x_t,a_t)\Big],$$

  where $x \in X$ is the initial state, and $0 < \alpha < 1$ is a discount factor.
- We want to approximate the solution of the constrained MDP

$$\text{minimize} \quad V(x_0,\pi,c) \quad \text{s.t.} \quad \pi \in \Pi \quad \text{and} \quad V(x_0,\pi,r) \le \theta_0,$$

  where $x_0 \in X$ is the initial state and $\theta_0 \in \mathbb{R}^q$ is the constraint constant.

Approximation of MDPs

- Consider a finite state and action discretization $\mathcal{M}_d$ of the control model $\mathcal{M}$, and use the optimal value of $\mathcal{M}_d$ as an approximation.
- If the state space X is compact, then we select a finite grid $x_k \in \mathbf{H}$ of states, with associated approximation error δ.
- Solve the MDP with state space $\mathbf{H}$ with an approximation error δ.

Approximation of MDPs

Main idea

- Here, we deal with a problem with noncompact state space X.
  1. Choose $\epsilon > 0$, and find a compact $K_\epsilon \subset X$ such that: "what happens outside K_ϵ has a weight less than ϵ".
  2. Discretize K_ϵ and obtain a δ-approximation of its optimal solution.
  3. Obtain a $(\delta + \epsilon)$-approximation.

Approximation of MDPs

Main idea

- Here, we deal with a problem with noncompact state space X.
  1. Choose $\epsilon > 0$, and find a compact $K_\epsilon \subset X$ such that: "what happens outside K_ϵ has a weight less than ϵ".
  2. Discretize K_ϵ and obtain a δ-approximation of its optimal solution.
  3. Obtain a $(\delta + \epsilon)$-approximation.
- Our approach: Use a discretization technique that proceeds in a single step (and not in two steps, as above).

Approximation of MDPs

Main idea

- Suppose that the stochastic kernel has a density with respect to some probability measure μ on X.
- There exists a function $p(y|x,a)$ on $X \times \mathbb{K}$ such that

$$P(B|x,a) = \int_B p(y|x,a)\mu(dy) \quad \text{for } B \subseteq X.$$

Approximation of MDPs

Main idea

- Suppose that the stochastic kernel has a density with respect to some probability measure μ on X.
- There exists a function $p(y|x,a)$ on $X \times \mathbb{K}$ such that

$$P(B|x,a) = \int_B p(y|x,a)\mu(dy) \quad \text{for } B \subseteq X.$$

- Obtain a discretization μ_N on a finite set $\mathbf{H}$ of the distribution μ, and consider the discretized kernels

$$P_N(B|x,a) = \int_B p(y|x,a)\mu_N(dy)$$

supported on $\mathbf{H}$.

Approximation of MDPs

Quantization

- Suppose that the state space X is a subset of $\mathbb{R}^d$.
- If Y is a random variable on $\mathbb{R}^d$ with distribution μ, let Y_N be the projection of Y (in the $L_2(\mathbb{R}^d)$ norm) in the space of random variables supported on N points in $\mathbb{R}^d$.
- We call Y_N the quantization of Y. We have explicit convergence rates:

$$||Y - Y_N||_2 = \mathrm{O}(N^{-1/d}).$$

- There are "toolboxes" that can find explicitly the random variable Y_N for a given distribution μ.

Approximation of MDPs

Plan of work

- Approximate the solution of the constrained MDP with transition kernel $P(B|x,a)$ by means of a constrained MDP with the quantized transition kernel $P_N(B|x,a)$.
- Obtain explicit bounds on the approximation error: given a precision $\varepsilon > 0$, determine *a priori* the number of points N needed in the quantization grid.
- We use a mixture of dynamic programming and linear programming.

Dynamic programming vs. linear programming

The DP approach

- In an unconstrained problem the optimal discounted cost is the solution of the discounted cost optimality equation (DCOE)

$$V^*(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V^*(y) P(dy|x,a) \right\}, \text{ for } x \in X.$$

In this case, we could study the DCOE for the quantized kernels P_N.

Dynamic programming vs. linear programming

The DP approach

- In an unconstrained problem the optimal discounted cost is the solution of the discounted cost optimality equation (DCOE)

$$V^*(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X V^*(y) P(dy|x,a) \right\}, \text{ for } x \in X.$$

  In this case, we could study the DCOE for the quantized kernels P_N.
- For a constrained problem, there exists $\lambda^* \in \mathbb{R}_+^q$ such that

$$\begin{aligned} V^*(x) &= \inf_{a \in A(x)} \Big\{ c(x,a) + \langle \lambda^*, r(x,a) - (1-\alpha)\theta_0 \rangle \\ &\qquad + \alpha \int_X V^*(y) P(dy|x,a) \Big\}. \end{aligned}$$

  This optimality equation is somehow useless because λ^* is unknown and, besides, a minimizing policy might not be constrained optimal.

Dynamic programming vs. linear programming

The LP approach

- Given a policy $\pi \in \Pi$, define the expected discounted state-action occupation measure for measurable $\Gamma \subseteq \mathbb{K} \subseteq X \times A$:

$$\nu_\pi(\Gamma) := \sum_{t=0}^{\infty} \alpha^t P_{x_0}^{\pi}\{(x_t, a_t) \in \Gamma\}$$

- The space of "feasible measures" $\{\nu_\pi\}_{\pi \in \Pi} = \mathcal{P}$ is characterized by means of linear constraints.
- The unconstrained and constrained control problems are respectively equivalent to the infinite dimensional LP problems

$$\begin{aligned} &\text{minimize} \quad \nu(c) \quad \text{s.t.} \quad \nu \in \mathcal{P} \\ &\text{minimize} \quad \nu(c) \quad \text{s.t.} \quad \nu \in \mathcal{P} \quad \text{and} \quad \nu(r) \leq \theta_0. \end{aligned}$$

- Both problems are of the "same nature".

Statement of the problem

Lipschitz continuity framework

- Given a function $v : X \to \mathbb{R}$ we want to compare

$$Pv(x,a) = \int_X v(y)p(y|x,a)\mu(dy) = E[v(Y)p(Y|x,a)]$$

and

$$P_N v(x,a) = \int_X v(y)p(y|x,a)\mu_N(dy) = E[v(Y_N)p(Y_N|x,a)].$$

- We know that Y_N is close to Y in the $L_2(\mathbb{R}^d)$ norm.
- Under adequate Lipschitz continuity conditions (in particular, v must be Lipschitz continuous), we can show that

$$P_N v(x,a) \quad \text{is close to} \quad Pv(x,a).$$

Statement of the problem

Lipschitz continuity framework

- Given a function $u : \mathbb{K} \to \mathbb{R}$ (interpreted as a cost function), define the dynamic programming operators:

$$(T^u v)(x) := \inf_{a \in A(x)} \left\{ u(x,a) + \alpha \int_X v(y) p(y|x,a) \mu(dy) \right\}$$

  and $T^u_N v$, with μ replaced with μ_N.
- We have that $T^u v$ and $T^u_N v$ are close provided that v is Lipschitz continuous.
- Hence, we place ourselves in the context of a Lipschitz continuous MDP.

Statement of the problem

Lipschitz continuity framework

- The elements $x \mapsto A(x)$, P, and u (the cost function) of the control model $\mathcal{M}$ are Lipschitz continuous.
- Then the optimal discounted cost V^*, i.e., the solution of the DCOE

$$V^*(x) = \inf_{a \in A(x)} \left\{ u(x,a) + \alpha \int_X V^*(y) P(dy|x,a) \right\}$$

is Lipschitz continuous.

Statement of the problem

Lipschitz continuity framework

- The elements $x \mapsto A(x)$, P, and u (the cost function) of the control model $\mathcal{M}$ are Lipschitz continuous.
- Then the optimal discounted cost V^*, i.e., the solution of the DCOE

$$V^*(x) = \inf_{a \in A(x)} \left\{ u(x,a) + \alpha \int_X V^*(y) P(dy|x,a) \right\}$$

  is Lipschitz continuous.
- Note that $x \mapsto V(x, \pi, u)$ is not, in general, continuous; but $x \mapsto \inf_{\pi \in \Pi} V(x, \pi, u)$ is continuous.

The linear programming approach

Main idea

- The LP that finds an optimal policy for the constrained MDP is $\mathbb{LP}$:

$$J^* = \min \ \nu(c) \quad \text{s.t.} \quad \nu(r-(1-\alpha)\theta_0) \leq \mathbf{0} \quad \text{and}$$

$$\nu(B\times A) = \delta_{x_0}(B) + \alpha \int_{\mathbb{K}} P(B|x,a)\nu(dx,da) \quad \text{for } B\subseteq X.$$

The linear programming approach

Main idea

- The LP that finds an optimal policy for the constrained MDP is $\mathbb{LP}$:

$$J^* = \min \ \nu(c) \quad \text{s.t.} \quad \nu(r-(1-\alpha)\theta_0) \leq \mathbf{0} \quad \text{and}$$

$$\nu(B\times A) = \delta_{x_0}(B) + \alpha\int_{\mathbb{K}} P(B|x,a)\nu(dx,da) \quad \text{for } B\subseteq X.$$

- We solve the finite state LP problem $\mathbb{LP}_N$

$$J^*_N := \min \ \nu(c) \quad \text{s.t.} \quad \nu(r-(1-\alpha)\theta_0) \leq \mathbf{0} \quad \text{and}$$

$$\nu(B\times A) = \delta_{x_0}(B) + \alpha\int_{\mathbb{K}} P_N(B|x,a)\nu(dx,da) \quad \text{for } B\subseteq X.$$

Steps of the proof

- The kernel P_N is not stochastic, and so there is no underlying Markov decision process.
- If $\mathbb{LP}$ verifies the Slater condition

$$\nu(r-(1-\alpha)\theta_0)<\mathbf{0} \quad \text{for some } \nu,$$

  then show that for large N the problem $\mathbb{LP}_N$ also satisfies the Slater condition.
- Consequently, both optima are the fixed points of the operators T^u and $T_N^{u_N}$ for

$$\begin{aligned} u(x,a) &= c(x,a)-\langle\lambda^*, r(x,a)-(1-\alpha)\theta_0\rangle \\ u_N(x,a) &= c(x,a)-\langle\lambda_N^*, r(x,a)-(1-\alpha)\theta_0\rangle. \end{aligned}$$

- Both cost functions being Lipschitz continuous, the corresponding fixed points are "close".

Main result

Theorem

Consider the Lipschitz continuous constrained MDP. Given an initial state $x_0 \in X$ and an arbitrary $\epsilon > 0$, there exists N such that

$$|J^* - J_N^*| < \epsilon.$$

Moreover, N depends on explicitly known data (the Lipschitz constants of the MDP, the norm of the cost functions, etc.).

Conclusions

- We have introduced a technique which allows to approximate explicitly the solution of a constrained MDP.
- We base our approach on the quantization of an underlying probability distribution.
- Our proofs are mainly based on finite state approximations of linear problems, with a digression to dynamic programming techniques.
- Numerical experimentation of this approach is in progress.

Thank you for your attention.