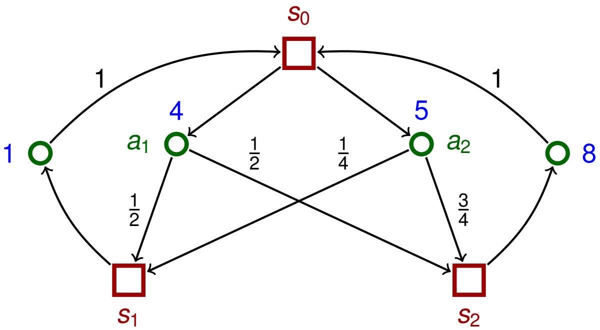
Trading Performance for Stability in Markov Decision Processes

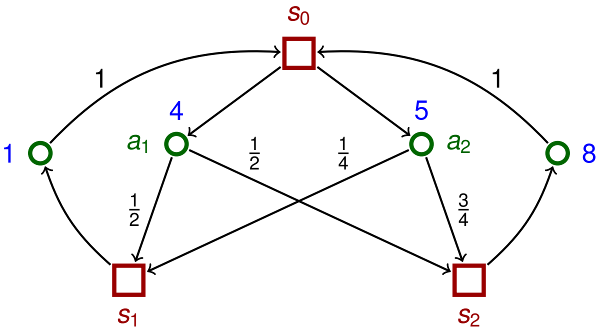
Tomáš Brázdil, K. Chatterjee, V. Forejt, A. Kučera

HIGHLIGHTS 2013

Markov Decision Processes

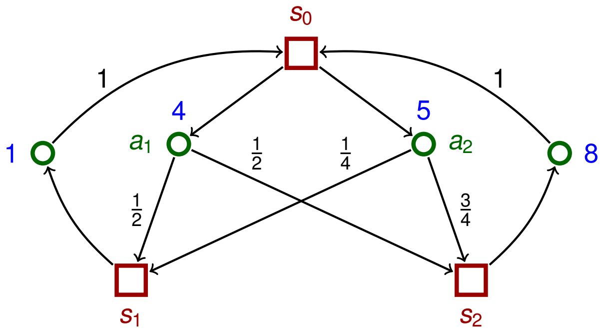


Markov Decision Processes



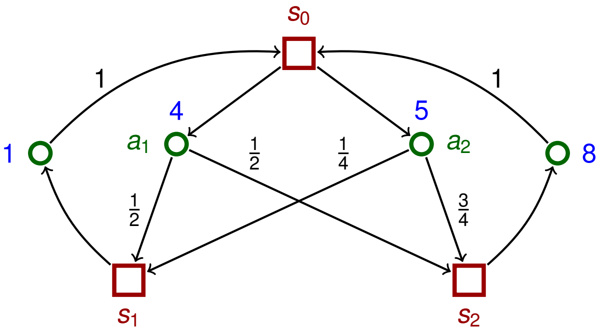
- A finite set of states together with one initial state
  - Ex: states s_0, s_1, s_2, and s_0 is initial

Markov Decision Processes



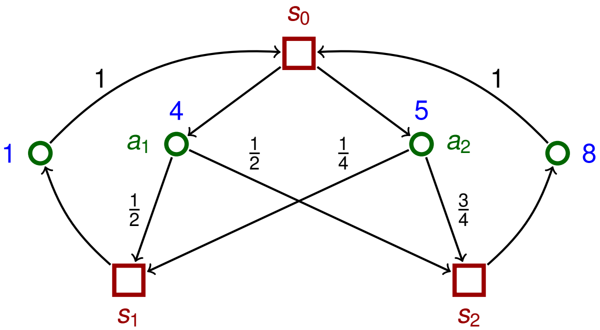
- A finite set of states together with one initial state
  - Ex: states s_0, s_1, s_2, and s_0 is initial
- A finite set of actions enabled in every state
  - each action is assigned a probability distribution on states
  - Ex: a_1, a_2 are enabled in s_0

Markov Decision Processes



- A finite set of states together with one initial state
  - Ex: states s_0, s_1, s_2, and s_0 is initial
- A finite set of actions enabled in every state
  - each action is assigned a probability distribution on states
  - Ex: a_1, a_2 are enabled in s_0
- Each action is assigned a (real) value v, e.g. $v(a_1) = 4$

Markov Decision Processes



- A finite set of states together with one initial state
  - Ex: states s_0, s_1, s_2, and s_0 is initial
- A finite set of actions enabled in every state
  - each action is assigned a probability distribution on states
  - Ex: a_1, a_2 are enabled in s_0
- Each action is assigned a (real) value v, e.g. $v(a_1) = 4$

Scheduler chooses actions, possibly in random, based on the history.

Mean payoff

Define a random variable **M** on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{M}(\omega) = \limsup_{T \to \infty} \frac{\sum_{n=0}^{T} v(a_n)}{T+1}$$

Mean payoff

Define a random variable **M** on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{M}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} v(a_n)}{T+1}$$

Minimize the expected value $\mathbb{E}_\sigma \mathbf{M}$, i.e.

a scheduler τ is optimal if $\mathbb{E}_\tau \mathbf{M} = \inf_\sigma \ \mathbb{E}_\sigma \mathbf{M}$.

Mean payoff

Define a random variable **M** on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{M}(\omega) \quad = \quad \limsup_{T \to \infty} \frac{\sum_{n=0}^{T} v(a_n)}{T+1}$$

Minimize the expected value $\mathbb{E}_\sigma \mathbf{M}$, i.e.

a scheduler τ is optimal if $\mathbb{E}_\tau \mathbf{M} = \inf_\sigma \ \mathbb{E}_\sigma \mathbf{M}$.

Fact

There exists a deterministic memory-less *optimal scheduler computable in polynomial time.*

Mean payoff

Define a random variable **M** on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{M}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} v(a_n)}{T+1}$$

Minimize the expected value $\mathbb{E}_\sigma \mathbf{M}$, i.e.

a scheduler τ is optimal if $\mathbb{E}_\tau \mathbf{M} = \inf_\sigma \ \mathbb{E}_\sigma \mathbf{M}$.

Fact
There exists a deterministic memory-less *optimal scheduler computable in polynomial time.*

Does $\mathbb{E}_\tau \mathbf{M}$ provide sufficient information about τ ?

Mean payoff

Define a random variable **M** on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{M}(\omega) \quad = \quad \limsup_{T \to \infty} \frac{\sum_{n=0}^{T} v(a_n)}{T+1}$$

Minimize the expected value $\mathbb{E}_\sigma \mathbf{M}$, i.e.

a scheduler τ is optimal if $\mathbb{E}_\tau \mathbf{M} = \inf_\sigma \ \mathbb{E}_\sigma \mathbf{M}$.

Fact

There exists a deterministic memory-less *optimal scheduler computable in polynomial time.*

Does $\mathbb{E}_\tau \mathbf{M}$ provide sufficient information about τ ? (No!)

Motivation

There are two weak points:

1. the *expected value* of **M**
2. **M** is defined by taking *averages*

Motivation

There are two weak points:

1. the *expected value* of **M**
2. **M** is defined by taking *averages*

Example:

Consider a streaming video.

A user of streaming video is promised bandwidth of 2 Mbit/sec, i.e. wants

1. 2 Mbit/sec whenever connected,
2. 2 Mbit/sec for the whole time.

Motivation

There are two weak points:

1. the *expected value* of **M**
2. **M** is defined by taking *averages*

Example:

Consider a streaming video.

A user of streaming video is promised bandwidth of 2 Mbit/sec, i.e. wants

1. 2 Mbit/sec whenever connected,
2. 2 Mbit/sec for the whole time.

$\mathbb{E}_\sigma \mathbf{M} \leq$ 2 Mbit/sec   does guarantee neither!

Motivation

There are two weak points:

1. the *expected value* of **M**
2. **M** is defined by taking *averages*

Example:
Consider a streaming video.

A user of streaming video is promised bandwidth of 2 Mbit/sec, i.e. wants

1. 2 Mbit/sec whenever connected,
2. 2 Mbit/sec for the whole time.

$\mathbb{E}_\sigma \mathbf{M} \leq 2$ Mbit/sec   does guarantee neither!

The user needs *stable* transmission!

Variance – The Main Definition

Stability can be expressed using variance of

Variance – The Main Definition

Stability can be expressed using variance of

1. the variable **M** (*global variance*)

$$\boxed{\mathbb{V}_\sigma \mathbf{M} \quad := \quad \mathbb{E}_\sigma \left(\mathbf{M} - \mathbb{E}_\sigma \mathbf{M}\right)^2 \quad = \quad \mathbb{E}_\sigma \mathbf{M}^2 - \left(\mathbb{E}_\sigma \mathbf{M}\right)^2}$$

Variance – The Main Definition

Stability can be expressed using variance of

1. the variable **M** (*global variance*)

$$\mathbb{V}_\sigma \mathbf{M} \quad := \quad \mathbb{E}_\sigma \left(\mathbf{M} - \mathbb{E}_\sigma \mathbf{M}\right)^2 \quad = \quad \mathbb{E}_\sigma \mathbf{M}^2 - \left(\mathbb{E}_\sigma \mathbf{M}\right)^2$$

2. individual values on a run (*local variance*) $\mathbb{E}_\sigma \mathbf{V}$ where **V** is a random variable on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{V}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} \left(v(a_n) - \mathbf{M}(\omega)\right)^2}{T+1}$$

Variance – The Main Definition

Stability can be expressed using variance of

1. the variable **M** (*global variance*)

$$\boxed{\mathbb{V}_\sigma \mathbf{M} \quad := \quad \mathbb{E}_\sigma \left(\mathbf{M} - \mathbb{E}_\sigma \mathbf{M}\right)^2 \quad = \quad \mathbb{E}_\sigma \mathbf{M}^2 - \left(\mathbb{E}_\sigma \mathbf{M}\right)^2}$$

2. individual values on a run (*local variance*) $\mathbb{E}_\sigma \mathbf{V}$ where **V** is a random variable on runs $\omega = s_0 a_0 s_1 a_1 \dots$

$$\boxed{\mathbf{V}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} \left(v(a_n) - \mathbf{M}(\omega)\right)^2}{T+1}}$$

Consider Pareto optimality w.r.t. both $\mathbb{E}_\sigma \mathbf{M}$ and a variance (either $\mathbb{V}_\sigma \mathbf{M}$, or $\mathbb{E}_\sigma \mathbf{V}$):

Variance – The Main Definition

Stability can be expressed using variance of

1. the variable **M** (*global variance*)

$$\boxed{\mathbb{V}_\sigma\mathbf{M} \quad := \quad \mathbb{E}_\sigma\left(\mathbf{M} - \mathbb{E}_\sigma\mathbf{M}\right)^2 \quad = \quad \mathbb{E}_\sigma\mathbf{M}^2 - \left(\mathbb{E}_\sigma\mathbf{M}\right)^2}$$

2. individual values on a run (*local variance*) $\mathbb{E}_\sigma\mathbf{V}$ where **V** is a random variable on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\boxed{\mathbf{V}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T}\left(v(a_n) - \mathbf{M}(\omega)\right)^2}{T+1}}$$

Consider Pareto optimality w.r.t. both $\mathbb{E}_\sigma\mathbf{M}$ and a variance (either $\mathbb{V}_\sigma\mathbf{M}$, or $\mathbb{E}_\sigma\mathbf{V}$):

- Small memory (2-memory for global, 3-memory for local) is sufficient for Pareto optimal schedulers

Variance – The Main Definition

Stability can be expressed using variance of

1. the variable **M** (*global variance*)

$$\mathbb{V}_\sigma \mathbf{M} \quad := \quad \mathbb{E}_\sigma \left(\mathbf{M} - \mathbb{E}_\sigma \mathbf{M}\right)^2 \quad = \quad \mathbb{E}_\sigma \mathbf{M}^2 - \left(\mathbb{E}_\sigma \mathbf{M}\right)^2$$

2. individual values on a run (*local variance*) $\mathbb{E}_\sigma \mathbf{V}$ where **V** is a random variable on runs $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\mathbf{V}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} \left(v(a_n) - \mathbf{M}(\omega)\right)^2}{T+1}$$

Consider Pareto optimality w.r.t. both $\mathbb{E}_\sigma \mathbf{M}$ and a variance (either $\mathbb{V}_\sigma \mathbf{M}$, or $\mathbb{E}_\sigma \mathbf{V}$):

- Small memory (2-memory for global, 3-memory for local) is sufficient for Pareto optimal schedulers
- Pareto optimal schedulers can be effectively approximated
  (pseudo-polynomial time in global case, exponential time in local case)

Hybrid Variance

What if we want to minimize both global and local variances?

Hybrid Variance

What if we want to minimize both global and local variances?

Proposal: Minimize the sum $\mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

Hybrid Variance

What if we want to minimize both global and local variances?

Proposal: Minimize the sum $\mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

But how to combine the two (very different) solutions?

Hybrid Variance

What if we want to minimize both global and local variances?

Proposal: Minimize the sum $\mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

But how to combine the two (very different) solutions?

Consider *hybrid variance* $\mathbb{E}_\sigma \mathbf{H}$ where for $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\boxed{\mathbf{H}(\omega) \quad = \quad \limsup_{T \to \infty} \frac{\sum_{n=0}^{T} (v(a_n) - \mathbb{E}_\sigma \mathbf{M})^2}{T+1}}$$

Hybrid Variance

What if we want to minimize both global and local variances?

Proposal: Minimize the sum $\mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

But how to combine the two (very different) solutions?

Consider *hybrid variance* $\mathbb{E}_\sigma \mathbf{H}$ where for $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\boxed{\mathbf{H}(\omega) \quad = \quad \limsup_{T\to\infty} \frac{\sum_{n=0}^{T} (v(a_n) - \mathbb{E}_\sigma \mathbf{M})^2}{T+1}}$$

Proposition

For all (reasonable) schedulers σ we have $\mathbb{E}_\sigma \mathbf{H} = \mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

Hybrid Variance

What if we want to minimize both global and local variances?

Proposal: Minimize the sum $\mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

But how to combine the two (very different) solutions?

Consider *hybrid variance* $\mathbb{E}_\sigma \mathbf{H}$ where for $\omega = s_0 a_0 s_1 a_1 \ldots$

$$\boxed{\mathbf{H}(\omega) \quad = \quad \limsup_{T \to \infty} \frac{\sum_{n=0}^{T} (v(a_n) - \mathbb{E}_\sigma \mathbf{M})^2}{T+1}}$$

Proposition

For all (reasonable) schedulers σ we have $\quad \mathbb{E}_\sigma \mathbf{H} = \mathbb{V}_\sigma \mathbf{M} + \mathbb{E}_\sigma \mathbf{V}$.

Consider Pareto optimality w.r.t. both $\mathbb{E}_\sigma \mathbf{M}$ and $\mathbb{E}_\sigma \mathbf{H}$:

- 2-memory is sufficient for Pareto optimal schedulers
- Pareto optimal schedulers can be approximated in pseudo-polynomial time

Conclusions & Future Work

- We have considered the problem of computing “stable” mean payoff in MDPs using
  - (global) variance
  - local variance
  - hybrid variance
- We have developed algorithms for Pareto optimization of expected mean payoff vs variance

Conclusions & Future Work

- We have considered the problem of computing “stable” mean payoff in MDPs using
  - (global) variance
  - local variance
  - hybrid variance
- We have developed algorithms for Pareto optimization of expected mean payoff vs variance

Future work

- Extension to vector mean payoffs (multiobjective)
- Other approaches to stabilization(?)
- Improve complexity bounds, matching lower bounds