Deterministic Policies for Constrained Reinforcement Learning in Polynomial Time

Jeremy McMahan

Why Deterministic Policies?

Why Deterministic Policies?

- Cheap [1]

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]
- Trust-worthy [3]

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]
- Trust-worthy [3]

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]
- Trust-worthy [3]
  - Predictable

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]
- Trust-worthy [3]
  - Predictable

Why Deterministic Policies?

- Cheap [1]
- Multi-agent coordination [2]
- Trust-worthy [3]
  - Predictable
- Optimal for modern constraints [4]

Modern Constraints

Modern Constraints

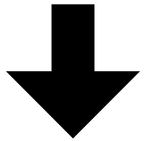
Expectation

Modern Constraints

Expectation — Unlimited Resource

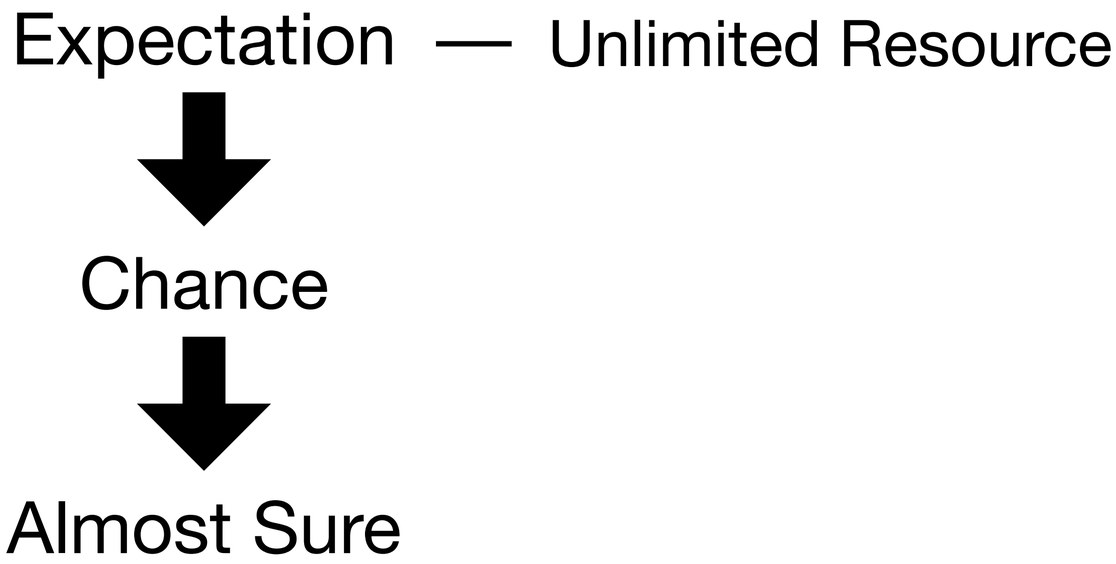
Modern Constraints

Expectation — Unlimited Resource

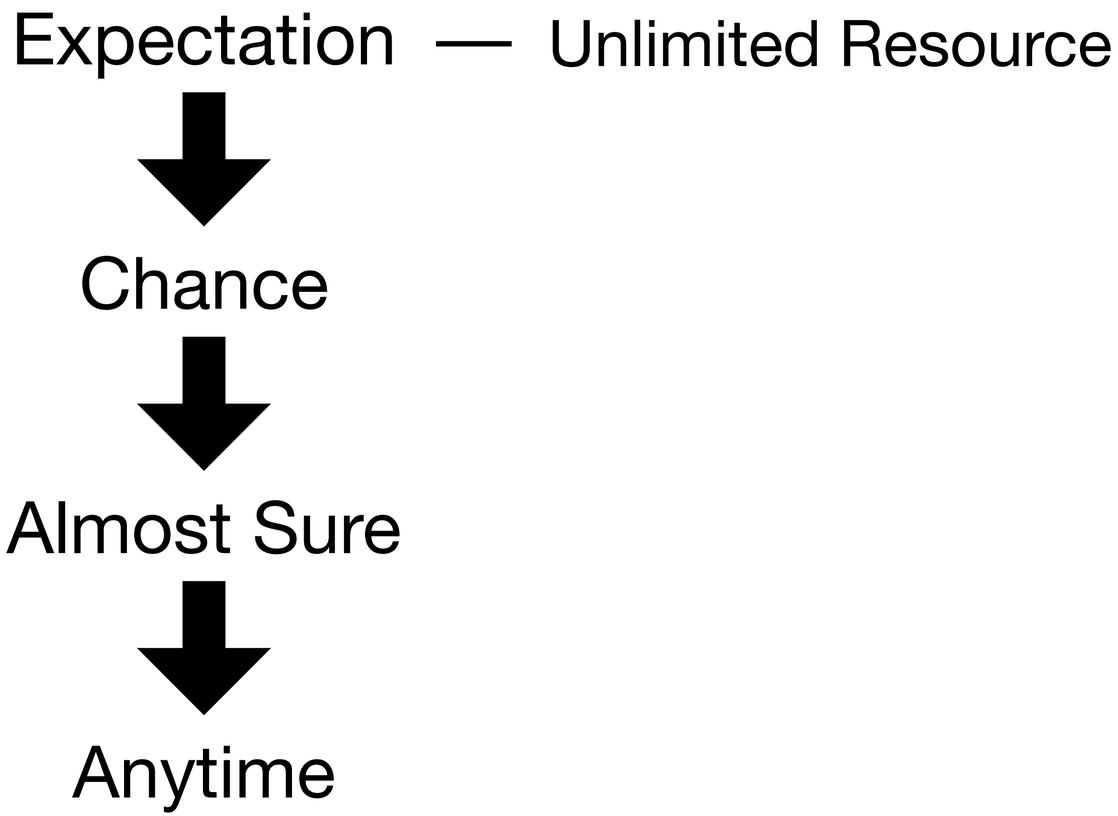


Chance

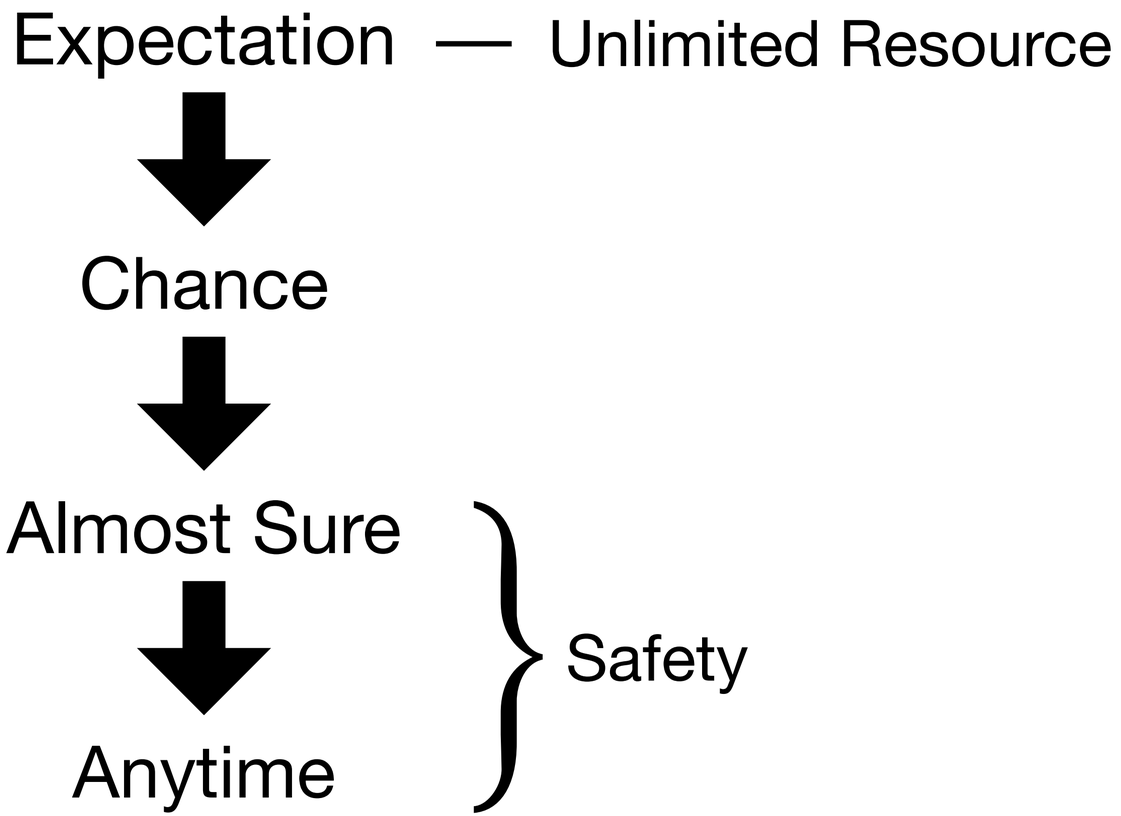
Modern Constraints



Modern Constraints



Modern Constraints



Problem

Problem

$$\max_{\pi \in \Pi} \mathbb{E}_M^{\pi} \left[\sum_{h=1}^{H} r_h(s_h, a_h) \right] \quad \text{s.t.} \quad \begin{cases} C_M^{\pi} \leq B \\ \pi \text{ deterministic} \end{cases}$$

Problem

$$\max_{\pi \in \Pi} \mathbb{E}_M^{\pi} \left[\sum_{h=1}^{H} r_h(s_h, a_h) \right] \quad \text{s.t.} \quad \begin{cases} C_M^{\pi} \leq B \\ \pi \text{ deterministic} \end{cases}$$

C belongs to a family of criteria including
expected, ***almost-sure***, and ***anytime*** criteria

Challenges

Challenges

- Problem is NP-hard

Challenges

- Problem is NP-hard
- Feasibility is NP-hard for > 1 constraint

Challenges

- Problem is NP-hard
- Feasibility is NP-hard for > 1 constraint
- Problem is not continuous

Challenges

- Problem is NP-hard
- Feasibility is NP-hard for > 1 constraint
- Problem is not continuous
- Dynamic programming fails

Can useful constraints be approximated efficiently?

Result

Result

Yes!

Result

Yes!

We design an additive and relative **FPTAS** for general cost criteria, including **expectation**, **almost-sure**, and **anytime**.

Result

Yes!

We design an additive and relative **FPTAS** for general cost criteria, including **expectation**, **almost-sure**, and **anytime**.

**Only chance constraints are left out, which are provably inapproximable*

Impact

Impact

Answers **three** long-standing open questions.

Impact

Answers **three** long-standing open questions.

Polynomial-time approximability is possible for:

Impact

Answers **three** long-standing open questions.

Polynomial-time approximability is possible for:

- *Almost-sure-constrained policies*

Impact

Answers **three** long-standing open questions.

Polynomial-time approximability is possible for:

- *Almost-sure-constrained policies*
- *Anytime-constrained policies*

Impact

Answers **three** long-standing open questions.

Polynomial-time approximability is possible for:

- *Almost-sure-constrained policies*
- *Anytime-constrained policies*
- *Deterministic, expectation-constrained policies*

Impact

Answers **three** long-standing open questions.

Polynomial-time approximability is possible for:

- *Almost-sure-constrained policies*
- *Anytime-constrained policies*
- *Deterministic, expectation-constrained policies*

Open for nearly 25 years!

Thank you!

References

1

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 25, No. 1, February 2000
Printed in U.S.A.

CONSTRAINED DISCOUNTED MARKOV DECISION PROCESSES AND HAMILTONIAN CYCLES

EUGENE A. FEINBERG

2

Towards a formalization of teamwork with resource constraints

Praveen Paruchuri, Milind Tambe, Fernando Ordonez
University of Southern California
Los Angeles, CA 90089
{paruchur,tambe,fordon}@usc.edu

Sarit Kraus
Bar-Ilan University
Ramat-Gan 52900, Israel
sarit@macs.biu.ac.il

3

Stationary Deterministic Policies for Constrained MDPs with Multiple Rewards, Costs, and Discount Factors

Dmitri Dolgov and Edmund Durfee
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109
{ddolgov, durfee}@umich.edu

4

Anytime-Constrained Reinforcement Learning

Jeremy McMahan **Xiaojin Zhu**
University of Wisconsin-Madison