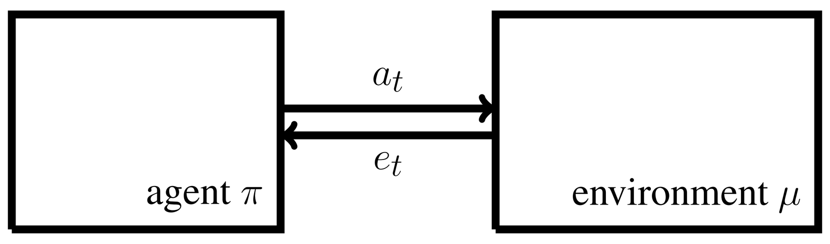
On the Computability of AIXI

Jan Leike and Marcus Hutter

General Reinforcement Learning



At every time step t the agent

- outputs action $a_t \in \mathcal{A}$
- receives percept $e_t \in \mathcal{E}$ including reward $r(e_t) \in \mathbb{R}$

Policy $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$
Environment $\mu : \mathcal{A}^* \to \Delta(\mathcal{E}^*)$
Discount function $\gamma : \mathbb{N} \to \mathbb{R}$ with $\gamma_t := \gamma(t) \geq 0$ and $\Gamma_t := \sum_{i=t}^{\infty} \gamma_i < \infty$; finite lifetime: $\Gamma_m = 0$
Assumptions:

- rewards are bounded between 0 and 1
- $\mathcal{A}$ and $\mathcal{E}$ are finite
- γ is lower semicomputable

Goal: maximize discounted rewards

How to Solve General Reinforcement Learning with Infinite Computation?

Answer: AIXI (Hutter, 2005)

Value Functions

Iterative value of policy π in environment ν: Don't use this

$$V_\nu^\pi(æ_{<t}) := \frac{1}{\Gamma_t} \lim_{m \to \infty} \sum_{e_{t:m}} \nu(e_{1:m} \mid e_{<t} \parallel a_{1:m}) \sum_{k=t}^{m} \gamma_k r_k$$

Recursive value of policy π in environment ν: Use this

$$W_\nu^\pi(æ_{<t}) := W_\nu^\pi(æ_{<t}\pi(æ_{<t}))$$
$$W_\nu^\pi(æ_{<t}a_t) := \frac{1}{\Gamma_t} \sum_{e_t} \big(\gamma_t r(e_t) + \Gamma_{t+1} W_\nu^\pi(æ_{1:t})\big) \nu(e_{1:t} \mid e_{<t} \parallel a_{1:t})$$

Optimal policy: $\arg\max_\pi V_\nu^\pi$ and $\arg\max_\pi W_\nu^\pi$

AIXI's Universal Prior

Universal prior akin to Solomonoff (1964, 1978), but for reactive environments:

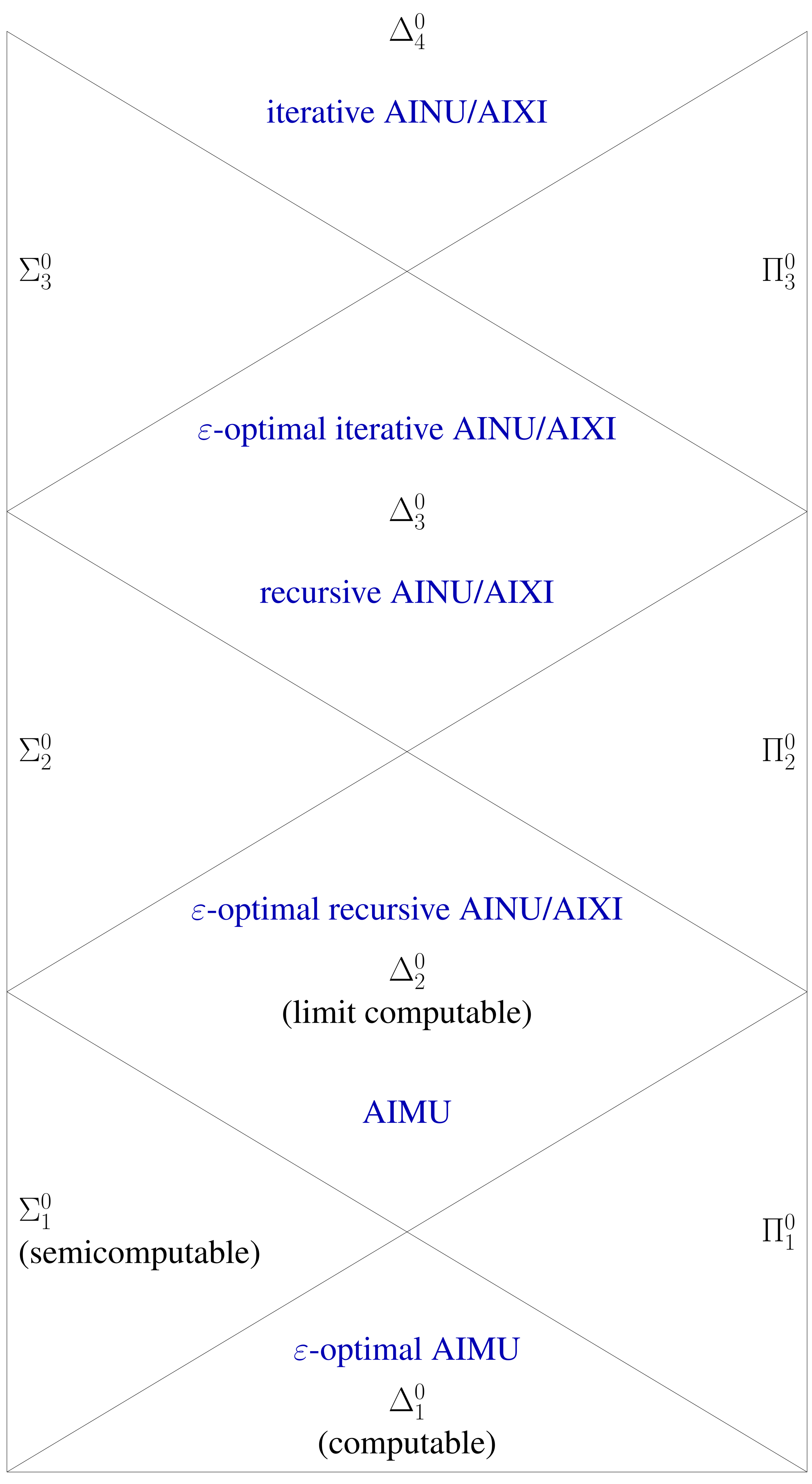
$$\xi(e_{<t} \parallel a_{<t}) := \sum_\nu w_\nu \nu(e_{<t} \parallel a_{<t}).$$

with $w_\nu > 0$ lower semicomputable and $\sum_\nu w_\nu \leq 1$
ξ returns the probability that the universal Turing machine U generates $e_{<t}$ when fed with $a_{<t}$ and uniformly random bits:

$$\xi(e_{<t} \parallel a_{<t}) = \sum_{p:\, e_{<t} \sqsubseteq U(p, a_{<t})} 2^{-|p|}$$

AIMU $=$ optimal policy for computable measure μ
AINU $=$ optimal policy for semicomputable semimeasure ν
AIXI $=$ optimal policy for universal prior ξ

Results



Except for iterative AIXI, all of these bounds are sharp!

The Arithemetical Hierarchy

$A \subseteq \mathbb{N}$ is Σ_n^0 (A^c is Π_n^0) $:\Longleftrightarrow$ $\exists$ computable relation S such that

$$k \in A \iff \exists k_1 \forall k_2 \ldots Q_n k_n \; S(k, k_1, \ldots, k_n)$$

A is Δ_n^0 $:\Longleftrightarrow$ A is Σ_n^0 and A is Π_n^0.

Complexity of Induction

Leike and Hutter (2015b)

	Plain	Conditional
M	Σ_1^0	Δ_2^0
M_{norm}	Δ_2^0	Δ_2^0
$\overline{M}$	Π_2^0	Δ_3^0
$\overline{M}_{\text{norm}}$	Δ_3^0	Δ_3^0

All of these bounds are tight
(for particular universal Turing machines)

Environments That End

Option 1: 10 cookies and the universe ends
Option 2: 1 cookie, but the universe continues forever, and there are no more cookies
What is the *rational* choice?
Iterative value function: option 2
Recursive value function: option 1

Australian National University

References

Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer, 2005.

Jan Leike and Marcus Hutter. On the computability of AIXI. In *Uncertainty in Artificial Intelligence*, 2015a.

Jan Leike and Marcus Hutter. On the computability of Solomonoff induction and knowledge-seeking. 2015b. Forthcoming.

Ray Solomonoff. A formal theory of inductive inference. Parts 1 and 2. *Information and Control*, 7(1):1–22 and 224–254, 1964.

Ray Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory*, 24(4):422–432, 1978.