A new upper bound for Policy Iteration

Romain Hollanders
Balázs Gerencsér
Jean-Charles Delvenne
Raphaël Jungers

Université catholique de Louvain, ICTEAM
Belgium

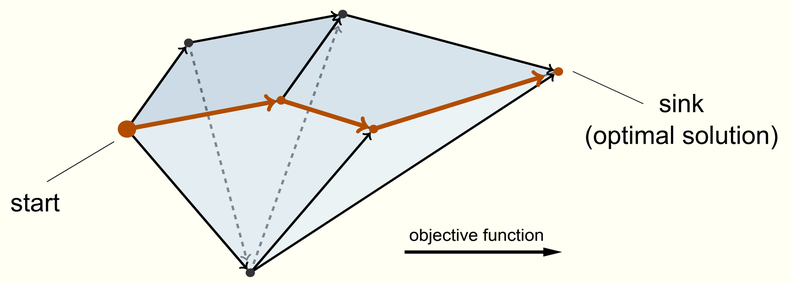
Linear Programming

Markov Decision Processes

Simplex algorithm

Follow the "best" outgoing edge around you and iterate until you reach the sink

Worst case : exponential complexity [Klee & Minty '72] [Friedmann '11]



Policy Iteration

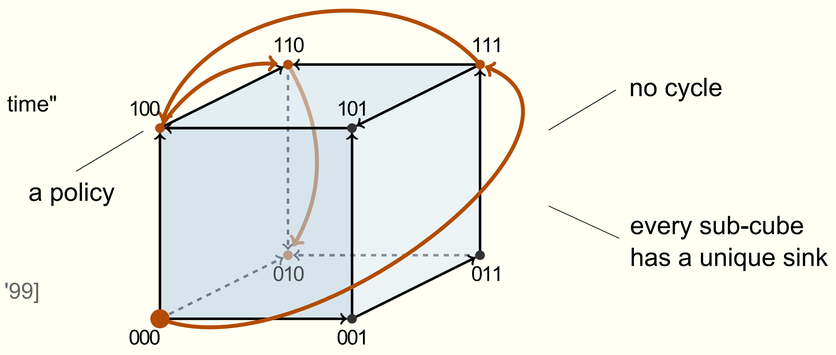
Follow all outgoing (improving) edges around you "at the same time" and iterate until you reach the sink

Worst case : exponential complexity [Fearnley '10]

Best lower bound : $\Omega\left(\sqrt[7]{2}^{n}\right)$ Iterations [Fearnley '10]

Best upper bound : $6 \cdot \frac{2^n}{n}$ Iterations [Mansour & Singh '99]

Can we get better bounds ?



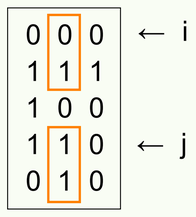
Relaxation 1 : Order-Regular matrices

Translate the cube structure into a matrix condition

$\Pi \in \{0,1\}^{m \times n}$ is Order-Regular

iff $\forall\ 1 \leq i < j < m,\ \exists\ 1 \leq k \leq n$

s.t.: $\Pi_{i,k} \neq \Pi_{i+1,k} = \Pi_{j,k} = \Pi_{j+1,k}$



Goal : Bound the number of rows (number of iterations or explored policies) of any Order-Regular matrix with n columns (dimensions or states)

Experimentally :

n	1	2	3	4	5	6	7
Maximum m	2	3	5	8	13	21	?

Conjecture [Hansen & Zwick '12]

Maximum $m = F_{n+2} = O(\phi^n)$ (Fibonacci numbers ; golden ratio)

Result 1 : for $n = 7$, we showed that $m \leq 33$ (using a "smart" exhaustive search ; deviates from the Fibonacci sequence)

The conjecture does not hold !

Using this relaxation, the best bound we can get for PI so far is $2^{n-1} + 1$...

Relaxation 2 : Pseudo-PI-sequences

Impose only two conditions on the sequence of policies explored by PI

Let $T_1 \prec ... \prec T_{2^n}$ be a total ordering of all distinct $T_i \in \{0,1\}^n$ (each T_i corresponds to the set of outgoing edges of a distinct policy of the cube)

Let $I \subseteq \{1, ..., 2^n\}$ (corresponds to the explored policies)

The set $\{i \in I\}$ is a Pseudo-PI-sequence iff I and the T_i's ensure :

The non-inclusion condition : $\forall\, i < j\ :\ \mathrm{supp}\,(T_i) \not\subseteq \mathrm{supp}\,(T_j)$

The jumping condition : $\forall\, i, j \in I, i < j\ :\ i + |\mathrm{supp}\,(T_i)| \leq j$

Goal : Bound the number of elements in any Pseudo-PI-sequence

Result 2 : the size of any Pseudo-PI-sequence is bounded above by $2 \cdot \frac{2^n}{n} + o\left(\frac{2^n}{n}\right)$

A new upper bound for Policy Iteration

n	1	2	3	4	5	6	7
Maximum m	2	3	5	8	13	21	33
$2 \cdot \frac{2^n}{n}$	4	4	5.3	8	12.8	21.3	36.6

Not such a bad fit for the real empirical bound

Result 3 : we can build a Pseudo-PI-sequence of size $2 \cdot \frac{2^n}{n} + o\left(\frac{2^n}{n}\right)$

The upper bound is optimal for the relaxation

Our construction of an extremal Pseudo-PI-sequence for n = 4 :

T_1 1111 $\prec$ T_2 1110 $\prec$ T_3 1101 $\prec$ T_4 1011 $\prec$ T_5 0111 $\prec$ T_6 1100 $\prec$ T_7 1010 $\prec$ T_8 1001 $\prec$ T_9 0110 $\prec$ T_{10} 0101 $\prec$ T_{11} 0011 $\prec$ T_{12} 1000 $\prec$ T_{13} 0100 $\prec$ T_{14} 0010 $\prec$ T_{15} 0001 $\prec$ T_{16} 0000

4 jumps ; 3 jumps ; 2 jumps ; 2 jumps ; 1 jump ; 1 jump ; 1 jump ; 1 jump

ordered by decreasing cardinality → non-inclusion condition OK

So $I = \{1, 5, 8, 10, 12, 13, 14, 15, 16\}$ and $|I| = 9 \sim 2 \cdot \frac{2^n}{n} + o\left(\frac{2^n}{n}\right)$ (chosen greedily such that the jumping condition is ensured)

More info on http:/perso.uclouvain.be/romain.hollanders