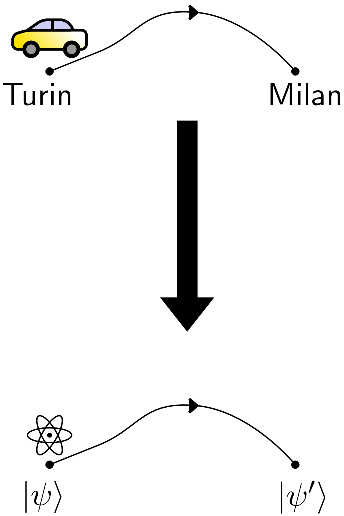
Quantum circuit design with Reinforcement Learning

Candidate: Francesco Montagna
Supervisor: Davide Girolami - DISAT

October 17, 2021

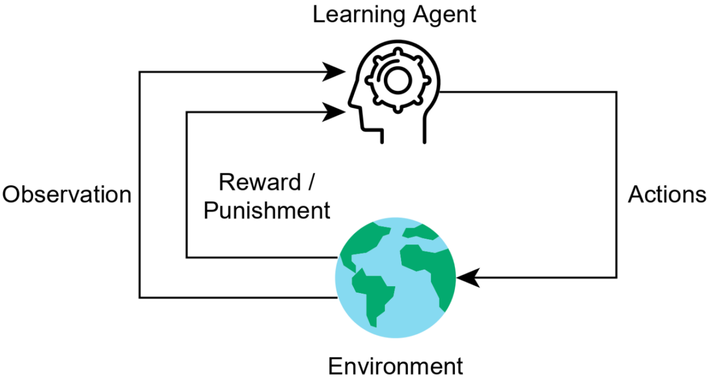
Politecnico di Torino

Overview



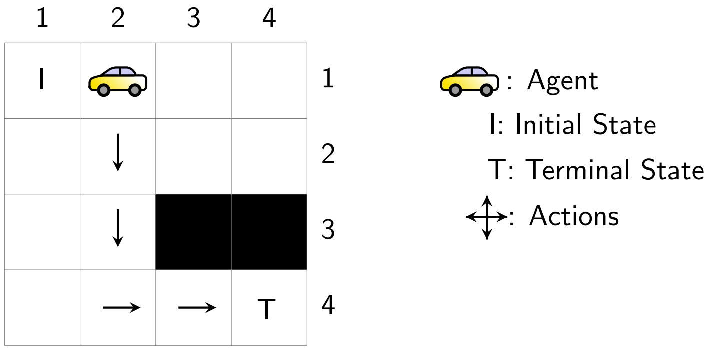
Reinforcement Learning

Agent and Environment interaction loop:



The Grid World Example

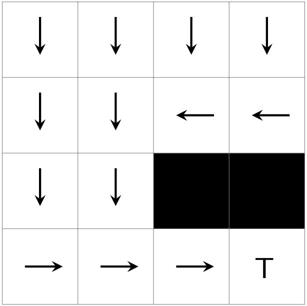
A canonical example of the reinforcement learning problem is the Grid World.



Reinforcement Learning

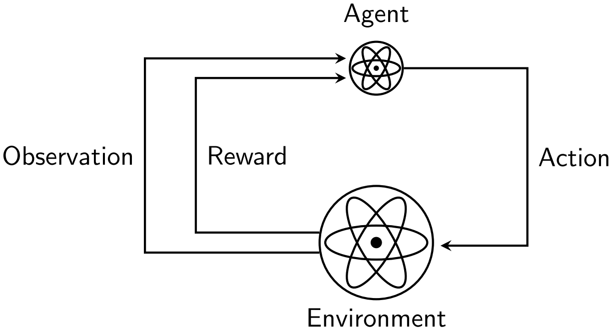
- Value function: $\mathbb{E}[\sum_t R_t | S_t = s]$
- Policy: $\pi(a|s)$

Policy example:



Quantum Mechanics and Computation

Agent and environment become *quantum systems* (e.g. an atom).



Postulate 1: *a physical system is associated to a complex vector space in which it is described by a vector.*

$$|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle \ ,$$

where $a_0, a_1 \in \mathbb{C}$.

Braket notation:

- $|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Postulate 2: *the state space of an N particles physical system is the tensor product of the state spaces of the single particle physical systems.*

Notation:

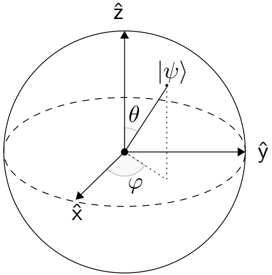
- $\otimes$: tensor product
- $|0\rangle \otimes |0\rangle = |00\rangle$

$$|\psi\rangle = a_{00}\,|00\rangle + a_{10}\,|10\rangle + a_{01}\,|01\rangle + a_{11}\,|11\rangle$$

The Qubit

- Classical bit: either 0 or 1.
- Qubit: superposition of $|0\rangle$ and $|1\rangle$, $|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle$

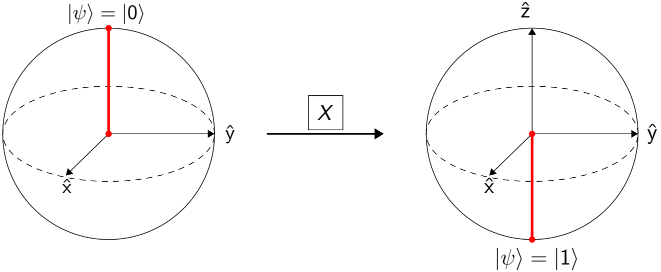
Bloch sphere representation of a qubit:



Gates and Circuits

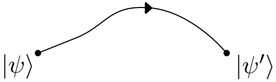
We define a quantum circuit as the combination of *wires* and *logic gates*.

$$|0\rangle \;\text{———}\; \boxed{X} \;\text{———}\; |1\rangle$$



Problem Statement

Prepare an initial state $|\psi\rangle$ into a target state $|\psi'\rangle$ minimizing the number of gates.



- Shortest path problem $\rightarrow$ *reinforcement learning* formulation.
- Algorithm limited to 2 qubits physical systems.

$$|00\rangle \longrightarrow \boxed{\text{Agent Designed Circuit}} \longrightarrow |11\rangle$$

The Method

Agent States

- $|\psi_t\rangle = a_{00,t}|00\rangle + a_{10,t}|10\rangle + a_{01,t}|01\rangle + a_{11,t}|11\rangle$.
- $a_{mn} \in \mathbb{C}$: $a_{mn} = \mathrm{Re}(a_{mn}) + j\,\mathrm{Im}(a_{mn})$

agent state at time t:

$$\begin{aligned} S_t =& [\mathrm{Re}(a_{00,t}), \mathrm{Re}(a_{10,t}), \mathrm{Re}(a_{01,t}), \mathrm{Re}(a_{11,t}), \\ & \mathrm{Im}(a_{00,t}), \mathrm{Im}(a_{10,t}), \mathrm{Im}(a_{01,t}), \mathrm{Im}(a_{11,t})] \end{aligned}$$

Actions

Actions are defined in terms of unitary gates applied to the quantum state:

- Rotations around X, Y, Z axes.
- $CNOT_{0,1}$: $CNOT_{0,1}|\alpha\rangle|\beta\rangle = |\alpha\rangle|\alpha \oplus \beta\rangle$

The Method

- $|\psi\rangle$: initial state vector.
- $|\psi_t\rangle = U|\psi\rangle = U_t U_{t-1} \dots U_1 |\psi\rangle$: state vector at time t.
- $|\psi'\rangle$: target state vector.
- $|\psi_T\rangle$: terminal state vector.

Fidelity: measures the similarity of two state vectors,

$$F(\psi', \psi_t) \ \in [0, 1] .$$

Terminal state: given a tolerance threshold ϵ, a state is terminal if it satisfies the condition

$$F(\psi', \psi_T) \in [1 - \epsilon, 1] .$$

Reward Function

$$R(s_t) = \begin{cases} 100, & \text{if } |\psi_t\rangle = |\psi_T\rangle \\ -1, & \text{else} \end{cases}$$

where s_t is the agent state at time t.
This choice forces the agent to minimize the number of exploited gates.

Results

Bell state	Fidelity	Number of gates
$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	1	2
$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	1	4
$\frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$	1	3
$\frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$	0.93	3

Table: Algorithm results on Bell states. Initial state: $|00\rangle$.

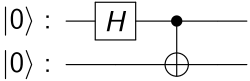


Figure: Circuit designed for the first Bell state as target.

IBM Quantum Lab

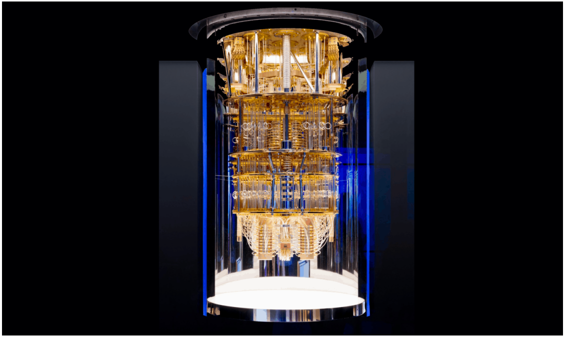


Figure: Quantum Computer at IBM (IBM Quantum Lab).

Experimental Results

The code for the experiments is written with Qiskit, a Python based framework for quantum computing, developed and maintained by IBM.

Fidelity between real and expected states:

$$F(\psi_{theory}, \psi_{exp}) = 0.94$$

Conclusion

Results:

- Optimal policy learned to prepare $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ from $|00\rangle$.
- Physical apparatus introduces noise: $F(\psi_{theory}, \psi_{exp}) = 0.94$.

Limitations:

- Low number of qubits
- Optimality not guaranteed

Future work

- Learn the reward given examples of optimal policies: Inverse Reinforcement Learning

Thank you for your attention