

# Baby Tetris First Part Report

ERGIN Seçkin Yağmur

GROUSSON Dylan

November 28, 2025

## 1 Question 1: Discounted MDP For Player

The objective of this part is to choose the optimal policy that maximizes the expected cumulative discounted reward over time for the "player".

### States

The state is represented by the current configuration of the grid, and the next piece to be placed.

$S := (\text{current grid, next piece to place})$

### Actions

The actions correspond to a position and orientation to place the piece on the grid.

$A := (\text{position, orientation})$

### Transition Function

The transition function is defined as:  $P(s'|s, a) := 1/2$

The reason for this is that there are two possible pieces (I or L) as the next piece to place, and considering our state representation, and our assumption of uniform distribution between the pieces, each transition has an equal probability of 1/2.

### Reward Function and Discount Factor

The reward function is defined as the number of lines cleared after placing the piece, and a discount factor  $\lambda$  is used to weigh future rewards.

$R(s, a, s') := \text{number of lines cleared and } \lambda \in [0, 1]$

### 1.1 Finding Optimal Policy

We use the value iteration method to find the optimal policy that maximizes the expected rewards. The value iteration algorithm iteratively updates the value function for each state based on the Bellman equation until it converges to the optimal value (margin  $\epsilon$ ).

$$V(s) = \max_{a \in A} \sum_{s' \in S} P(s'|s, a) \cdot (R(s, a, s') + \lambda V(s'))$$

## 2 Game Structure