Game Theory in Robotics: Pursuit-Evasion

Tomáš Kroupa

Artificial Intelligence Center
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague

2021

Pursuit-Evasion in Mobile Robotics

One or more pursuers try to capture one or more evaders who try to avoid capture.

- The study of motion planning problems in adversarial settings
  - Detecting intruders
  - Playing hide-and-seek
  - Catching burglars
- The planner seeks an optimal strategy against the worst-case adversary

Classes of Pursuit-Evasion Games

Differential

- Hamilton-Jacobi-Isaacs differential equations model the dynamics
- Their solutions are players' strategies as control inputs for achieving the objectives
- 👍 Velocity or acceleration are expressed explicitly as differential constraints
- 👎 The resulting equations are very complicated and difficult to solve

Combinatorial

- A real environment is modeled as a polygon or graph
  - *The Cops and Robbers Game*
  - *Parson's game*
  - *The lion-and-man game*
- 👍 Complexity results and guarantees in terms of the size of game
- 👎 Abstraction from the continuous features of environment

Lecture Goals/Outline

To understand how

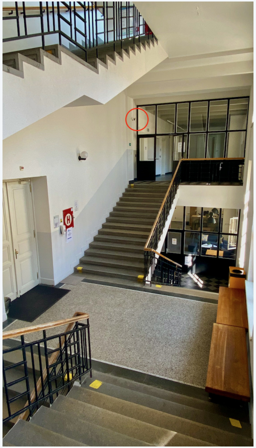
- *the robotic motion planning changes in the presence of an adversary pursuing their own goals and*
- *the robot's navigation can be enhanced using the game-theoretic methods in this case.*

1. Motivation: A simple path planning problem
2. The robust path planning problem as a two-player zero-sum game
3. Dynamics of pursuers/evaders can be modeled as a stochastic game

Robust Path Planning Problem

What Path Should the Robot Follow to Avoid CCTV?

👉 The position of cameras is known



What Path Should the Robot Follow to Avoid CCTV?

👉 The position of cameras is known

The planner navigates a robot to a goal location in a previously mapped environment. The adversary is a part of the environment!

Planner

- Models the problem as a single-agent *Markov decision process*
- Must find a path minimizing the robot's visibility to cameras

Adversary

- Places cameras to detect the robot
- Has no strategic goals

What Path Should the Robot Follow to Avoid CCTV?

👉 What are optimal camera locations?

Robust Path Planning Problem

What Path Should the Robot Follow to Avoid CCTV?

👉 What are optimal camera locations?

Both the planner and adversary can control the environment actively.

Planner

- Path π for the robot
- Finite set of paths Π
- Probability distribution $p \in \Delta_\Pi$

Adversary

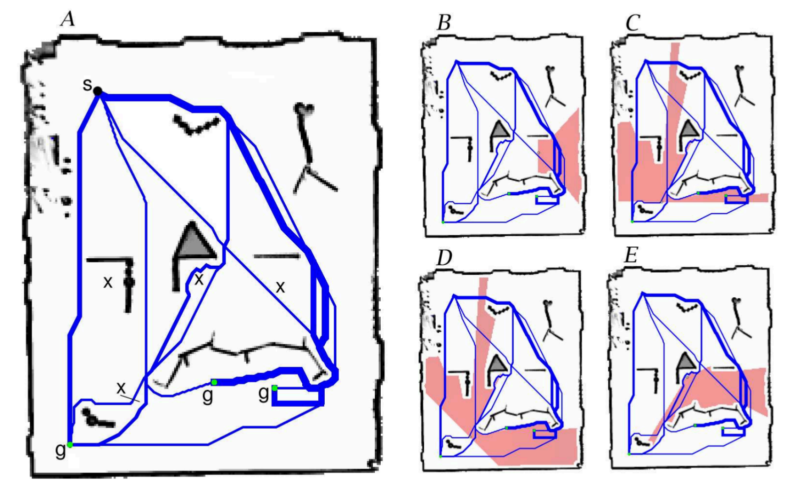
- Cost vector $\mathbf{c}$
- Finite set of cost vectors C
- Probability distribution $q \in \Delta_C$

Let $V(\pi, \mathbf{c})$ be the value of policy π and cost vector $\mathbf{c}$. Solve:

$$\min_{p \in \Delta_\Pi} \max_{q \in \Delta_C} \sum_{\pi \in \Pi} \sum_{\mathbf{c} \in C} p(\pi) q(\mathbf{c}) V(\pi, \mathbf{c})$$

Example of Solution

Blum et al. (2003)



Robust Path Planning Problem

Planning Paths: Experiments

Blum et al. (2003)

- The gridworld of size up to 269×226
- The robot can move in any of 16 compas directions
- Each cell has cost 1 and a cost proportional to the distance of camera

Computational limits

- Sets Π and C should be reasonably small
- Already $\binom{100}{2} = 4\,950$ positions exist for 2 cameras in the gridworld 10×10

Two-Player Zero-Sum Games

Two-Player Zero-Sum Game

1. Players/agents are the planner and the adversary
2. Strategy sets are M and N
3. The matrix $\mathbf{C} = [c_{ij}]_{i \in M,\, j \in N}$ of costs for the planner

For example:

$$|M| = 2, \qquad |N| = 4, \qquad \mathbf{C} = \begin{bmatrix} 1 & 0 & 4 & -1 \\ -1 & 1 & -2 & 5 \end{bmatrix}$$

The zero-sum assumption means

the loss of 4 for the planner $=$ the gain of 4 for the adversary.

Minmax/Maxmin Strategies

👉 We seek the optimal performance against the worst-case adversary

- Assume that the agents adopt maxmin/minmax strategies $\bar{i} \in M$ and $\bar{j} \in N$:

$$\bar{i} = 1, \qquad \bar{j} = 2, \qquad \mathbf{C} = \begin{bmatrix} 1 & 0 & 4 & -1 \\ -1 & 1 & -2 & 5 \end{bmatrix}$$

- The floor on the profit of adversary $(0) \leq$ the ceiling of the cost of planner (4):

$$\max_{j \in N} \min_{i \in M} c_{ij} \;\leq\; c_{\bar{i}\,\bar{j}} \;\leq\; \min_{i \in M} \max_{j \in N} c_{ij}$$

- However, the adversary can increase the profit by playing $j = 3$
- In this case the planner would adopt $i = 2$
- Then the adversary would play $j = 4$ etc.

Mixed Strategies

A mixed strategy of a player is a probability distribution over the strategy set.

- Let Δ_M and Δ_N be the sets of mixed strategies of planner/adversary
- If the agents play $\mathbf{x} \in \Delta_M$ and $\mathbf{y} \in \Delta_N$, the expected loss of planner is

$$\sum_{i \in M} \sum_{j \in N} x_i y_j c_{ij} = \mathbf{x}^\top \mathbf{C} \mathbf{y}$$

In particular, if the adversary uses a pure strategy $\mathbf{e}_j \in \Delta_N$ with $j \in N$,

$$\sum_{i \in M} x_i c_{ij} = \mathbf{x}^\top \mathbf{C} \mathbf{e}_j$$

Minmax/Maxmin in Mixed Strategies

1. A minmax strategy of the planner is a mixed strategy $\bar{\mathbf{x}} \in \Delta_M$ such that

$$\max_{\mathbf{y}\in\Delta_N} \bar{\mathbf{x}}^\top \mathbf{C}\mathbf{y} = \min_{\mathbf{x}\in\Delta_M} \max_{\mathbf{y}\in\Delta_N} \mathbf{x}^\top \mathbf{C}\mathbf{y}$$

2. A maxmin strategy of the adversary is a mixed strategy $\bar{\mathbf{y}} \in \Delta_N$ such that

$$\min_{\mathbf{x}\in\Delta_M} \mathbf{x}^\top \mathbf{C}\bar{\mathbf{y}} = \max_{\mathbf{y}\in\Delta_N} \min_{\mathbf{x}\in\Delta_M} \mathbf{x}^\top \mathbf{C}\mathbf{y}$$

It is easy to show that

$$\underbrace{\max_{\mathbf{y}\in\Delta_N} \min_{\mathbf{x}\in\Delta_M} \mathbf{x}^\top \mathbf{C}\mathbf{y}}_{\text{The lower bound on the profit}} \leq \bar{\mathbf{x}}^\top \mathbf{C}\bar{\mathbf{y}} \leq \underbrace{\min_{\mathbf{x}\in\Delta_M} \max_{\mathbf{y}\in\Delta_N} \mathbf{x}^\top \mathbf{C}\mathbf{y}}_{\text{The upper bound on the cost}}$$

Minimax Theorem

👉 von Neumann, 1928

$$\underbrace{\min_{\mathbf{x}\in\Delta_M}\max_{\mathbf{y}\in\Delta_N}\ \mathbf{x}^\top\mathbf{C}\mathbf{y}}_{\text{The value of the game}} = \max_{\mathbf{y}\in\Delta_N}\min_{\mathbf{x}\in\Delta_M}\ \mathbf{x}^\top\mathbf{C}\mathbf{y}$$

- An equilibrium is a pair of minmax/maxmin strategies $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$
- For any equilibrium $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, we obtain

$$\bar{\mathbf{x}}^\top\mathbf{C}\bar{\mathbf{y}} = \text{the value of the game}$$

Computing Minmax Strategy

Linear programming

We can write

$$\min_{\mathbf{x}\in\Delta_M} \max_{\mathbf{y}\in\Delta_N} \mathbf{x}^\top \mathbf{C}\mathbf{y} = \min_{\mathbf{x}\in\Delta_M} \max_{j\in N} \mathbf{x}^\top \mathbf{C}\mathbf{e}_j.$$

This minmax problem is equivalent to a linear program:

$$\begin{aligned} \text{Minimize} \quad & \max_{j\in N} \mathbf{x}^\top \mathbf{C}\mathbf{e}_j \\ \text{subject to} \quad & \mathbf{x} \in \Delta_M \end{aligned}$$

$$\begin{aligned} \text{Minimize} \quad & v \\ \text{subject to} \quad & \mathbf{x}^\top \mathbf{C}\mathbf{e}_j \leq v, \quad \forall j \in N \\ & x_i \geq 0, \quad \forall i \in M \\ & \sum_{i\in M} x_i = 1 \\ & v \in \mathbb{R} \end{aligned}$$

Computing Minmax Strategy

Example

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 4 & -1 \\ -1 & 1 & -2 & 5 \end{bmatrix}$$

$$\begin{aligned} \text{Minimize} \quad & v \\ \text{subject to} \quad & x_1 - x_2 \le v \\ & x_2 \le v \\ & 4x_1 - 2x_2 \le v \\ & -x_1 + 5x_2 \le v \\ & x_1, x_2 \ge 0 \\ & x_1 + x_2 = 1 \\ & v \in \mathbb{R} \end{aligned}$$

The equilibrium strategies are $\bar{\mathbf{x}} = (\frac{7}{12}, \frac{5}{12})$, $\bar{\mathbf{y}} = (0, 0, \frac{1}{2}, \frac{1}{2})$, and $\bar{v} = \frac{3}{2}$.

- 👉 The strategy sets M and N are too large in the path planning problems
- 👉 The cost matrix $\mathbf{C}$ may not be explicitly given

We show an iterative method relying on 2 principles:

1. Small subgames can be solved efficiently
2. Subgames are expanded with best responses

The best response of planner to a mixed strategy $\mathbf{y} \in \Delta_N$ is $i \in M$ minimizing $\mathbf{e}_i^\intercal \mathbf{C}\mathbf{y}$.

Double Oracle Algorithm

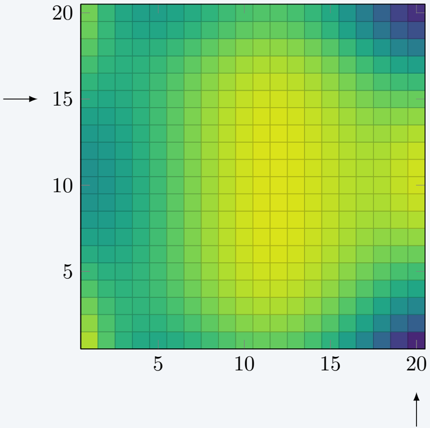
Blum et al. (2003)

1. Pick initial subsets of strategies for each player
2. Compute an equilibrium of the subgame — MASTER PROBLEM
3. Expand the current strategy sets with the best responses — SUB-PROBLEM
4. Repeat 2. and 3. until the current equilibrium is good enough

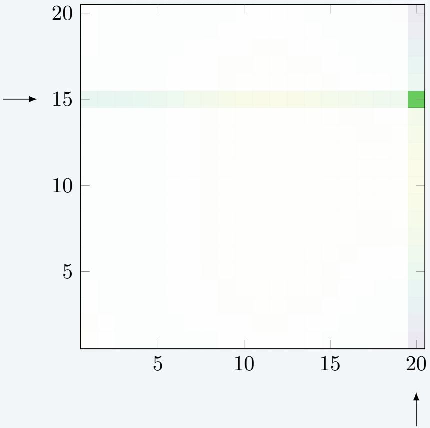
Double Oracle Algorithm

Initialize

Initialize with random pure strategies.



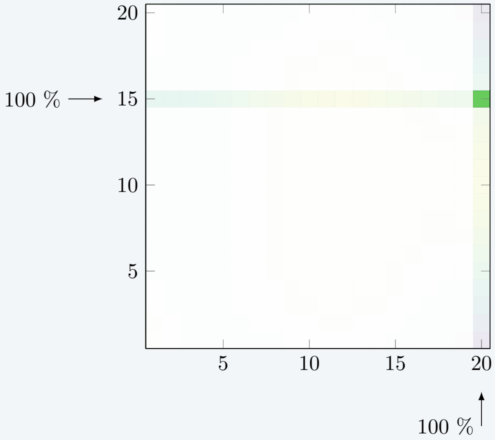
Find an equilibrium of the 1×1 subgame.



Double Oracle Algorithm

Find an equilibrium of the 1×1 subgame.

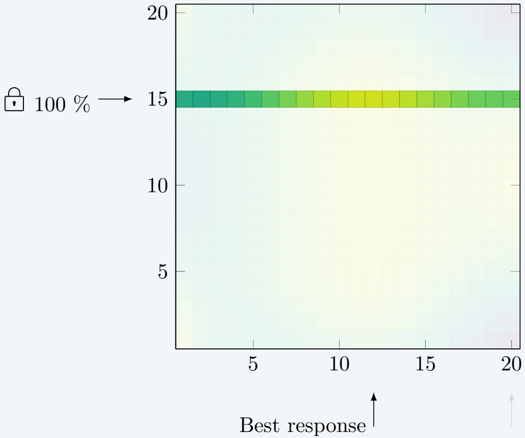
Master Problem



Double Oracle Algorithm

Best Response (adversary)

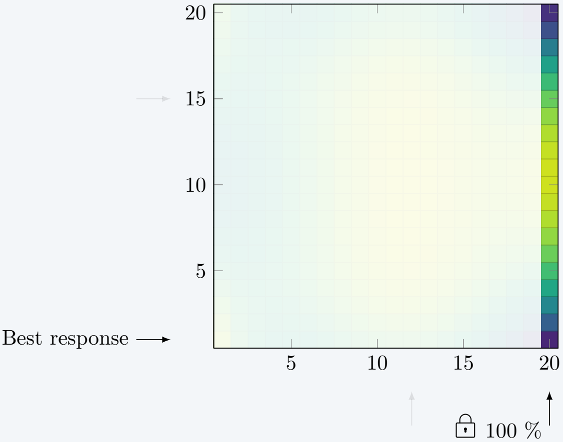
Find adversary's best response against a fixed strategy of the planner.



Double Oracle Algorithm

Find planner's best response against a fixed strategy of the adversary.

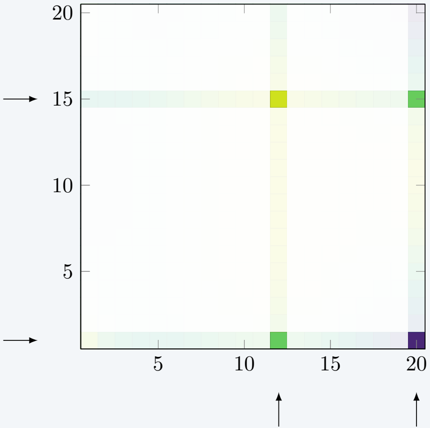
Best Response (planner)



Double Oracle Algorithm

Master Problem (Iteration 2)

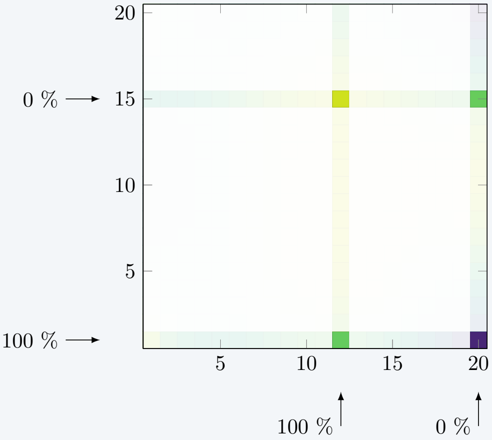
Find an equilibrium of the 2×2 subgame.



Double Oracle Algorithm

Find an equilibrium of the 2×2 subgame.

Master Problem (Iteration 2)

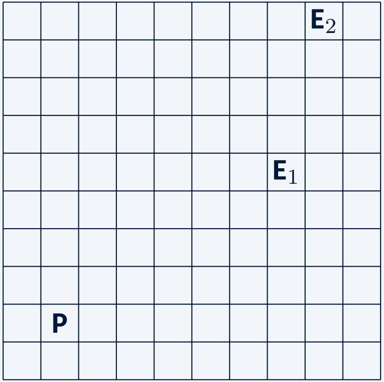


- 👍 The algorithm recovers an exact equilibrium in finitely many steps
- 👍 The approximation of equilibrium/value of the game
- 👍 Easy to implement using efficient LP solvers
- 👎 It may need $O(|M| + |N|)$ iterations

Stochastic Games

Repeating Zero-Sum Games

👉 Or making MDPs competitive?



- The pursuer **P** tries to capture evaders $\mathbf{E}_1$ and $\mathbf{E}_2$
- Stochastic policy describes the mixed strategy of each player in every state
- We are seeking a common generalization of
  - two-person zero-sum games and
  - Markov decision processes (MDPs)

1. The planner and the adversary
2. Strategy sets M and N
3. The set S of states
4. The transition function

$$T : S \times M \times N \to \Delta_S$$

where $T(s, i, j)$ denotes the probability distribution on S

5. The reward function

$$R : S \times M \times N \to \mathbb{R}$$

where $R(s, i, j)$ is the reward to the planner

Policy

A policy of the planner is a mapping

$$\pi : S \to \Delta_M$$

from states to mixed strategies, and analogously for the adversary.

Special cases

- If $|S| = 1$, then we obtain a two-person zero-sum game and the concept of mixed strategy
- If $|N| = 1$, then we get an MDP with the concept of stochastic policy

How to evaluate policies?

Discounting

👉 How much future rewards effect optimal decisions?

- The policies of the planner and adversary determine a random reward R_{t+k} received k steps into the future at time t
- A discount factor is a number $0 \leq \gamma < 1$
- The goal of planner is to maximize the expected discounted reward at time t,

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k}\right]$$

Value/Quality Function

- Let $\mathcal{V}(s)$ be the expected reward for the optimal policy from state $s \in S$
- Let $\mathcal{Q}(s, i, j)$ be the expected reward for action $i \in M$ from state $s \in S$ when
  1. the adversary selects strategy $j \in M$ and
  2. the planner continues optimally thereafter

$$\mathcal{Q}(s,i,j) = R(s,i,j) + \gamma \sum_{s' \in S} T(s,i,j)(s') \cdot \mathcal{V}(s')$$

$$\mathcal{V}(s) = \max_{\pi_s \in \Delta_M} \min_{j \in N} \sum_{i \in M} \mathcal{Q}(s,i,j) \cdot \pi_s(i)$$

Value Iteration

A variant for stochastic games

$$\mathcal{Q}(s,i,j) := R(s,i,j) + \gamma \sum_{s' \in S} T(s,i,j)(s') \cdot \mathcal{V}(s') \tag{1}$$

$$\mathcal{V}(s) := \max_{\pi_s \in \Delta_M} \min_{j \in N} \sum_{i \in M} \mathcal{Q}(s,i,j) \cdot \pi_s(i) \tag{2}$$

- Start with an estimate of value function $\mathcal{V}$
- Update (1)–(2) iteratively
- This procedure converges to the optimal values (Shapley, 1953)

References

- Chung, Timothy H., Geoffrey A. Hollinger, and Volkan Isler. Search and Pursuit-Evasion in Mobile Robotics. *Autonomous Robots* 31 (4): 299–316, 2011.
- McMahan, H. Brendan, Geoffrey J. Gordon, and Avrim Blum. Planning in the Presence of Cost Functions Controlled by an Adversary. In *Proceedings of the 20th International Conference on Machine Learning* (ICML-03), 2003.
- Littman, Michael L. 1994. Markov Games as a Framework for Multi-Agent Reinforcement Learning. *Machine Learning Proceedings*, 1994. https://doi.org/10.1016/b978-1-55860-335-6.50027-1.