15-780: Graduate AI

Lecture 8. Games

Geoff Gordon (this lecture)
Ziv Bar-Joseph
TAs Michael Benisch, Yang Gu
Last time, on Grad AI
Optimization

- Unconstrained optimization: gradient $= 0$
- Equality-constrained optimization
  - Lagrange multipliers
- Inequality-constrained: either
  - nonnegative multipliers (last $W$), or
  - search through bases (simplex, on $M$)
Duality

- How to express path planning as an LP
- Dual of path planning LP
Optimization in ILPs

- DFS, with pruning by:
  - constraint propagation
  - best solution so far
  - dual feasible solution
  - dual feasible solution for relaxation of ILP with some variables set (branch and bound)
Optimization in ILPs

- Duality gap and Slater’s condition
- Cutting planes (how to use, how to find)
  - generally, e.g., Gomory
  - problem specific, e.g., subtour elimination for TSPs
- Branch and cut
Historical note

- Gomory’s cuts weren’t the first poly-time cuts: e.g., Dantzig in 1959
- But they were first to guarantee finite termination of cutting plane method for ILPs
- Proven by Gomory in 1963
Game search
Synthetic example
Principal variation
Making it work

- Minimax is all well and good for small games
- But what about bigger ones? 2 answers:
  - cutting off search early (big win)
  - pruning (smaller win but still useful)
Heuristics

- Quickly and approximately evaluate a position without search
- E.g., \( Q = 9, R = 5, B = N = 3, P = 1 \)
- Build out game tree as far as we can, use heuristic at leaves in lieu of real value
  - might want to build it out unevenly
  (more below)
Heuristics

- Deep Blue used: materiel, mobility, king position, center control, open file for rook, paired bishops/rooks, … (> 6000 total features!)
- Weights are context dependent, learned from DB of grandmaster games then hand tweaked
Quiescence
Pruning

- Idea: don’t bother looking at parts of the tree we can prove are irrelevant
Pruning example
Pruning example

```
max
min
max
```

![Pruning example diagram](image)
Alpha-beta pruning

- Do a DFS through game tree
- At each node $n$ on stack, keep bounds
  - $\alpha(n)$: value of best deviation so far for MAX along path to $n$
  - $\beta(n)$: value of best deviation so far for MIN along path to $n$
Alpha-beta pruning

- Deviation = way of leaving the path to $n$
- So, to get $\alpha$,
  - take all MAX nodes on path to $n$
  - look at all their children that we’ve finished evaluating
  - best (highest) of these children is $\alpha$
- Lowest of children of MIN nodes is $\beta$
Example of alpha and beta
Alpha-beta pruning

- At max node:
  - receive $\alpha$ and $\beta$ values from parent
  - expand children one by one
  - update $\alpha$ as we go
  - if $\alpha$ ever gets higher than $\beta$, stop
  - won’t ever reach this node (return $\alpha$)
Alpha-beta pruning

- At min node:
  - receive $\alpha$ and $\beta$ values from parent
  - expand children one by one
  - update $\beta$ as we go
  - if $\beta$ ever gets lower than $\alpha$, stop
  - won’t ever reach this node (return $\beta$)
Example
How much do we save?

- **Original tree:** $b^d$ nodes
  - $b = \text{branching factor}$
  - $d = \text{depth}$

- If we expand children in random order, pruning will touch $b^{3d/4}$ nodes

- Lower bound (best node first): $b^{d/2}$

- Can often get close to lower bound w/ **move ordering heuristics**
Matrix games

- *Games where each player chooses a single move (simultaneously with other players)*
- *Also called normal form games*
- *Simultaneous moves cause uncertainty: we don’t know what other player(s) will do*
Acting in a matrix game

- One of the simplest kinds of games; we’ll get more complicated later in course
- But still will make us talk about
  - negotiation
  - cooperation
  - threats, promises
Matrix game: prisoner’s dilemma

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
<td>-1</td>
<td>-9</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>0</td>
<td>-5</td>
</tr>
</tbody>
</table>

Payoff to Row

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>-9</td>
<td>-5</td>
</tr>
</tbody>
</table>

Payoff to Col
Matrix game: prisoner’s dilemma

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>-1, -1</td>
<td>-9, 0</td>
</tr>
<tr>
<td>$D$</td>
<td>0, -9</td>
<td>-5, -5</td>
</tr>
</tbody>
</table>
Can also have n-player games

<table>
<thead>
<tr>
<th></th>
<th>$H$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0, 0, 1</td>
<td>0, 0, 1</td>
</tr>
<tr>
<td>$T$</td>
<td>0, 0, 1</td>
<td>1, 1, 0</td>
</tr>
</tbody>
</table>

if Layer plays $H$

<table>
<thead>
<tr>
<th></th>
<th>$H$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>1, 1, 0</td>
<td>0, 0, 1</td>
</tr>
<tr>
<td>$T$</td>
<td>0, 0, 1</td>
<td>0, 0, 1</td>
</tr>
</tbody>
</table>

if Layer plays $T$
Analyzing a game

- What do we want to know about a game?
- Value of a joint action: just read it off of the table
- Value of a mixed joint strategy: almost as simple
Value of a mixed joint strategy

Suppose Row plays 30-70, Col plays 60-40

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>.6*.3*w</td>
<td>.4*.3*x</td>
</tr>
<tr>
<td>D</td>
<td>.6*.7*y</td>
<td>.4*.7*z</td>
</tr>
</tbody>
</table>
Payoff of joint strategy

- *Just an average over elements of payoff matrices $M_R$ and $M_C$*

- *If $x$ and $y$ are strategy vectors like $(.3, .7)'$ then we can write $x'M_Ry$ and $x'M_Cy$*
What else?

- Could ask for value of a strategy $x$ under various weaker assumptions about other players’ strategies $y, z, \ldots$

- Weakest assumption: other players might do absolutely anything!

- How much does a strategy guarantee us in the most paranoid of all possible worlds?
Safety value

- **Worst-case value of a row strategy** $x$ in 2-player game is
  - $\min_y x' M_R y$
- **More than two players**, $\min$ over $y$, $z$, …
- **Best worst-case value is safety value or minimax value of game**
  - $\max_x \min_y x' M_R y$
What else?

- If the world really is out to get us, the safety value is the end of the story
- This is the case in...
Zero-sum games
Zero-sum game

- A 2-player matrix game where
  \[(\text{payoff to A}) = -(\text{payoff to B})\] for all combinations of actions

- Note: 3-player games are never called zero-sum, even if payoffs add to 0

- But if \[(\text{payoff to A}) = 7 - (\text{payoff to B})\] we sometimes fudge and call it zero-sum
Zero-sum: matching pennies

<table>
<thead>
<tr>
<th></th>
<th>$H$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$T$</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>
Minimax

- In zero-sum games, safety value for Row is negative of safety value for Col (famous theorem of Nash)
- A strategy that guarantees minimax value is a minimax strategy
- If both players play such strategies, we are in a minimax equilibrium
  - no incentive for either player to switch
Finding minimax

\[ \min_x \max_y x'My \text{ subject to} \]
\[ 1'x = 1 \]
\[ 1'y = 1 \]
\[ x, y \geq 0 \]
For example

\[ \begin{align*}
\min & \quad \max \quad x \quad y \\
\text{s.t.} & \quad x_n y_n + x_T y_T - x_n y_T - x_T y_n \\
& \quad x_n + x_T = 1 \\
& \quad y_n + y_T = 1 \\
& \quad x, y \geq 0
\end{align*} \]
Finding minimax

- Eliminate x’s equality constraint:
  - \( \min_x \max_y, z \ z(1 - 1'x) + x'My \) subject to
    - \( 1'y = l \)
    - \( x, y \geq 0 \)
Finding minimax

- Gradient wrt $x$ is
  - $My - 1z$

- $\max y, z \ z \ \text{subject to}$
  - $My - 1z \geq 0$
  - $1'y = 1$
  - $y \geq 0$
For example

\[
\begin{align*}
\max_z & \quad z \\
\text{s.t.} & \quad y_i^z \\
& \quad z \leq y_i^f - y_i^r \\
& \quad z \geq y_i^r + y_i^l \\
& \quad y_i^l + y_i^r = 1 \\
& \quad x, y \geq 0
\end{align*}
\]
Interpreting LP

- \( \max y, z \) subject to
  
  \[ My \geq 1z \]
  
  \[ 1'y = 1 \]
  
  \[ y \geq 0 \]

- \( y \) is a strategy for Col; \( z \) is value of this strategy
\[ z \leq J_H - (1 - y_H) \]

\[ z \leq y_H + (1 - y_H) \]

Value if \( x \) plays \( T \)

Value if \( x \) plays \( H \)

Optimal

Feasible
Duality

- $x$ is dual variable for $My \geq 1z$
- Complementarity: Row can only play strategies where $My = 1z$
- Makes sense: others cost more
- Dual of this LP looks the same, so Col can only play strategies where $x'M$ is maximal
Back to general-sum

- What if the world isn’t really out to get us?
- Minimax strategy is unnecessarily pessimistic
General-sum equilibria
Lunch

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>3, 4</td>
<td>0, 0</td>
</tr>
<tr>
<td>$U$</td>
<td>0, 0</td>
<td>4, 3</td>
</tr>
</tbody>
</table>

$A = \text{Ali Baba}, \ U = \text{Union Grill}$
Pessimism

- In Lunch, safety value is $12/7 < 2$
- Could get 3 by suggesting less-preferred restaurant
- Any halfway-rational player will cooperate with this suggestion
Rationality

- Trust the other player to look out for his/her own best interests
- Stronger assumption than “s/he might do anything”
- Results in possibility of higher-than-safety payoff
Dominated strategies

- First step towards being rational: if a strategy is bad no matter what the other player does, don’t play it!
- Such a strategy is (strictly) dominated
- Strict = always worse (not just the same)
- Weak = sometimes worse, never better
### Eliminating dominated strategies

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1, -1</td>
<td>-9, 0</td>
</tr>
<tr>
<td>D</td>
<td>0, -9</td>
<td>-5, -5</td>
</tr>
</tbody>
</table>

The dominated strategy is eliminated by removing the red lines.
Do we always get a unique answer?

- No: try Lunch
- What can we do instead?
- Well, what was special about Row offering to play A?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 4</td>
<td>0, 0</td>
</tr>
<tr>
<td>U</td>
<td>0, 0</td>
<td>4, 3</td>
</tr>
</tbody>
</table>
Equilibrium

- If Row says s/he will play A, Col’s best response is to play A as well.
- And if Col plays A, then Row’s best response is also A.
- So (A, A) is a mutually reinforcing pair of strategies—an equilibrium.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 4</td>
<td>0, 0</td>
</tr>
<tr>
<td>U</td>
<td>0, 0</td>
<td>4, 3</td>
</tr>
</tbody>
</table>
Finding equilibria

- The idea of equilibrium allows us to rule out some more joint strategies beyond what dominance gave us.
- The particular type of equilibrium we are about to describe is due to Nash.
  - his name keeps coming up...
Finding equilibria

- In a Nash equilibrium, we have a (mixed) strategy for each player
- Each strategy is a best response to others
  - puts zero weight on suboptimal actions
  - therefore zero weight on dominated actions
\[ z = y_n + (1 - y_n) \]

\[ z = \frac{1}{1 - y_n} \]

\[ y_n \]

If \( x \) plays

70% \( T \)
How good is equilibrium?

- Does an equilibrium tell you how to play?
- Sadly, no.
- To get further, we’ll need additional assumptions
Bargaining
Bargaining

- In the standard model of a matrix game, players can’t communicate
- To allow for bargaining, we will extend the model two ways:
  - first, *cheap talk*
  - second, a *moderator*
Cheap talk

- Players get a chance to talk to one another before picking their actions
- They can say whatever they want—lie, threaten, cajole, or even be honest
- What will happen?
Coordination

- Certainly the players will try to coordinate
- That is, they will try to agree on an equilibrium
  - agreeing on a non-equilibrium will lead to deviation
- But which one?
Pareto dominance
Pareto dominance

- In Lunch, there are 3 Nash equilibria
- Players could agree on any one, or agree to randomize among them
  - e.g., each simultaneously say a binary number, XOR together, use result to pick equilibrium
Pareto dominance

- Not all equilibria are created equal
- For any in brown triangle’s interior, there is one on red line that’s better for both players
- Red line = Pareto dominant
Beyond Pareto

- We still haven’t achieved our goal of actually predicting what will happen
- We’ve narrowed it down a lot: Pareto-dominant equilibria
- Further narrowing is the subject of much argument among game theorists
Nash designed a model of the bargaining process (there’s that name again…)

Rubinstein later made the model more detailed and implementable

Model includes offers, threats, and impatience to reach an agreement

In this model, we finally have a unique answer to “what will happen?”
Nash bargaining solution

- Predicts players will agree on the point on Pareto frontier that maximizes product of extra utility
- Invariant to axis rescaling, player exchanging
Moderator

- A moderator has a big deck of cards
- Each card has a recommended action for each player
- Moderator draws a card, whispers actions to corresponding players
  - actions may be correlated
  - only find out your own
Correlated equilibrium

- Since players can have correlated actions, an equilibrium with a moderator is called a correlated equilibrium

- Example: 5-way stoplight

- All NE are CE

- At least as many CE as NE in every game (often strictly more)
Realism?

- Moderators are often available
- Sometimes have to be kind of clever
- E.g., can simulate a moderator using cheap talk and some crypto
Correlated equilibrium

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 4</td>
<td>0, 0</td>
</tr>
<tr>
<td>U</td>
<td>0, 0</td>
<td>4, 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>U</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>
Correlated equilibrium

- Probability that Row is recommended to play A = a + b
- Given recommendation for A, probability that Col also plays A = a / (a + b)
- Rationality: when I’m recommended to play A, I don’t want to play B instead
Correlated equilibrium

\[
4 \frac{a}{a+b} + 0 \frac{b}{a+b} \geq 0 \frac{a}{a+b} + 3 \frac{b}{a+b} \quad \text{if} \ a + b > 0
\]

\[
4a + 0b \geq 0a + 3b
\]

\[
0c + 3d \geq 4c + 0d
\]

\[
0b + 4d \geq 3b + 0d
\]

\[
3a + 0c \geq 0a + 4c
\]
Correlated equilibrium

Haykin, Principe, Sejnowski, and McWhirter: New Directions in Statistical Signal Processing: From Systems to Brain...
Bargaining

- Can use Nash bargaining model to select among CE
- Same results hold: unique answer on Pareto frontier (but now Pareto frontier might be better)