Simple Stochastic Games

Lecture notes for the course "Games on Graphs"

B. Srivathsan

Chennai Mathematical Institute, India

We will use definitions from [Con92].

Definition 1 (Simple Stochastic Games (SSG)) A *Simple Stochastic Game (SSG) G* is given by a graph $(V = V_{avg} \sqcup V_{max} \sqcup V_{min}, E)$. The vertex set is of the form $\{1, 2, \ldots, n-1, n\}$. From every vertex, there are two outgoing edges. Note that both the outgoing edges can lead to the same vertex. Each outgoing edge from vertices in V_{avg} is marked with probability $\frac{1}{2}$. Vertex $n-1$ is called the 0-sink and vertex n is called the 1-sink. The only edges from $n-1$ and n are self loops.

For convenience, we will call vertices in V_{max} as *max vertices*, the ones in V_{min} as *min vertices* and vertices in V_{avg} as *average vertices*.

The game is played by two players Max and Min. A token is placed at some vertex. If it is a max (resp. min) vertex, player Max (resp. Min) moves it to one of the successors. If it is an average vertex, the token moves to each successor with probability $\frac{1}{2}$.

Definition 2 (Strategy/Policy) A *strategy* σ for Max is a function $\sigma : V_{max} \to V$ such that $(v, \sigma(v) \in E$. The strategy chooses an edge for every max vertex. Strategies are also called *policies*. Similarly, we can define a strategy τ for Min as a function $\tau : V_{min} \to V$.

Each pair of strategies (σ, τ) for G gives a Markov Chain $G_{\sigma,\tau}$ obtained by restricting the graph to the edges given by the strategies. An SSG is said to be stopping if for every σ, τ, the SSG $G_{\sigma,\tau}$ is stopping. For the Markov Chain $G_{\sigma,\tau}$, we know the vector denoting reachability probabilities $v_{\sigma,\tau}$.

We will now define a value vector for SSGs.

Definition 3 (Value vector for SSGs) For an SSG G, we define its value vector as:

$$\overline{v} := \begin{bmatrix} \max_\sigma \min_\tau v_{\sigma,\tau}(1) \\ \max_\sigma \min_\tau v_{\sigma,\tau}(2) \\ \vdots \\ \max_\sigma \min_\tau v_{\sigma,\tau}(n) \end{bmatrix}$$

Notice that we have chosen max min, whereas we could have also considered min max. Later, we will show that this order does not matter: replacing max min in the above definition with min max gives the same value.

1 Charactering value vector using constraints

Similar to the constraints for MDPs, we will now give constraints for SSGs. Given an SSG G, consider the following set of constraints over the variables $\langle w_1, w_2, \ldots, w_n \rangle$:

$$\begin{aligned}
& 0 \leq w_i \leq 1 && \text{for every } i \\
& w_n = 1 && \\
& w_{n-1} = 0 && \\
\text{for each } i \leq n-2 \text{ in } V_{max} \quad & w_i = \max(w_j, w_k) && \text{where } j \text{ and } k \text{ are children of } i \\
\text{for each } i \leq n-2 \text{ in } V_{min} \quad & w_i = \min(w_j, w_k) && \text{where } j \text{ and } k \text{ are children of } i \\
\text{for each } i \leq n-2 \text{ in } V_{avg} \quad & w_i = \frac{1}{2} w_j + \frac{1}{2} w_k && \text{where } j \text{ and } k \text{ are children of } i
\end{aligned} \tag{1.1}$$

Theorem 4 *For a stopping SSG, its value vector $\overline{v}$ is the unique solution to (1.1).*

Before proving the above theorem, let us mention that just from the definition of $\overline{v}$, each of its components need not be related, that is, $\overline{v}(1)$ and $\overline{v}(2)$ can arise out of different strategies. But Theorem 4 relates all these values. Therefore, in order to prove the above theorem, it will be convenient if we can get a link between the components of $\overline{v}$: we will show that the whole of $\overline{v}$ can be obtained using special types of strategies.

Definition 5 (Optimal strategies) We consider two notions of optimality:

- Let σ be a strategy for Max. A strategy τ for Min is said to be *optimal w.r.t.* σ if τ is an optimal strategy in the (min)-MDP G_σ: for every min vertex i, the value $v_{\sigma,\tau}(i) = \min(v_{\sigma,\tau}(j), v_{\sigma,\tau}(k))$ where j and k are its children.
- A pair of strategies (σ, τ) is said to be optimal if $v_{\sigma,\tau}$ satisfies the optimality equations.

Lemma 6 Let (σ_1, τ_1) and (σ_2, τ_2) be optimal strategies. Then $v_{\sigma_1,\tau_1} = v_{\sigma_2,\tau_2}$[1].

Proof
Denote v_{σ_1,τ_1} as v_1 and v_{σ_2,τ_2} as v_2. Let $U = \{i \in V \mid v_1(i) > v_2(i)\}$. Wlog, assume this set U is non-empty. Let $U' = \{i \in U \mid v_1(i) - v_2(i) \geq v_1(j) - v_2(j) \text{ for all vertices } j\}$. Pick $i \in U'$. If $i \in V_{avg}$, we can show that both of its children are in U'.

Let $i \in V_{max}$. Suppose $\sigma_1(i) = \sigma_2(i) = j$, then we have $j \in U'$. Suppose $\sigma_1(i) = j$ and $\sigma_2(i) = k \neq j$ such that $v_2(k) > v_2(j)$. We have: $v_1(i) = v_1(j)$, $v_2(i) = v_2(k) > v_2(j)$. So, $v_1(j) - v_2(j) = v_1(i) - v_2(j) > v_1(i) - v_2(k) = v_1(i) - v_2(i)$, contradicting that i gives a maximum difference. Therefore, if $\sigma_1(i) = j$ and $\sigma_2(i) = k$ then we necessarily have $v_2(k) = v_2(j)$. Moreover, from the above calculation, $v_1(j) - v_2(j) = v_1(i) - v_2(i)$ and hence $j \in U'$.

[1]Proof suggested by Rohan Goyal (B. Sc Math and Comp. Sci, second year)

Let $i \in V_{min}$. Similar argument as above, with a minor change. Suppose $\sigma_1(i) = \sigma_2(i) = j$, then we have $j \in U'$. Suppose $\sigma_1(i) = j$ such that $\sigma_1(j) < \sigma_1(k)$ and $\sigma_2(i) = k$. We have: $v_1(i) = v_1(j) < v_1(k)$, $v_2(i) = v_2(k)$. So, $v_1(k) - v_2(k) > v_1(j) - v_2(k) = v_1(i) - v_2(i)$. The rest of the argument proceeds as above.

uild a Markov chain G' as follows. Pick some arbitrary vertex $i \in U'$ and add it to G'. If $i \in V_{avg}$ add both its successors to G'. If $i \in V_{max}$, add the successor $j \in U'$ as in the argument above, and if $i \in V_{min}$, add the successor $k \in U'$ as above to G'. This process should terminate at some point. Notice that we are building strategies for Max and Min in this process. Since G is stopping, at some point, we should hit one of the sink vertices. This gives a contradiction as the sink vertices cannot be present in U'. ☐

The above lemma shows that optimality equations have a unique solution. We will now show that the value vector equals this unique solution.

Lemma 7 Let σ be a strategy for Max and τ a strategy for Min which is optimal w.r.t. σ. Let i be a Max vertex with children j and k, such that $v_{\sigma,\tau}(i) \neq \max(v_{\sigma,\tau}(j), v_{\sigma,\tau}(k))$.

Let σ' be the strategy obtained from σ by switching the successor at i: $\sigma'(i) = \arg\max(v_{\sigma,\tau}(j), v_{\sigma,\tau}(k))$, and $\sigma'(i') = \sigma(i')$ for other $i' \in V_{max}$. Let τ' be an optimal strategy w.r.t σ'.

Then: $v_{\sigma,\tau}(i) < v_{\sigma',\tau'}(i)$ and $v_{\sigma,\tau}(i'') \leq v_{\sigma',\tau'}(i'')$ for all other $i'' \in V$.

Proof

We have $v_{\sigma,\tau} = Q_{\sigma,\tau} v_{\sigma,\tau} + b_{\sigma,\tau}$ and $v_{\sigma',\tau'} = Q_{\sigma',\tau'} v_{\sigma',\tau'} + b_{\sigma',\tau'}$. We have:

$$\begin{aligned} v_{\sigma',\tau'} - v_{\sigma,\tau} &= Q_{\sigma',\tau'} v_{\sigma',\tau'} - Q_{\sigma,\tau} v_{\sigma,\tau} + b_{\sigma',\tau'} - b_{\sigma,\tau} \\ &= Q_{\sigma',\tau'}(v_{\sigma',\tau'} - v_{\sigma,\tau}) + (Q_{\sigma',\tau'} - Q_{\sigma,\tau}) v_{\sigma,\tau} + b_{\sigma',\tau'} - b_{\sigma,\tau} \end{aligned}$$

Denote by M the column vector $(Q_{\sigma',\tau'} - Q_{\sigma,\tau}) v_{\sigma,\tau} + b_{\sigma',\tau'} - b_{\sigma,\tau}$. From the above equations:

$$v_{\sigma',\tau'} - v_{\sigma,\tau} = (I - Q_{\sigma',\tau'})^{-1} M$$

Let us look at what $M(i'')$ is for each vertex i''. Assume j'' and k'' are the children of i''.

- When $i'' \in V_{avg}$, we have the row $Q_{\sigma',\tau'}[i'']$ to be equal to $Q_{\sigma,\tau}[i'']$ and $b_{\sigma',\tau'}(i'') = b_{\sigma,\tau}(i'')$. Hence $M(i'') = 0$.
- Let $i'' \in V_{min}$. Notice that $M(i'') = v_{\sigma,\tau}(\tau'(i'')) - v_{\sigma,\tau}(\tau(i''))$. Since τ is optimal wrt σ, we have $\tau(i'') = \min(v_{\sigma,\tau}(j''), v_{\sigma,\tau}(k''))$. This proves that $M(i'') \geq 0$.
- Let $i'' \in V_{max}$. We have $M(i'') = v_{\sigma,\tau}(\sigma'(i'')) - v_{\sigma,\tau}(\sigma(i''))$. By definition of σ', this will mean $M(i) > 0$ and $M(i'') = 0$ for all $i'' \neq i$.

We have seen earlier that the inverse $(I - Q_{\sigma',\tau'})^{-1}$ equals $I + Q_{\sigma',\tau'} + Q^2_{\sigma',\tau'} + \dots$. This shows that the product $(I - Q_{\sigma',\tau'})^{-1} M$ has all entries non-negative, and in particular the entry at i is strictly bigger than 0. This proves the required result. ☐

Proof of Theorem 4. Firstly, for every optimal strategy (σ, τ), the value vector $v_{\sigma,\tau}$ satisfies the optimality equations, just by definition. Moreover, by Lemma 6, every optimal pair

Algorithm 1.1: Strategy improvement algorithm for MDPs, also known as policy iteration

```
algorithm strategy − improvement(G)
σ ← an arbitrary positional strategy
τ ← an optimal strategy wrt σ
  v_{σ,τ} ← probabilities to reach 1−sink in G_{σ,τ}
  repeat
      pick  i ∈ V_max  s.t.  v_{σ,τ}(i) < max(v_{σ,τ}(j), v_{σ,τ}(k))  where j,k are its children
      σ'(i) ← argmax{v_{σ,τ}(j), v_{σ,τ}(k)} ,  σ'(i'') = σ(i'')  when  i'' ∈ V_max  with  i'' ≠ i
      τ' ← an optimal strategy wrt σ'
      σ ← σ'
      τ ← τ'
      v_{σ,τ} ← probabilities to reach 1−sink in G_{σ,τ}
  until (σ,τ) is optimal
```

gives the same value. It remains to show that the value vector $\overline{v}$ is obtained by an optimal pair.

For each strategy σ of max, the value $\min_\tau v_{\sigma,\tau}(i)$ is given by an optimal strategy $\tau(\sigma)$ w.r.t σ in the min-MDP G_σ (from the analogous theorem studied in the MDP case). Therefore, $\max_\sigma \min_\tau v_{\sigma,\tau}(i'') = \max_\sigma v_{\sigma,\tau(\sigma)}(i'')$ for all vertices i''. Suppose σ is a strategy that gives the maximum at some i''. If $(\sigma, \tau(\sigma))$ is not optimal, then there is a max vertex i where the optimality equation is not satisfied. We can then get a strategy σ' as in Lemma 7 such that $v_{\sigma,\tau}(i'') \leq v_{\sigma',\tau(\sigma')}(i'')$, with the inequality being strict when $i'' = i$. Therefore, we can assume that σ satisfies optimality equations. In particular, the same optimal pair $(\sigma, \tau(\sigma))$ gives the maximum at every vertex i.

Lemma 8 For every vertex i, we have $\max_\sigma \min_\tau v_{\sigma,\tau}(i) = \min_\tau \max_\sigma v_{\sigma,\tau}(i)$.

Proof
For each strategy σ, we have $\min_\tau v_{\sigma,\tau}(i) \leq \min_\tau \max_\sigma v_{\sigma,\tau}(i)$. Therefore, $\max_\sigma \min_\tau v_{\sigma,\tau}(i) \leq \min_\tau \max_\sigma v_{\sigma,\tau}(i)$. We need to show the other direction.

Suppose (σ^*, τ^*) is an optimal pair. We have $\max_\sigma \min_\tau v_{\sigma,\tau} = v_{\sigma^*,\tau^*}$ from Theorem 4. Now: $\min_\tau \max_\sigma v_{\sigma,\tau}(i) \leq \max_\sigma v_{\sigma,\tau^*}(i)$. But, $\max_\sigma v_{\sigma,\tau^*}$ will be given by an optimal strategy in the MDP G_{τ^*}. From the fact that every pair of optimal strategies gives the same value vector (Lemma 6), we can say $\max_\sigma v_{\sigma,\tau^*} = v_{\sigma^*,\tau^*}$. This proves $\max\min$ equals $\min\max$.

□

2 Algorithms

Strategy improvement. Algorithm 1.1 gives the procedure. At each iteration, we get a switched strategy of Max. From Lemma 7, this gives a strictly better value vector. This shows that no pair is repeated, and hence the process terminates. When it terminates, it necessarily gives an optimal pair.

Quadratic programming.

Value iteration.

References

[Con92] Anne Condon. The complexity of stochastic games. *Inf. Comput.*, 96(2):203–224, 1992.