Game Forms for Coalition Effectivity Functions

Colm Baston and Venanzio Capretta

School of Computer Science, University of Nottingham
{colm.baston,venanzio.capretta}@nottingham.ac.uk

Introduction Coalition logic, introduced by Pauly,[1] is a multi-agent modal logic for reasoning about what groups of agents can achieve if they act collectively, as a coalition. The semantics for coalition logic is based on *game forms*, which are essentially perfect-information strategic games where the players act simultaneously. From a game form, we can derive an *effectivity function* which defines which subsets of outcomes a particular coalition of players can guarantee, regardless of how all other players act.

Pauly proves that there is a set of properties, *playability*, that precisely describe when an arbitrary effectivity function is the effectivity function for some strategic game. Our goal is to formalise this equivalence in the logic of a type-theoretic proof assistant, specifically Coq and Agda. Proving the playability of an effectivity function that is derived from a game form is straightforward, provided that we develop good libraries for decidable subsets of agents and states. The other direction is more complex, requiring the construction of a game form from a playable effectivity function, then proving that the derived effectivity function is equivalent to the function that we started with. We believe that Pauly's proof of the second direction can be simplified in addition to adapting it for type-theoretic formalisation, and we give a sketch of this below.

Game Forms A game form G is a tuple $\langle N, \{\mathcal{A}_i\}_{i\in N}, S, o\rangle$ where: N is a finite, non-empty set of agents (for n agents, we simply use the natural numbers $\{0,\ldots,n-1\}$); $\{\mathcal{A}_i\}_{i\in N}$ is a family of non-empty sets of actions for each agent i (a *strategy profile* $\sigma : \Pi_{i\in N}\mathcal{A}_i$ is a choice of actions for every agent); S is a set of possible outcome states; o is a function $(\Pi_{i\in N}\mathcal{A}_i) \to S$ that selects an outcome for every strategy profile.

A coalition C is a decidable subset of N. Let $\sigma_C : \Pi_{i\in C}\mathcal{A}_i$ be a strategy profile for C and $\sigma_{\overline{C}} : \Pi_{i\in\overline{C}}\mathcal{A}_i$ a strategy profile for the complement coalition $\overline{C} = N \setminus C$. We denote by $\sigma_C \oplus \sigma_{\overline{C}}$ the global strategy profile which joins the actions of both coalitions.

The effectivity function for game form G is a function $E_G : \mathcal{P}_{\mathsf{dec}}(N) \to \mathcal{P}(\mathcal{P}_{\mathsf{dec}}(S))$ which associates each coalition with a set of *goals*: each goal is a decidable set of states that the coalition can achieve by working together; that is, $X \in E_G(C)$ iff there is a strategy profile for C that guarantees an outcome in X. The effectivity function for a game form G is therefore defined by:

$$E_G(C) = \{X \in \mathcal{P}_{\mathsf{dec}}(S) \mid \exists\sigma_C, \forall\sigma_{\overline{C}}, o(\sigma_C \oplus \sigma_{\overline{C}}) \in X\}$$

In the semantics of coalition logic, it is very convenient to work abstractly with an effectivity function rather than directly with the game definition. Therefore we need a characterisation of those effectivity functions that come from games.

Playable Effectivity Functions An effectivity function $E : \mathcal{P}_{\mathsf{dec}}(N) \to \mathcal{P}(\mathcal{P}_{\mathsf{dec}}(S))$ is playable iff it satisfies the following properties: For any $C \subseteq N$, $\varnothing \notin E(C)$; For any $C \subseteq N$, $S \in E(C)$; E is *N-maximal*: for any $X \subseteq S$, $\overline{X} \notin E(\varnothing) \Rightarrow X \in E(N)$; E is *outcome-monotonic*: for any $C \subseteq N$ and any $X_1 \subseteq X_2 \subseteq S$, $X_1 \in E(C) \Rightarrow X_2 \in E(C)$; E is *superadditive*: for any

[1] Marc Pauly, "A modal logic for coalitional power in games", *J. of Logic and Computation*, 12, 02 2002.

disjoint pair $C_1, C_2 \subseteq N$, and any pair $X_1, X_2 \subseteq S$, $X_1 \in E(C_1) \wedge X_2 \in E(C_2) \Rightarrow X_1 \cap X_2 \in E(C_1 \cup C_2)$.

Two more properties follow from from the above: E *is regular*: for any $C \subseteq N$ and any $X \subseteq S$, $X \in E(C) \Rightarrow \overline{X} \notin E(\overline{C})$; E is *coalition-monotonic*: for any $C_1 \subseteq C_2 \subseteq N$, $E(C_1) \subseteq E(C_2)$.

The class of playable effectivity functions consists exactly of those functions that come from some game. Proving that for a game form G, E_G is playable is just a routine question of checking the properties. The inverse requires that for every playable E we construct a game form G such that $E = E_G$. The original construction by Pauly is rather involved and would be difficult to formalise in type theory. We have found a simpler way of constructing the game.

Game Form Construction Given sets N and S and a playable effectivity function $E : \mathcal{P}_{\mathsf{dec}}(N) \to \mathcal{P}(\mathcal{P}_{\mathsf{dec}}(S))$, we construct a game form G such that $E_G = E$. The set of agents and the set of states remain unchanged from N and S respectively, so we just need to define a family of sets of actions $\{\mathcal{A}_i\}_{i \in N}$, and an outcome function o.

An action for an agent $i \in N$ consists of a choice of a coalition C that i would like to be part of, a goal X that i would like the coalition to aim for, a selected outcome $x \in X$, and a natural number t which will be used in determining which agent gets to make the final decision:

$$\mathcal{A}_i = \{\langle C, X, x, t\rangle \mid C \subseteq N, i \in C, X \in E(C), x \in X, t \in \mathbb{N}\}$$

Let a strategy profile σ be given: we have a choice $\sigma_i = \langle C_i, X_i, x_i, t_i\rangle$ for every $i \in N$. A coalition $C \subseteq N$ is called *σ-cooperative* if for every $i \in C$, $C_i = C$ and for every $i, j \in C$, $X_i = X_j$. Let $X_C = X_i$ for any $i \in C$. Intuitively, a coalition C is σ-cooperative if all its members want to be in the coalition, and they agree on the goal X_C they want to aim for.

Let $\langle C_1, \ldots, C_m\rangle$ be all the σ-cooperative coalitions and let C_0 be the set of agents that are not in a σ-cooperative coalition. $\langle C_0, \ldots, C_m\rangle$ is a partition of N. For every $k = 1 \ldots m$, we already defined $X_{C_k} = X_i$ for any $i \in C$. Define $X_{C_0} = S$ and

$$O(\sigma) = \bigcap_{k=0}^{m} X_{C_k} = \bigcap_{k=1}^{m} X_{C_k}$$

The outcome of the game will be defined to be a state in $O(\sigma)$. The choice of the specific state will depend again on σ. We use the numbers t_i to determine an agent that will make the final decision: let $d = (\sum_{i \in N} t_i) \mod |N|$. The outcome will be the state chosen by this agent, x_d. However, this is not guaranteed to be an element of $O(\sigma)$: it is an element of X_d which is a superset of $O(\sigma)$. In case it isn't we revert to an arbitrary choice function $H : \Pi_{X \in E(N)} X$. This exists constructively because by definition of playable effectivity function every $X \in E(N)$ is non-empty.

We can prove that $O(\sigma) \in E(N)$, so we can define:

$$o(\sigma) = \begin{cases} x_d & \text{if } x_d \in O(\sigma) \\ H(O(\sigma)) & \text{otherwise} \end{cases}$$

Theorem. $E_G = E$