EURECOM
Department of Mobile Communications
2229, route des Crêtes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-11-248

Cooperative Markov Decision Processes: Time Consistency, Greedy Players Satisfaction, and Cooperation Maintenance

January 5th, 2012

Prof. Konstantin AVRACHENKOV
Prof. Laura COTTATELLUCCI
Lorenzo MAGGI

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : k.avrachenkov@sophia.inria.fr, {laura.cottatellucci, lorenzo.maggi}@eurecom.fr,

[1]EURECOM's research is partially supported by its industrial members: BMW Group, Cisco, Monaco Telecom, Orange, SAP, SFR, Sharp, STEricsson, Swisscom, Symantec, Thales.

Cooperative Markov Decision Processes: Time Consistency, Greedy Players Satisfaction, and Cooperation Maintenance

Konstantin Avrachenkov, Laura Cottatellucci, Lorenzo Maggi

Abstract

We deal with multi-agent Markov Decision Processes (MDPs) in which cooperation among players is allowed. We find a cooperative payoff distribution procedure (MDP-CPDP) that distributes in the course of the game the payoff that players would get in the long run static game. We show under which conditions such a MDP-CPDP fulfills a time consistency property, contents greedy players, and strengthen the coalition cohesiveness throughout the game.

Index Terms

Cooperative Markov decision processes, stochastic games, payoff distribution procedure, time consistency, greedy players, cooperation maintenance.

Contents

1 Introduction

Repeated cooperative games constitute one of the most recent and interesting topics in game theory. They attempt to model real situations in which the *same* game is repeated over time and players can cooperate and form coalitions throughout the duration of the game. The papers by Oviedo (2000) and by Kranich, Perea, and Peters (2001) are the two independent pioneering works in this field.

While the theory of competitive Markov decision processes (MDPs), otherwisely called non-cooperative stochastic games, has been thoroughly studied (Filar and Vrieze 1996 for an extensive survey), to the best of the authors' knowledge, there is very little work in the literature on cooperative MDPs. Unlike classic repeated games, there are several *different* stage games that follow one another according to a discrete-time Markov chain, whose transition probabilities depend on the players' actions in each stage game. Players can decide whether to join the grand coalition or, throughout the game, forming coalitions. The payoff gained by a coalition is, under the transferable utility (TU) assumption, shared among its participants. Once a group of players has withdrawn from the grand coalition, it cannot rejoin it later on. Petrosjan (2002), in his pioneering work, proposed a cooperative payoff distribution procedure (CPDP) in cooperative games on finite trees.

In this paper we deal with discount cooperative MDPs, in which the payoffs at each stage are multiplied by a discount factor and summed up over time. Our game model is in fact more general than the one by Petrosjan (2002), since we allow for cycles on the state space and we do not impose the finiteness of the game horizon. We also point out that our model is different from the one proposed by Predtetchinski (2007), since we assume that the utility of the coalitions is transferable and the probability transitions among the single stage games does depend on the players' actions in each stage.

In static cooperative game theory (e.g. Peleg and Sudhölter 2007), in which only one stage game is played, the main challenge is to find a payoff sharing procedure among all players such that it is both optimum for the whole community of players and it does not prompt any subset of players to withdraw from the grand coalition. On the contrary, in our framework of cooperative MDPs, since the horizon of the game is not even finite, then it is legitimate to suppose that all players demand to be rewarded at each stage, and not at the end of the whole game. Therefore, the situation is more tricky than in the classic static setting, because we need to find a stage-wise payoff distribution such that all the players are content with it at *each* stage of the game.

The paper is organized as follows. Section 2 is a short survey on non-cooperative and cooperative multi-agent MDPs. Following the lines of Petrosjan's work, in Section 3 we propose a stationary stage-wise CPDP for cooperative discounted MDPs (MDP-CPDP). In Section 4 we prove that our MDP-CPDP satisfies what

we call the "terminal fairness property", i.e. the expected discounted sum of payoff allocations belongs to a cooperative solution (i.e. Shapley Value, Core, etc.) of the whole discounted game. In Section 5 we show that our MDP-CPDP fulfills the time consistency property, which is a crucial one in repeated games theory (e.g. Filar and Petrosjan 2000): it suggests that a CPDP should respect the terminal fairness property in a subgame starting from any time step. In Section 6 we show that, under some conditions, for all discount factors small enough, also the greedy players having a myopic perspective of the game are satisfied with our MDP-CPDP. Section 7 deals with a special case of our model, entaling that the transition probabilities among the states do not depend on the players' strategies. In Section 8 we deal perhaps with the most meaningful attribute for a CPDP, which is the n-tuple step cooperation maintenance property. It claims that, at each stage of the game, the long run reward that each group of players expects to get by withdrawing from the grand coalition after n step should be less than what it would get by sticking to the grand coalition forever. In some sense, if such a condition is fulfilled for all integers n's, then no players are enticed to withdraw from the grand coalition. We find that the single step cooperation maintenance property, earliest introduced in a deterministic setting by Mazalov and Rettieva (2010), is the strongest one among all n's. Furthermore, we give a necessary and sufficient condition, inspired by the celebrated Bondareva-Shapley Theorem (Bondareva 1963; Shapley 1967), for our MDP-CPDP to satisfy the n-tuple step cooperation maintenance property, for all integers n.

Some notation remarks. The ordering relations $<, >$, if referred to vectors, are component-wise, as well as the $\max$ and $\min$ operators. The entry that lies in the i-th row and in the j-th column of matrix $\mathbf{A}$ is written as $\mathbf{A}_{i,j}$. An equivalent notation for the n-by-m matrix $\mathbf{A}$ is $[\mathbf{A}_{i,j}]_{i=1,j=1}^{n,m}$. The i-th element of column vector $\mathbf{a}$ is denoted by $\mathbf{a}_i$. The expression $\mathrm{val}(\mathbf{A})$ stands for the value (e.g. Filar and Vrieze 1996) of the matrix $\mathbf{A}$. Let $\{C_i\}_i$ be a collection of sets; we define the set $\sum_i C_i$ as $\{\sum_i c_i : \ c_i \in C_i, \ \forall\, i\}$.

2 Discounted Cooperative Markov Decision Processes

In a multi-agent Markov Decision Process (MDP) Γ with $P > 1$ players there is a finite set of states $S = \{s_1, s_2, \ldots, s_N\}$, and for each state s the set of actions available to the i-th player is denoted by $A_i(s)$, $i = 1, \ldots, P$, and $|A_i(s)| = m_i(s)$. To each $(P+1)$-tuple $(s, a_1, \ldots, a_P)$, with $a_i \in A_i(s)$, an immediate reward $r_i(s, a_1, \ldots, a_P)$ for player $i = 1, \ldots, P$ and a transition probability distribution $p(.|s, a_1, \ldots, a_P)$ on the state space S are assigned.

Let $\mathcal{C} = \{1, \ldots, P\}$ be the grand coalition. We assume that any subset of players $\Lambda \subseteq \mathcal{C}$ can withdraw from the grand coalition and form a coalition at any time stage of the game, and all the players are compelled to play throughout the

whole duration of the game. Moreover, once a coalition is formed, it can no longer rejoin the grand coalition in the future.

Let $A_\Lambda(s) = \prod_{i\in\Lambda} A_i(s)$ be the set of actions available to coalition Λ in state s, for all $s \in S$. A stationary strategy $\mathbf{f}_\Lambda$ for the coalition Λ determines the probability $\mathbf{f}_\Lambda(a|s)$ that in state s the coalition Λ chooses the action $a \in A_\Lambda(s)$. We define with $\mathbf{F}_\Lambda$ the set of stationary strategies for coalition $\Lambda \subseteq \mathcal{C}$. If for every $s \in S$ there exists $a(s)$ such that $\mathbf{f}_\Lambda(a(s)|s) = 1$, then the stationary strategy $\mathbf{f}_\Lambda$ is called pure (or deterministic).
Let us define the transition probability distribution on the state space S, given the independent strategies $\mathbf{f}_\Lambda \in \mathbf{F}_\Lambda$, $\mathbf{f}_{\mathcal{C}\setminus\Lambda} \in \mathbf{F}_{\mathcal{C}\setminus\Lambda}$, as

$$p(s'|s, \mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}) = \sum_{a_\Lambda\in A_\Lambda(s)} \sum_{a_{\mathcal{C}\setminus\Lambda}\in A_{\mathcal{C}\setminus\Lambda}(s)} p(s'|s, a_\Lambda, a_{\mathcal{C}\setminus\Lambda})\, \mathbf{f}_\Lambda(a_\Lambda|s)\, \mathbf{f}_{\mathcal{C}\setminus\Lambda}(a_{\mathcal{C}\setminus\Lambda}|s),$$

for all $s, s' \in S$. Analogously, let $r_i(s, \mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$ be the expected instantaneous reward for player i in state s.

Let $\beta \in [0; 1)$ be the discount factor and let

$$r_\Lambda(s, \mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}) = \sum_{i\in\Lambda} r_i(s, \mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$$

be the instantaneous reward gained by the coalition Λ in state s. We define $\mathbf{\Phi}_\Lambda^{(\beta)}(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$ as the N-by-1 vector whose k-th component equals the expected β-discounted long run reward for coalition $\Lambda \subseteq \mathcal{C}$, when the initial state of the game is s_k, i.e.

$$\mathbf{\Phi}_\Lambda^{(\beta)}(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}) = \sum_{t=0}^{\infty} \beta^t \mathbf{P}^t(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})\, \mathbf{r}_\Lambda(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}), \tag{1}$$

where $\mathbf{P}(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$ is the *N*-by-*N* transition probability matrix and $\mathbf{r}_\Lambda(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$ is a *N*-by-1 vector, whose k-th component is $r_\Lambda(s_k, \mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})$.

Let Γ_s be the game Γ starting in state $s \in S$. For any $\beta \in [0; 1)$ and for every state s, we assign to each coalition Λ a real utility $v^{(\beta)}(\Lambda, \Gamma_s)$. Under the transferable utility (TU) condition, the coalition values can be shared in any manner among the members of the coalition. Hence, the set of feasible allocations for coalition $\Lambda \subseteq \mathcal{C}$ in the game Γ_s is $\mathcal{V}^{(\beta)}(\Lambda, \Gamma_s)$, where

$$\mathcal{V}^{(\beta)}(\Lambda, \Gamma_s) = \left\{ \mathbf{x} \in \mathbb{R}^P : \sum_{i\in\Lambda} \mathbf{x}_i \le v^{(\beta)}(\Lambda, \Gamma_s) \right\}.$$

It is widely accepted to assign to the empty coalition the null utility, i.e.

$$v^{(\beta)}(\{\emptyset\}, \Gamma_s) = 0.$$

We consider the value associated to the grand coalition $v^{(\beta)}(\mathcal{C}, \Gamma_s)$ to be the biggest achievable discounted sum of reward in the game Γ_s:

$$\begin{aligned} v^{(\beta)}(\mathcal{C}, \Gamma_s) &= \Phi_{\Lambda}^{(\beta)}(s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) \\ \mathbf{f}_{\mathcal{C}}^{(\beta)*} &= \underset{\mathbf{f}_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}}{\operatorname{argmax}} \mathbf{\Phi}_{\mathcal{C}}^{(\beta)}(\mathbf{f}_{\mathcal{C}}), \qquad \forall\, \beta \in [0; 1) \end{aligned} \tag{2}$$

where $\mathbf{f}_{\mathcal{C}}^{(\beta)*}$ the global optimum strategy for the grand coalition, for all $\Gamma_s,\ s \in S$. In most applications it makes sense to define the coalition value $v^{(\beta)}(\Lambda, \Gamma_s)$ as the maximum total reward that coalition Λ can ensure for itself in the β-discounted long run game Γ_s (von Neumann and Morgenstern 1944), i.e.

$$\begin{aligned} v^{(\beta)}(\Lambda, \Gamma_s) &= \max_{\mathbf{f}_{\Lambda} \in \mathbf{F}_{\Lambda}} \min_{\mathbf{f}_{\mathcal{C}\setminus\Lambda} \in \mathbf{F}_{\mathcal{C}\setminus\Lambda}} \Phi_{\Lambda}^{(\beta)}(s, \mathbf{f}_{\Lambda}, \mathbf{f}_{\mathcal{C}\setminus\Lambda}) \\ &= \min_{\mathbf{f}_{\mathcal{C}\setminus\Lambda} \in \mathbf{F}_{\mathcal{C}\setminus\Lambda}} \max_{\mathbf{f}_{\Lambda} \in \mathbf{F}_{\Lambda}} \Phi_{\Lambda}^{(\beta)}(s, \mathbf{f}_{\Lambda}, \mathbf{f}_{\mathcal{C}\setminus\Lambda}), \qquad \forall\, \Lambda \subseteq \mathcal{C}/\{\emptyset\}. \end{aligned} \tag{3}$$

Throughout the paper, if not specified, we always consider nonempty coalitions. We now provide some useful definitions and results.

Definition 1 (Linear combination of games)**.** *Let $\mathcal{V}(\Delta_i, \Lambda)$ be the set of feasible allocations for the coalition $\Lambda \subseteq \mathcal{C}$ in the game Δ_i, for $i = 1, \ldots, N$. The linear combination $\sum_i b_i \Delta_i$ is a game in which the set of feasible allocations for the coalition Λ is the Minkowski sum $\mathcal{V}(\sum_i b_i \Delta_i, \Lambda) \equiv \sum_i b_i \mathcal{V}(\Delta_i, \Lambda)$.*

Proposition 1. *Let $\Delta_1, \ldots, \Delta_N$ be N games with transferable utilities. Let $v(\Lambda, \Delta_i)$ be the value of coalition $\Lambda \subseteq \mathcal{C}$ in the game Δ_i. Let $b_1, \ldots, b_N$ be non negative coefficients. Then, $\sum_i b_i \Delta_i$ is a TU game such that the value of the coalition $\Lambda \subseteq \mathcal{C}$ is*

$$v\left(\Lambda, \sum_{i=1}^{N} b_i \Delta_i\right) = \sum_i b_i v(\Lambda, \Delta_i).$$

Proof. Let

$$\widetilde{\mathcal{V}}(\Lambda) = \left\{ \mathbf{x} \in \mathbb{R}^P : \sum_{i:\{i\} \in \Lambda} \mathbf{x}_i \leq \sum_i b_i v(\Lambda, \Delta_i) \right\}.$$

We have to prove that, for all $\Lambda \subseteq \mathcal{C}$, $\mathcal{V}(\sum_i b_i \Delta_i, \Lambda) \equiv \sum_i b_i \mathcal{V}(\Delta_i, \Lambda) = \widetilde{\mathcal{V}}(\Lambda)$. Let the real P-tuple $\mathbf{c}(i) \in \mathcal{V}(\Delta_i, \Lambda)$, for all i. It is straightforward to see that $\sum_i b_i \mathbf{c}(i) \in \widetilde{\mathcal{V}}(\Lambda)$. Then, $\sum_i b_i \mathcal{V}(\Delta_i, \Lambda) \subseteq \widetilde{\mathcal{V}}(\Lambda)$. Let us fix the real P-tuple $\widetilde{\mathbf{c}} \in \widetilde{\mathcal{V}}(\Lambda)$. We define $I = \{i : b_i > 0\}$. We need to find $\{\mathbf{c}'(i) \in \mathcal{V}(\Delta_i, \Lambda)\}_{i \in I}$ such that $\sum_{i \in I} b_i \mathbf{c}'(i) = \widetilde{\mathbf{c}}$. Let $\mathbf{c}'_j(i) = \widetilde{\mathbf{c}}_j / (|I| b_i)$ for all j such that $\{j\} \notin \Lambda$. To determine the remaining $|I||\Lambda|$ elements $\{\mathbf{c}'_j(i),\ \forall\, i \in I,\ j : \{j\} \in \Lambda\}$, we introduce the following set of inequalities:

$$\begin{cases} \sum_{i \in I} b_i \mathbf{c}'_j(i) = \widetilde{\mathbf{c}}_j & \forall\, j : \{j\} \in \Lambda \\ \sum_{j:\{j\} \in \Lambda} \mathbf{c}'_j(i) \leq v(\Lambda, \Delta_i) & \forall\, i \in I \end{cases} \tag{4}$$

Let us prove that (4) admits a solution. Let $\epsilon_i \geq 0$, for all $i \in I$, be such that

$$\sum_{i\in I} \epsilon_i = \sum_{i\in I} b_i v(\Lambda, \Delta_i) - \sum_{j:\{j\}\in\Lambda} \widetilde{\mathbf{c}}_j \geq 0 \tag{5}$$

We write the following linear system

$$\begin{cases} \sum_{i\in I} b_i \mathbf{c}'_j(i) = \widetilde{\mathbf{c}}_j & \forall\, j:\{j\}\in\Lambda \\ b_i \sum_{j:\{j\}\in\Lambda} \mathbf{c}'_j(i) = b_i v(\Lambda,\Delta_i) - \epsilon_i & \forall\, i \in I \end{cases} \tag{6}$$

Evidently, any solution to (6) is also a solution to (4). Thanks to (5), the sum of the first $|\Lambda|$ equations of (6) equals the sum of the remaining $|I|$ equations. By discarding the last equation of (6) we get a linear system with $|\Lambda|+|I|-1$ linearly independent equations in $|\Lambda||I| > |\Lambda|+|I|-1$ unknowns. Hence, a solution to (6) exists and $\sum_i b_i \mathcal{V}(\Delta_i,\Lambda) \supseteq \widetilde{V}(\Lambda)$. Then, $\sum_i b_i \mathcal{V}(\Delta_i,\Lambda) = \widetilde{\mathcal{V}}(\Lambda)$ and the thesis is proved. □

Definition 2 (Terminal cooperative solution)**.** *Set $\beta \in [0;1)$. The terminal cooperative solution $\mathbf{T}^{(\beta)}(\Gamma_s)$ is a set-valued function which represents a static cooperative solution (e.g. Shapley value, Core, etc.) of the whole game starting in state s, i.e.*

$$\mathbf{T}^{(\beta)}(\Gamma_s) \equiv \mathbf{T}^{(\beta)}\left(\Gamma_s, \{v^{(\beta)}(\Lambda,\Gamma_s)\}_{\Lambda\subseteq\mathcal{C}}\right) : \mathbb{R}^{2^P-1} \to \mathbb{R}^P, \qquad \forall\, s \in S.$$

Analogously, we define $\mathbf{T}^{(\beta)}(\sum_i b_i \Gamma_{s_i})$ as the terminal cooperative solution of the cooperative game with coalition values $\{v^{(\beta)}(\Lambda, \sum_i b_i \Gamma_{s_i})\}_{\Lambda\subseteq\mathcal{C}}$.

The terminal cooperative solution $\mathbf{T}^{(\beta)}$ can represent any of the classical cooperative solutions. For example, $\mathbf{T} \equiv \mathbf{Co}$ represents the Core of the β-discounted game Γ_s, that is the set, possibly empty, of the real P-tuples $\mathbf{x}$ satisfying

$$\begin{cases} \sum_{i\in\mathcal{C}} \mathbf{x}_i = v^{(\beta)}(\mathcal{C},\Gamma_s) \\ \sum_{i\in\Lambda} \mathbf{x}_i \geq v^{(\beta)}(\Lambda,\Gamma_s),\ \forall\, \Lambda \subset \mathcal{C}. \end{cases} \tag{7}$$

The strict Core $\mathbf{sCo}^{(\beta)}(\Gamma_s)$ is defined in (7), but with the strict inequality signs. The terminal cooperative solution $\mathbf{T} \equiv \mathbf{Sh}^{(\beta)}(\Gamma_s)$ stands for the Shapley value of the β-discounted game Γ_s, i.e. for all $i = 1,\ldots,P$,

$$\mathbf{Sh}_i^{(\beta)}(\Gamma_s) = \sum_{\Lambda\subseteq\mathcal{C}/\{i\}} \frac{|\Lambda|!\,(P-|\Lambda|-1)!}{P!} \left[v^{(\beta)}(\Lambda\cup\{i\},\Gamma_s) - v^{(\beta)}(\Lambda,\Gamma_s)\right].$$

We now state the following results, used in the following sections.

Proposition 2. *Let $\Delta_1,\ldots,\Delta_N$ be games with transferable utilities with non empty Cores $\mathbf{Co}(\Delta_1),\ldots,\mathbf{Co}(\Delta_N)$, respectively. Let $b_1,\ldots,b_N$ be non negative coefficients. Then, $\sum_{i=1}^N b_i \mathbf{Co}(\Delta_i) \subseteq \mathbf{Co}(\sum_{i=1}^N b_i \Delta_i)$.*

Proof. Let $\mathbf{x}_1(i), \ldots, \mathbf{x}_P(i)$ be an allocation belonging to the Core $\mathbf{Co}(\Delta_i)$. Thanks to the linearity property of coalition values shown in Proposition 1, we can write

$$\sum_{i=1}^{N}\sum_{k\in\mathcal{C}} b_i\mathbf{x}_k(i) = \sum_{i=1}^{N} b_i v(\mathcal{C},\Delta_i) = v\left(\mathcal{C},\sum_{i=1}^{N} b_i\Delta_i\right)$$
$$\sum_{i=1}^{N}\sum_{k\in\Lambda} b_i\mathbf{x}_k(i) \geq \sum_{i=1}^{N} b_i v(\Lambda,\Delta_i) = v\left(\Lambda,\sum_{i=1}^{N} b_i\Delta_i\right), \qquad \forall\,\Lambda\subset\mathcal{C}.$$

Then, any point belonging to $\sum_{i=1}^{N} b_i\mathbf{Co}(\Delta_i)$ is also in $\mathbf{Co}(\sum_{i=1}^{N} b_i\Delta_i)$. Hence, the thesis is proved. □

Proposition 3. *For all* $\beta \in [0;1)$, $\sum_{i=1}^{N} b_i\mathbf{Sh}^{(\beta)}(\Gamma_{s_i}) = \mathbf{Sh}^{(\beta)}(\sum_{i=1}^{N} b_i\Gamma_{s_i})$, *where* $b_i \geq 0,\ \forall\, i$.

Proof. The proof follows straightforward from Proposition 1 and from the linearity property of the Shapley value. □

3 Cooperative Payoff Distribution Procedure

In cooperative MDPs, different stage games follow one another in time; the game may have an infinite length, or the players may not know when the game reaches the end. This is the case of *transient* games, for which

$$\sum_{t=0}^{\infty}\sum_{s'\in S} p_t(s'|s,\mathbf{f}_{\mathcal{C}}) < \infty, \quad \forall\, s\in S,\ \mathbf{f}_{\mathcal{C}}\in\mathbf{F}_{\mathcal{C}}. \tag{8}$$

where $p_t(s'|s) = p(S_t = s'|S_0 = s)$ is the probability of being in state s' at the t-th step, knowing that the starting state was s. Therefore, it is reasonable to assume that all the players demand to be rewarded at each stage of the game, and not only at its conclusion. With respect to static cooperative game theory, an additional complication lies in satisfying all the players at each time stage of the game, since coalitions are allowed to form throughout the game unfolding.
According to classic cooperative game theory, player i gets the terminal cooperative solution $\mathbf{T}_i^{(\beta)}(\Gamma_s)$ at the end of the β-discounted game Γ_s. The *goal* here is to find a way to stage-wisely share among the participants the value of the grand coalition.

Remark: All the results presented in the current section, as well as the ones in Sections 4, 5, 8, can be easily extended to undiscounted transient MDPs, i.e. games for which equation (8) holds and $\beta = 1$. Note in fact that, mathematically, introducing a discount factor $\beta \in [0;1)$ is equivalent to multiplying each transition probability by β, which automatically ensures the transient condition (8).

In his pioneering work, Petrosjan (2002) introduced a cooperative payoff distribution procedure (CPDP) for games on finite trees. Following his lines, in this section we propose a CPDP for cooperative MDPs with β-discounted criterion, with $\beta \in [0;1)$ fixed *a priori*.

Definition 3 (CPDP)**.** *The cooperative payoff distribution procedure (CPDP)* $\mathbf{g}^{(\beta)} = [\mathbf{g}_1^{(\beta)}, \ldots, \mathbf{g}_P^{(\beta)}]$ *is a recursive function that, for each time step* $t \geq 0$*, associates a real* P*-tuple* $\mathbf{g}^{(\beta)}(\mathbf{h}_t)$ *to the past history* $\mathbf{h}_t = [S_0, \mathbf{g}^{(\beta)}(\mathbf{h}_0), S_1, \ldots, \mathbf{g}^{(\beta)}(\mathbf{h}_{t-1}), S_t]$ *of states succession and stage-wise allocations up to time* t*.*

The following are two alternative interpretations for $\mathbf{g}_i^{(\beta)}$:

i) $\beta^t \mathbf{g}_i^{(\beta)}(\mathbf{h}_t)$ is the payoff that player $i \in \mathcal{C}$ gets at the stage t of the game, when $\mathbf{h}_t$ is the history of the process;

ii) $\mathbf{g}_i^{(\beta)}(\mathbf{h}_t)$ is the payoff that player i gets at time t when the new transition probabilities p' are reduced by a factor β, i.e. $p'(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) = \beta p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})$. Hence, $1-\beta$ is the stopping probability in each state.

Let us now define stationary CPDPs.

Definition 4 (Stationarity)**.** *Set* $\beta \in [0;1)$*. A CPDP* $\mathbf{g}^{(\beta)}$ *is stationary iff* $\mathbf{g}^{(\beta)}(\mathbf{h}_t) = \mathbf{g}^{(\beta)}(S_t = s) = \mathbf{g}^{(\beta)}(s)$*, for all* $t \geq 0$ *and* $\mathbf{h}_t$*.*

Hence, a stationary CPDP $\mathbf{g}^{(\beta)} : S \to \mathbb{R}^P$ is a stage-wise payoff distribution law that does not depend on the whole history of the process up to time t, but only on the state at time t.

We finally propose a CPDP for cooperative MDPs (MDP-CPDP).

Definition 5 (MDP-CPDP)**.** *Set* $\beta \in [0;1)$*. Select the real* P*-tuple* $\overline{\mathbf{T}}^{(\beta)}(\Gamma_s) \in \mathbf{T}^{(\beta)}(\Gamma_s)$*,* $\forall\, s \in S$*. Our MDPs cooperative payoff distribution procedure (MDP-CPDP) is the function* $\boldsymbol{\gamma}^{(\beta)}(s)$ *between the Euclidean spaces* $\mathbb{R} \to \mathbb{R}^N$ *defined by*

$$\boldsymbol{\gamma}^{(\beta)}(s) = \sum_{s' \in S} \left[\delta_{s,s'} - \beta\, p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\right] \overline{\mathbf{T}}^{(\beta)}(\Gamma_{s'}), \qquad \forall\, s \in S. \tag{9}$$

In the following sections we will illustrate some appealing properties of such a CPDP.

4 Terminal Fairness

In this section, we let the terminal cooperative solution $\mathbf{T}$ be any of the classic cooperative solution (Core, Shapley value, Nucleolus, etc.). We now propose two desirable properties for a CPDP and we prove that the MDP-CPDP defined in (9) fulfills both of them.

The first fundamental feasibility property of a stationary CPDP consists in sharing among the players the total payoff attained by the grand coalition at each stage of the game. In order to ensure always such a property, we also require that the instantaneous rewards are deterministic.

Property 1 (Stage-wise efficiency)**.** *Set $\beta \in [0;1)$. The CPDP $\mathbf{g}^{(\beta)}$ is stage-wise efficient iff $\sum_{i\in\mathcal{C}} \mathbf{g}_i^{(\beta)}(s) = \sum_{i\in\mathcal{C}} r_i(s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})$ for all $s \in S$, where $\mathbf{f}_{\mathcal{C}}^{(\beta)*}$ is a pure stationary strategy.*

Theorem 1. *The MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$, defined in (9), fulfills the stage-wise efficiency Property 1, for all $\beta \in [0;1)$.*

Proof. The global optimum strategy $\mathbf{f}_{\mathcal{C}}^{(\beta)*}$ is pure, since the optimization problem (2) that it solves can be formulated as a Markov Decision Process (Puterman 1994). Hence, $r_i(s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})$ is deterministic as a function of s, for all $i \in \mathcal{C}$.
Let us sum (9) over all possible $i \in \mathcal{C}$, for all $s \in S$:

$$v^{(\beta)}(\mathcal{C}, \Gamma_s) = \sum_{i\in\mathcal{C}} \boldsymbol{\gamma}_i^{(\beta)}(s) + \beta \sum_{s'\in S} p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) v^{(\beta)}(\mathcal{C}, \Gamma_{s'}).$$

Since the following is also valid for all $s \in S$ from the definition of $v^{(\beta)}$:

$$v^{(\beta)}(\mathcal{C}, \Gamma_s) = \sum_{i\in\mathcal{C}} \mathbf{r}_i(s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) + \beta \sum_{s'\in S} p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) v^{\beta}(\mathcal{C}, \Gamma_{s'}),$$

then, $\sum_{i\in\mathcal{C}} \boldsymbol{\gamma}_i^{(\beta)}(s) = \sum_{i\in\mathcal{C}} \mathbf{r}_i(s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})$, surely. □

In order to guarantee a continuity between static cooperative game theory and dynamic payoff allocation, we require the expected discounted sum of the stage-wise allocations to be equal to the terminal cooperative solution of the game.

Property 2 (Terminal fairness)**.** *Set $\beta \in [0;1)$. The CPDP $\mathbf{g}^{(\beta)}$ is said to be terminal fair iff the terminal cooperative solution is stage-wisely distributed in the course of the game, i.e. $E\Big[\sum_{t\geq 0} \beta^t \mathbf{g}^{(\beta)}(\mathbf{h}_t)|S_0 = s\Big] \in \mathbf{T}^{(\beta)}(\Gamma_s)$, for all $s \in S$.*

Theorem 2. *The MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}(s) \in \mathbb{R}^P$, defined in (9) is the unique stationary CPDP that satisfies the terminal fairness Property 2, for all $\beta \in [0;1)$.*

Proof. We know from Filar and Vrieze (1996) that, for all $i \in \mathcal{C}$,

$$\begin{bmatrix} E[\sum_{t\geq 0} \beta^t \boldsymbol{\gamma}_i^{(\beta)}(S_t)|S_0 = s_1] \\ \vdots \\ E[\sum_{t\geq 0} \beta^t \boldsymbol{\gamma}_i^{(\beta)}(S_t)|S_0 = s_N] \end{bmatrix} = \sum_{t\geq 0} \beta^t \mathbf{P}^t(\mathbf{f}_{\mathcal{C}}^{(\beta)*}) \begin{bmatrix} \boldsymbol{\gamma}_i^{(\beta)}(s_1) \\ \vdots \\ \boldsymbol{\gamma}_i^{(\beta)}(s_N) \end{bmatrix}.$$

If we substitute (9) in the equation above, we find that $\boldsymbol{\gamma}_i^{(\beta)}$ defined in (9) satisfies the relation:

$$E\Big[\sum_{t\geq 0} \beta^t \boldsymbol{\gamma}^{(\beta)}(S_t)|S_0 = s\Big] = \overline{\mathbf{T}}^{(\beta)}(\Gamma_s), \;\; \forall\, s \in S,\; i \in \mathcal{C}.$$

Since the matrix $\sum_{t\geq 0}\beta^t\mathbf{P}^t(\mathbf{f}_{\mathcal{C}}^{(\beta)*}) = (\mathbf{I}-\beta\mathbf{P}(\mathbf{f}_{\mathcal{C}}^{(\beta)*}))^{-1}$ is invertible, then such $\boldsymbol{\gamma}^{(\beta)}$ is also unique. □

It is straightforward to verify that the MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ defined in (9) also fulfills a *terminal efficiency* property, i.e.

$$\sum_{i\in\mathcal{C}} E\Big[\sum_{t\geq 0}\beta^t\gamma_i^{(\beta)}(S_t|S_0=s)\Big] = v^{(\beta)}(\mathcal{C},\Gamma_s), \qquad \forall\, s\in S.$$

5 Time Consistency

Time consistency is a well known concept in dynamic cooperative theory (Filar and Petrosjan 2000 and references therein). It captures the idea that the stage-wise allocation must respect the terminal fairness Property 2 even from a later starting time of the game, for any possible trajectory of the game up to that time. In other words, if players renegotiate the agreement on CPDP at any intermediate time step, assuming that cooperation has prevailed from initial date until that instant, then the payoff distribution procedure would remain the same. This property can be formalized as follows.

Property 3 (Time consistency)**.** *Set $\beta\in[0;1)$. The CPDP $\mathbf{g}^{(\beta)}$ in (9) is said to be time consistent iff, for all $n\geq 1$ and for all possible allocation/state histories $\mathbf{h}_{n-1}$ up to time $n-1$,*

$$E\left[\sum_{t=n}^{\infty}\beta^t\mathbf{g}^{(\beta)}(S_t,\mathbf{h}_{t-1})\Big|\mathbf{h}_{n-1}\right] \in \beta^n\mathbf{T}^{(\beta)}\left(\sum_{s'\in S}p(s'|S_{n-1}=\overline{s},\mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right), \tag{10}$$

where $\overline{s}$ is the latest state of history $\mathbf{h}_{n-1}$.

Now we are ready to state the main result of this section.

Theorem 3. *The stationary MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies the time consistency Property 3 for all $\beta\in[0;1)$, where $\mathbf{T}$ represents the Shapley Value, or the Core if we suppose that $\mathbf{Co}^{(\beta)}(\Gamma_s)$ is nonempty for any $s\in S$.*

Proof. Since $\boldsymbol{\gamma}^{(\beta)}$ is stationary, we can rewrite (10) as

$$E\left[\sum_{t=0}^{\infty}\beta^t\boldsymbol{\gamma}^{(\beta)}(S_{t+n})\Big|S_{n-1}=\overline{s}\right] \in \mathbf{T}^{(\beta)}\left(\sum_{s'\in S}p(s'|\overline{s},\mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right). \tag{11}$$

Let us rewrite now equation (9), for all $s\in S$, as

$$\overline{\mathbf{T}}^{(\beta)}(\Gamma_s) = \boldsymbol{\gamma}^{(\beta)}(s) + \beta\sum_{s'\in S}p(s'|s,\mathbf{f}_{\mathcal{C}}^{(\beta)*})\overline{\mathbf{T}}^{(\beta)}(\Gamma_{s'}), \tag{12}$$

where $\boldsymbol{\gamma}(s) = [\gamma_1(s), \ldots, \gamma_P(s)]^T$ and $\overline{\mathbf{T}}^{(\beta)}(\Gamma_s) \in \mathbf{T}^{(\beta)}(\Gamma_s)$. Thanks to (12), we can write

$$E\left[\sum_{t=0}^{\infty} \beta^t \boldsymbol{\gamma}^{(\beta)}(S_{t+n}) \Big| S_{n-1} = \overline{s}\right] = \sum_{s' \in S} p(s'|\overline{s}, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) \overline{\mathbf{T}}^{(\beta)}(\Gamma_{s'}).$$

It is implicit that any player, after being rewarded with $\boldsymbol{\gamma}^{(\beta)}(\overline{s})$ in state $\overline{s}$ at step $n-1$, can withdraw from the grand coalition only in the following time step n. Then, also the transition probabilities from state $\overline{s}$ are invariant with respect to a change of strategy. Therefore, we can exploit Proposition 2 to claim that, if $\mathbf{T} \equiv \mathbf{Co}$, then

$$E\left[\sum_{t=0}^{\infty} \beta^t \boldsymbol{\gamma}^{(\beta)}(S_{t+n}) \Big| S_{n-1} = \overline{s}\right] \in \mathbf{Co}^{(\beta)}\left(\sum_{s' \in S} p(s'|\overline{s}, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) \Gamma_{s'}\right).$$

Thanks to Proposition 3 we can state that, if $\mathbf{T} \equiv \mathbf{Sh}$, then

$$E\left[\sum_{t=0}^{\infty} \beta^t \boldsymbol{\gamma}^{(\beta)}(S_{t+n}) \Big| S_{n-1} = \overline{s}\right] = \mathbf{Sh}^{(\beta)}\left(\sum_{s' \in S} p(s'|\overline{s}, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) \Gamma_{s'}\right)$$

So, (11) is verified, and the thesis is proved. □

6 Greedy Players Satisfaction

We now consider the presence of greedy players, i.e. players having a myopic perspective of the game and who only look to get the highest reward in the single stage game. We try to find conditions under which greedy players are satisfied as well.
In this section we consider the coalition value $v^{(\beta)}(\Lambda, \Gamma_s)$ to be the β-discounted value of the two player zero sum game of coalition Λ against $\mathcal{C}\backslash\Lambda$ in the game Γ_s. This concept is expressed by Condition 1.

Condition 1 (Maxmin coalition values)**.** *The coalition value $v^{(\beta)}(\Lambda, \Gamma_s)$ is computed as the max-min expression in (3).*

Let Ω_s be the single stage game in state s, for any $s \in S$. We assume that Ω_s is also a TU game, in which the coalition value $v(\Lambda, \Omega_s)$ is, analogously to (3), the value of the zero sum game played by the coalition Λ against $\mathcal{C}\backslash\Lambda$, for each $\Lambda \subseteq \mathcal{C}$. Obviously, $v^{(0)}(\Lambda, \Gamma_s) \equiv v(\Lambda, \Omega_s)$.

The new property that we are seeking for in this section can be summarized as follows.

Property 4 (Greedy players satisfaction)**.** *Set $\beta \in [0; 1)$. For all $s \in S$, the CPDP $\mathbf{g}^{(\beta)}(s)$ belongs to Core of the stage-wise game Ω_s, i.e. $\mathbf{g}^{(\beta)}(s) \in \mathbf{Co}(\Omega_s)$.*

The intuition here is to let the discount factor β tend to zero and to probe under which conditions $\boldsymbol{\gamma}^{(\beta)}(s)$ lies in $\mathbf{Co}(\Omega_s)$. For this purpose, in the current section we consider $\mathbf{T} \equiv \mathbf{Sh}$.

Lemma 1. *There exists a pure strategy $\underline{\mathbf{f}}^*_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}$ and $\beta^* > 0$ such that $\underline{\mathbf{f}}^*_{\mathcal{C}}$ is optimal for all $\beta \in [0; \beta^*)$.*

Proof. The global optimization problem is a Markov Decision Process (MDP) having $\boldsymbol{\Phi}^{(\beta)}_{\mathcal{C}}$ as discounted reward. Take a strictly decreasing sequence $\{\beta_k\}$ such that $\lim_{k\to\infty} \beta_k = 0$. Since both the actions and the states have a finite cardinality, then there exists a pure strategy $\underline{\mathbf{f}}^*_{\mathcal{C}}$ and an infinite subsequence of $\{\beta_k\}$, namely $\{\beta_{n_k}\}$, with $n_k < n_{k+1}\ \forall\, k$, such that $\underline{\mathbf{f}}^*_{\mathcal{C}}$ is optimal for all the discount factors $\{\beta_{n_k}\}$. Fix a pure strategy $\mathbf{f}_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}$. Then

$$y^{(\beta_{n_k})}(s, \mathbf{f}_{\mathcal{C}}) = \Phi^{(\beta_{n_k})}_{\mathcal{C}}(s, \underline{\mathbf{f}}^*_{\mathcal{C}}) - \Phi^{(\beta_{n_k})}_{\mathcal{C}}(s, \mathbf{f}_{\mathcal{C}}) \geq 0, \qquad \forall\, k \in \mathbb{N}. \tag{13}$$

It is easy to see that $y^{(\beta)}$, with $\beta \in (0; 1)$, is a continuous rational function. Then, either it is identically zero for all $\beta \in (0; 1)$ or $y^{(\beta)} = 0$ in a finite number of points in the interval $(0; 1)$. Hence, for (13), there exists $\beta^*(s, \mathbf{f}_{\mathcal{C}}) > 0$ such that $y^{(\beta)}(s, \mathbf{f}_{\mathcal{C}}) \geq 0$, for all $\beta \in (0; \beta^*(s, \mathbf{f}_{\mathcal{C}}))$. Take $\beta^* = \min_{s, \mathbf{f}_{\mathcal{C}}} \beta^*(s, \mathbf{f}_{\mathcal{C}}) > 0$.
Since $\Phi^{(\beta)}_{\mathcal{C}}(s, \underline{\mathbf{f}}^*_{\mathcal{C}})$ is also continuous in $\beta = 0$ from the right, then $\underline{\mathbf{f}}^*_{\mathcal{C}}$ is also optimal for $\beta = 0$. The thesis is proved. □

Define now Θ_s as the affine space:

$$\Theta_s : \left\{ \mathbf{x} \in \mathbb{R}^P : \sum_{i \in \mathcal{C}} \mathbf{x}_i = \sum_{i \in \mathcal{C}} \mathbf{r}_i(s, \underline{\mathbf{f}}^*_{\mathcal{C}}) \right\}, \tag{14}$$

where $\underline{\mathbf{f}}^*_{\mathcal{C}}$ is the global optimal strategy for all discount factors sufficiently close to 0.

Corollary 1. *For any $s \in S$, $\boldsymbol{\gamma}^{(\beta)}(s)$ belongs to the affine space Θ_s, for all β sufficiently close to 0.*

Proof. The proof follows straightforward from Theorem 1 and from Lemma 1. □

Here we present a useful result.

Lemma 2. *Let $\mathbf{T} \equiv \mathbf{Sh}$. Under Condition 1, $\lim_{\beta \downarrow 0} \boldsymbol{\gamma}^{(\beta)}(s) = \mathbf{Sh}^{(0)}(\Gamma_s) \equiv \mathbf{Sh}(\Omega_s)$.*

Proof. Recall the expression (9) of $\boldsymbol{\gamma}^{(\beta)}$, that we rewrite as

$$\boldsymbol{\gamma}^{(\beta)}(s) = \sum_{s' \in S} \left[\delta_{s,s'} - \beta\, p(s' | s, \mathbf{f}^{(\beta)*}_{\mathcal{C}}) \right] \mathbf{Sh}^{(\beta)}(\Gamma_{s'}),\ \forall\, s \in S.$$

It is sufficient to prove that $\lim_{\beta\downarrow 0}\mathbf{Sh}^{(\beta)}(\Gamma_s)=\mathbf{Sh}^{(0)}(\Gamma_s)$, $\forall\, s\in S$. Since each component of the vector $\mathbf{Sh}^{(\beta)}(\Gamma_s)$ is a linear combination of the discounted values $\{v_\beta(\Lambda,\Gamma_s)\}_{\Lambda\subseteq\mathcal{C}}$, then we only need to show that

$$\lim_{\beta\downarrow 0} v^{(\beta)}(\Lambda,\Gamma_s)=v^{(0)}(\Lambda,\Gamma_s)\equiv v(\Lambda,\Omega_s),\ \forall\, s\in S,\ \forall\,\Lambda\subseteq\mathcal{C}.$$

First of all we recall the relation (Filar and Vrieze 1996)

$$|\operatorname{val}(\mathbf{B})-\operatorname{val}(\mathbf{C})|\le\max_{i,j}|\mathbf{B}_{i,j}-\mathbf{C}_{i,j}| \tag{15}$$

where $\mathbf{B},\mathbf{C}$ are matrices with the same size. We know from (Filar and Vrieze 1996) that

$$\begin{aligned} v^{(\beta)}(\Lambda,\Gamma_s)=\ \operatorname{val}\Bigg(\Big[\sum_{i\in\Lambda}\mathbf{r}_i(s,a_\Lambda,a_{\mathcal{C}\setminus\Lambda})\ +\ldots \\ +\beta\sum_{s'\in S}p(s'|s,a_\Lambda,a_{\mathcal{C}\setminus\Lambda})\,v^{(\beta)}(\Lambda,\Gamma_{s'})\Big]_{a_\Lambda=1,a_{\mathcal{C}\setminus\Lambda}=1}^{m_\Lambda(s),m_{\mathcal{C}\setminus\Lambda}(s)}\Bigg), \end{aligned} \tag{16}$$

where $a_\Lambda\in A_\Lambda(s)$ and $a_{\mathcal{C}\setminus\Lambda}\in A_{\mathcal{C}\setminus\Lambda}(s)$. Thus, from (15,16) we can say that, for all $\Lambda\subseteq\mathcal{C}$,

$$\begin{aligned} |v^{(\beta)}(\Lambda,\Gamma_s)-v^{(0)}(\Lambda,\Gamma_s)|&\le\max_{a_\Lambda,a_{\mathcal{C}\setminus\Lambda}}\Big|\beta\sum_{s'\in S}p(s'|s,a_\Lambda,a_{\mathcal{C}\setminus\Lambda})\,v^{(\beta)}(\Lambda,\Gamma_{s'})\Big| \\ &\le\frac{\beta}{1-\beta}\,M \end{aligned}$$

where $M=\max_{s,a_\Lambda,a_{\mathcal{C}\setminus\Lambda}}|r_\Lambda(s,a_\Lambda,a_{\mathcal{C}\setminus\Lambda})|$. Fix $\epsilon>0$. Set $\delta=\epsilon/(M+\epsilon)$. Then for all $\beta\in[0;\delta)$, we have $|v^{(\beta)}(\Lambda,\Gamma_s)-v^{(0)}(\Lambda,\Gamma_s)|<\epsilon$. Hence, $v^{(\beta)}(\Lambda,\Gamma_s)$ is right continuous in β at $\beta=0$ for all $s\in S$, $\Lambda\subseteq\mathcal{C}$. □

Let us formulate an additional condition, which holds only in the current section.

Condition 2 (Stage-wise strict convexity)**.** *The single stage games $\{\Omega_s\}_{s\in S}$ are strictly convex, i.e. $v(\Lambda_1\cup\Lambda_2,\Omega_s)+v(\Lambda_1\cap\Lambda_2,\Omega_s)>v(\Lambda_1,\Omega_s)+v(\Lambda_2,\Omega_s)$, $\forall\, s\in S,\ \forall\,\Lambda_1,\Lambda_2\subseteq\mathcal{C}$.*

We know from Shapley (1971) that, if Condition 2 holds, then the Core of Ω_s is $(P-1)$-dimensional for any $s\in S$, i.e. the affine hull of $\mathbf{Co}(\Omega_s)$ coincides with Θ_s in (14), for any $s\in S$. Note that, in general, the affine hull of $\mathbf{Co}(\Omega_s)$ could be a strict subset of Θ_s.

Corollary 2. *Suppose that the stage-wise strict convexity Condition 2 holds. Then*

(i) *the Shapley value of Ω_s lie in the relative interior of $\mathbf{Co}(\Omega_s)$, for any $s\in S$;*

(ii) *the interior of* $\mathbf{Co}(\Omega_s)$ *relative to* Θ_s *coincides with the strict Core* $\mathbf{sCo}(\Omega_s)$*, for any* $s \in S$*.*

Proof. For the proof of *(i)*, see Shapley (1971). Now we prove *(ii)*. Fix a generic $s \in S$. If for a coalition $\underline{\Lambda} \subset \mathcal{C}$, $\sum_{i \in \underline{\Lambda}} \mathbf{x}_i = v(\underline{\Lambda}, \Omega_s)$, then take (k, j) such that $j \in \underline{\Lambda},\ k \notin \underline{\Lambda}$. For all $\alpha \in \mathbb{R}$, the vector $\mathbf{x}^{(kj)} = \mathbf{x} + \alpha[\mathbf{e}^{(k)} - \mathbf{e}^{(j)}]$ does not lie in $\mathbf{Co}(\Omega_s)$, where $\mathbf{e}^{(i)} \in \mathbb{R}^P$ is 1 in its i-th component and 0 elsewhere. Hence, $\mathbf{x}$ does not belong to the relative interior of $\mathbf{Co}(\Omega_s)$.
Conversely, if a vector $\mathbf{x} \in \mathbf{sCo}(\Omega_s)$, then it is straightforward to see that it also belongs to the relative interior of $\mathbf{Co}(\Omega_s)$. □

Theorem 4. *Let* $\boldsymbol{\gamma}^\beta$ *be the MDP-CPDP associated to the terminal cooperative solution* $\mathbf{T}$*. Consider* $\mathbf{T}(\Gamma_s) \equiv \mathbf{Sh}(\Gamma_s)$*, for all* $s \in S$*. Then, under Conditions 1 and 2, the greedy players satisfaction Property 4 is verified by* $\boldsymbol{\gamma}^{(\beta)}$ *for all discount factors* β *sufficiently close to* 0*.*

Proof. Take $\beta^* > 0$, such that $\underline{\mathbf{f}}^*_{\mathcal{C}}$ is global optimum for all $\beta \in [0, \beta^*)$. Fix $s \in S$. We know from Corollary 2 that $\mathbf{Sh}(\Omega_s)$ lies in the relative interior of $\mathbf{Co}(\Omega_s)$. The affine hull of $\mathbf{Co}(\Omega_s)$ coincides with the hyperplane Θ_s for Condition 2. Moreover, from Corollary 1 we know that, for all $s \in S$, $\boldsymbol{\gamma}^{(\beta)}(s)$ belongs to the affine space Θ_s for all $\beta \in [0, \beta^*)$. Hence, for Lemma 2 we can say that for all $\epsilon > 0$ there exists $\delta_s \in (0, \beta^*)$ such that

$$\forall\, \beta \in [0; \delta_s),\ \boldsymbol{\gamma}^\beta(s) \in [B_{\delta_s} \cap \Theta_s] \subseteq \mathbf{Co}(\Omega_s),$$

where B_{δ_s} is the ball belonging to $\mathbb{R}^P$ having radius of δ_s. Take $\delta = \min_{s \in S} \delta_s$. The thesis is proved. □

Hence, under Condition 2, for all $\beta \in [0; \delta)$, all the greedy players are content with the stage-wise allocation as well.

7 Transition probabilities not depending on the actions

In this section we deal with a special case of our model, entaling that the transition probabilities among the states do not depend on the players' strategies.

Condition 3. *The actions taken by players in state* s *do not influence the transition probabilities from state* s*, i.e.* $p(s'|s, a_1, \ldots, a_P) = p(s'|s)$*, for all* $a_i \in A_i(s)$ *and for each* $s, s' \in S$*.*

Like in Section 6, we consider the single stage game Ω_s to possess transferable utilities $\{v(\Lambda, \Omega_s)\}_{s \in S, \Lambda \subseteq \mathcal{C}}$. Nevertheless, we no longer impose the maxmin Condition 1 on the coalition values. This model is equivalent to the one of Predtetchinski (2007), except for the TU assumption. Let us provide our main result of this section. It states that, under Condition 3, if we choose a stage-wise allocation belonging to the Core of each single stage game, this is actually a MDP-CPDP,

fulfilling the greedy players satifaction Property 4 and whose discounted long run sum belongs to the Core of each long run game Γ_s, $s \in S$.

Theorem 5. *Set $\beta \in [0;1)$. Let $\overline{\mathbf{T}}^{(\beta)}(\Gamma_s) \in \mathbb{R}^P$ be a terminal cooperative solution, for all $s \in S$. Let the stage wise allocation $\boldsymbol{\gamma}^{(\beta)}$ be the MDP-CPDP associated to $\overline{\mathbf{T}}^{(\beta)}$. Under Condition 3, if $\boldsymbol{\gamma}^{(\beta)}$ fulfills the greedy players satisfaction Property 4 for all $s \in S$, then $\overline{\mathbf{T}}(\Gamma_s) \in \mathbf{Co}^{(\beta)}(\Gamma_s)$, for all $s \in S$.*

Proof. For each $\Lambda \subseteq \mathcal{C}$, let $\mathcal{V}(\Omega_s, \Lambda)$ and $\mathcal{V}(\Lambda, \Gamma_s)$ be the set of feasible allocations for coalition Λ in the games Ω_s and Γ_s, respectively. Since the transition probability matrix does not depend on the players' actions, we can write

$$\begin{bmatrix} \mathcal{V}(\Gamma_{s_1}, \Lambda) \\ \vdots \\ \mathcal{V}(\Gamma_{s_N}, \Lambda) \end{bmatrix} = (\mathbf{I} - \beta \mathbf{P})^{-1} \begin{bmatrix} \mathcal{V}(\Omega_{s_1}, \Lambda) \\ \vdots \\ \mathcal{V}(\Omega_{s_N}, \Lambda) \end{bmatrix}, \qquad \forall\, \Lambda \subseteq \mathcal{C}. \tag{17}$$

Since the matrix $(\mathbf{I} - \beta \mathbf{P})^{-1}$ is non negative, the thesis follows straighforward from Proposition 2. □

In Section 6 we showed that, when the transition probabilities among the states depend on the players' actions, a MDP-CPDP fulfills the greedy players satisfaction Property 4 provided that $\mathbf{T} \equiv \mathbf{Sh}$, the single stage games $\{\Omega_s\}_{s \in S}$ are strictly convex and β is sufficiently close to zero. It is interesting that instead, in this case, we only need to assume that the games $\{\Omega_s\}_{s \in S}$ all possess a non empty Core, in order to fulfill Property 4 for *all* $\beta \in [0;1)$.
The reader should also notice that the converse of Theorem 5 is not true. Indeed, it is possible to find a terminal cooperative solution belonging to the Core of the long run games Γ_s, for all $s \in S$, to which it is associated a MDP-CPDP outside the Core of at least one single stage games Ω_s.

We conclude here by providing the analogous result of Theorem 5 for the Shapley value. The proof follows straightforward from (17) and from Proposition 3.

Corollary 3. *Set $\beta \in [0;1)$. Let $\overline{\mathbf{T}}^{(\beta)}(\Gamma_s) \in \mathbb{R}^P$ be a terminal cooperative solution, for all $s \in S$. Let $\boldsymbol{\gamma}^{(\beta)}$ be the MDP-CPDP associated to $\overline{\mathbf{T}}^{(\beta)}$. Under Condition 3, $\boldsymbol{\gamma}^{(\beta)}(s) = \mathbf{Sh}(\Omega_s)$, for all $s \in S$, if and only if $\overline{\mathbf{T}}(\Gamma_s) = \mathbf{Sh}(\Gamma_s)$, for all $s \in S$.*

It is now interesting to investigate about the loss incurred in the long run game by a greedy coalition of players which withdraws from the grand coalition in a stage of the game.

8 Cooperation Maintenance

The (single step) cooperation maintenance property was first introduced by Mazalov and Rettieva (2010), who employed it in a deterministic fish war setting.

Such a property helps to preserve the cooperation agreement throughout the game, since the long run payoff that each coalition expects to get by deviating in the next stage of the game is not smaller than the payoff that the coalition receives by deviating in the current stage. We now adapt it to our cooperative MDP model. For simplicity, we restrict the following definitions to stationary CPDPs.

Property 5 (First step cooperation maintenance)**.** *Set $\beta \in [0;1)$. The stationary CPDP $\mathbf{g}^{(\beta)}$ satisfies, for any initial state $s \in S$ and for each coalition $\Lambda \subset \mathcal{C}$,*

$$\sum_{i\in\Lambda} \mathbf{g}_i^{(\beta)}(s) + \beta v^{(\beta)}\left(\Lambda, \sum_{s'\in S} p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right) \geq v^{(\beta)}(\Lambda, \Gamma_s).$$

In other words, Property 5 claims that each coalition is always incentivated to postpone the moment in which it will withdraw from the grand coalition, under the condition that, once a coalition $\Lambda \subset \mathcal{C}$ is formed, it can no longer rejoin the grand coalition in the future. By induction, we can say that the cooperation maintenance property enforces the grand coalition agreement throughout the whole game.

8.1 *n*-tuple step cooperation maintenance

We now generalize Property 5, by considering the dilemma faced by a coalition which decides whether deviating in the current stage or after n steps. Hence, let us then define the n-tuple step cooperation maintenance property, with $n \geq 1$.

Property 6 (*n*-tuple step cooperation maintenance)**.** *Set $\beta \in [0;1)$. Let the integer $n \geq 1$. The stationary CPDP $\mathbf{g}^{(\beta)}$ satisfies the n-tuple step cooperation maintenance property iff, for any initial state $s \in S$ and for each coalition $\Lambda \subset \mathcal{C}$,*

$$\sum_{t=0}^{n-1} \beta^t p_t(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*}) \sum_{i\in\Lambda} \mathbf{g}_i^{(\beta)}(s') + \beta^n v^{(\beta)}\left(\Lambda, \sum_{s'\in S} p_n(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right) \geq v^{(\beta)}(\Lambda, \Gamma_s).$$

Let $\mathbf{P}^{*(\beta)} \equiv \mathbf{P}^{(\beta)}(\mathbf{f}_{\mathcal{C}}^{(\beta)*})$ be the transition probability matrix associated to the global optimal stationary strategy $\mathbf{f}_{\mathcal{C}}^{(\beta)*}$, whose (i,j) element is $p(s_j|s_i, \mathbf{f}_{\mathcal{C}}^{(\beta)*})$.

We now find a necessary and sufficient condition on the coalition values $v^{(\beta)}$ to ensure the existence of our MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$, defined in (9), satisfying the n-tuple step cooperation maintenance property, for any $n \geq 1$. Let us denote $\mathbf{v}^{(\beta)}(\Lambda)$ as

$$\mathbf{v}^{(\beta)}(\Lambda) \equiv \left[v^{(\beta)}(\Lambda, \Gamma_{s_1}) \ \ldots \ v^{(\beta)}(\Lambda, \Gamma_{s_N})\right]^T, \qquad \forall\, \Lambda \subseteq \mathcal{C}.$$

Theorem 6. *Fix an integer $n \geq 1$, $\beta \in [0;1)$. The set of stationary CPDPs $\boldsymbol{\gamma}^{(\beta)}$ satisfying the n-tuple step cooperation maintenance Property 6 is nonempty if and only if the vectors*

$$\widetilde{\mathbf{v}}^{(\beta,n)}(\Lambda) = \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \mathbf{v}^{(\beta)}(\Lambda), \qquad \Lambda \subseteq \mathcal{C}$$

are component-wisely balanced, i.e. for every function $\alpha_s : 2^P/\{\emptyset\} \to [0;1]$ *such that:*

$$\forall\, i \in \mathcal{C} : \sum_{\substack{\Lambda \subseteq \mathcal{C}: \\ \Lambda \ni i}} \alpha_s(\Lambda) = 1,$$

the following condition holds:

$$\sum_{\Lambda \subseteq \mathcal{C}} \alpha_s(\Lambda) \widetilde{\mathbf{v}}_k^{(\beta,n)}(\Lambda) \leq \widetilde{\mathbf{v}}_k^{(\beta,n)}(\mathcal{C}), \qquad \forall\, k \in [1;N],$$

where $\widetilde{\mathbf{v}}_k^{(\beta,n)}(\Lambda)$ *is the* k*-th component of* $\widetilde{\mathbf{v}}^{(\beta,n)}(\Lambda)$.

Proof. Recall the expression of $\boldsymbol{\gamma}^{(\beta)}$ in equation (9), that can be rewritten as:

$$\boldsymbol{\gamma}_i^{(\beta)} = \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \overline{\mathbf{T}}_i^{(\beta)}, \quad \forall\, i \in \mathcal{C} \tag{18}$$

where $\boldsymbol{\gamma}_i^{(\beta)} = [\boldsymbol{\gamma}_i^{(\beta)}(s_1) \dots \boldsymbol{\gamma}_i^{(\beta)}(s_N)]^T$, $\overline{\mathbf{T}}_i^{(\beta)} = [\overline{\mathbf{T}}_i^{(\beta)}(\Gamma_{s_1}) \dots \overline{\mathbf{T}}_i^{(\beta)}(\Gamma_{s_N})]^T \in \mathbf{T}^{(\beta)}(\Gamma_s)$ for each state $s \in S$. By exploiting twice the well known formula for matrix geometric series:

$$\sum_{k=0}^{n-1} \left[\beta \mathbf{P}^{*(\beta)}\right]^k = \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right]^{-1} \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right]$$

we can reformulate Property 6 as

$$\begin{cases} \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} \geq \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \mathbf{v}^{(\beta)}(\Lambda), \quad \forall\, \Lambda \subset \mathcal{C} \\ \sum_{i \in \mathcal{C}} \overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}) \end{cases} \tag{19}$$

where the second relation in (19) comes from the classic efficiency property of a cooperative solution. Since the matrix $(\mathbf{I} - [\beta \mathbf{P}^{*(\beta)}]^n)$ is invertible, then we can equivalently rewrite (19) as

$$\begin{cases} \sum_{i \in \Lambda} \widetilde{\overline{\mathbf{T}}}_i^{(\beta,n)} \geq \widetilde{\mathbf{v}}^{(\beta,n)}(\Lambda), \ \forall\, \Lambda \subset \mathcal{C} \\ \sum_{i \in \mathcal{C}} \widetilde{\overline{\mathbf{T}}}_i^{(\beta,n)} = \widetilde{\mathbf{v}}^{(\beta,n)}(\mathcal{C}) \end{cases} \tag{20}$$

where

$$\widetilde{\overline{\mathbf{T}}}_i^{(\beta,n)} = \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \overline{\mathbf{T}}_i^{(\beta)}$$

Since the relations in the systems of inequalities in (20) are component-wise, for the Bondareva-Shapley Theorem (Bondareva 1963; Shapley 1967) the thesis is proved. □

The reader should note that, in the limit for $n \to \infty$, the result of Theorem 6 coincides with the Bondareva-Shapley Theorem for static cooperative games.

We now state an important and intuitive result which further reinforces the importance of the single step cooperation maintenance property.

Theorem 7. *Set $\beta \in [0;1)$. If the MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies the single step cooperation maintenance Property 5, then it satisfies the n-tuple step cooperation maintenance Property 6, for all $n > 1$.*

Proof. Let $\boldsymbol{\gamma}^{(\beta)}$ be defined in (18), where $\overline{\mathbf{T}}^{(\beta)}$ satisfies the single step cooperation maintenance Property 5, i.e., from (19),

$$\begin{cases} \beta \mathbf{P}^{*(\beta)} \left[\sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda) \right] \geq \sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda), \quad \forall\, \Lambda \subset \mathcal{C} \\ \sum_{i \in \mathcal{C}} \overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}) \end{cases} \tag{21}$$

By iteratively left multiplying by the nonnegative matrix $\beta \mathbf{P}^{*(\beta)}$ both sides of the first relation in (21), for each coalition $\Lambda \subset \mathcal{C}$, we obtain

$$\sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda) \leq \beta \mathbf{P}^{*(\beta)} \left[\sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda) \right] \leq \left[\beta \mathbf{P}^{*(\beta)} \right]^2 \left[\sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda) \right] \leq \dots$$

Hence, the thesis is proved. □

8.2 Core selection criterion

In the following we prove that the single step cooperation maintenance Property 5 also implies that the discounted sum of allocations for each player, when s is the initial state, belongs to the Core of the game Γ_s,

Corollary 4. *Set $\beta \in [0;1)$. If a MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies the single step cooperation maintenance Property 5, then*

$$E \left[\sum_{t \geq 0} \beta^t \boldsymbol{\gamma}^{(\beta)}(S_t) | S_0 = s \right] \in \mathbf{Co}^{(\beta)}(\Gamma_s), \quad \forall\, s \in S. \tag{22}$$

Proof. Let us define $\boldsymbol{\gamma}^{(\beta)}$ as in (18). We reformulate (22) as

$$\begin{cases} \sum_{i \in \Lambda} \overline{\mathbf{T}}_i^{(\beta)} \leq \mathbf{v}^{(\beta)}(\Lambda), & \forall\, \Lambda \subset \mathcal{C}, \\ \sum_{i \in \mathcal{C}} \overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}). \end{cases} \tag{23}$$

Since $\boldsymbol{\gamma}^{(\beta)}$ satisfies Property 5, then (19) is verified, with $n = 1$. By left multiplying each set of inequalties in (19) by the nonnegative matrix $(\mathbf{I} - \beta \mathbf{P}^{*(\beta)})^{-1}$, we obtain the system of inequalities in (23). □

In this section we showed how appealing the single step cooperation maintenance property is. For Theorem 7, if our MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ fulfills it, then each coalition always prefers to withdraw from the grand coalition in the future, other than at the current stage.

In the case we consider the Core as the terminal cooperative solution ($\mathbf{T} \equiv \mathbf{Co}$), Corollary 4 suggests that the point of the Core $\overline{\mathbf{T}}^{(\beta)}$ used to compute the

MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ in equation (9) should be picked such that $\overline{\mathbf{T}}^{(\beta)}$ also satisfies the single step cooperation maintenance property. In this sense, Property 5 is also a *Core selection* criterion.

8.2.1 Counterexample for the converse of Corollary 4

It is natural to ask whether the converse of Corollary 4 is true. We will show in the following example that it does not hold in general, i.e. if a MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies (23), then not necessarily the single step cooperation maintenance Property 5 holds.

Let us consider a cooperative MDP with only two players ($P = 2$), four states ($N = 4$) and with perfect information, i.e. in each state at most one player has more than one action available. Player 1 controls states (s_1, s_2), and the remaining states (s_3, s_4) are controlled by player 2. Let the discount factor $\beta = 0.8$. The immediate rewards for each player and the transition probabilities for each state/action pair are shown in the following table.

	(s,a)	r_1	r_2	$p(s_1\|s,a)$	$p(s_2\|s,a)$	$p(s_3\|s,a)$	$p(s_4\|s,a)$
pl. 1	(s_1,a_1)	1	3	0.1	0.4	0.1	0.4
	(s_1,a_2)	2	1	0.4	0.1	0.1	0.3
	(s_1,a_3)	1	0	0.4	0.2	0.4	0.1
	(s_2,a_4)	2	1	0.1	0	0.4	0.4
	(s_2,a_5)	3	1	0.2	0.2	0.2	0.5
	(s_2,a_6)	4	3	0.2	0	0.2	0.3
pl. 2	(s_3,a_7)	5	1	0.3	0.6	0.4	0.1
	(s_3,a_8)	1	3	0.3	0.4	0.2	0
	(s_3,a_9)	2	6	0.3	0.3	0.1	0
	(s_4,a_{10})	0	1	0.5	0	0.1	0.1
	(s_4,a_{11})	2	2	0.1	0.3	0.5	0.2
	(s_4,a_{12})	3	0	0.1	0.5	0.3	0.6

Table 1: *Immediate rewards and transition probabilities for each player, state, and strategy.*

In this case, the state-wise value vectors for all the possible coalitions $\{1\}$, $\{2\}$ and $\mathcal{C} = \{1, 2\}$, rounded off to the second decimal, are

$$\mathbf{v}^{(0.8)}(\{1\}) \approx \begin{bmatrix} 8.73 \\ 10.03 \\ 7.34 \\ 7.16 \end{bmatrix}, \quad \mathbf{v}^{(0.8)}(\{2\}) \approx \begin{bmatrix} 9.57 \\ 8.65 \\ 10.93 \\ 11.23 \end{bmatrix}, \quad \mathbf{v}^{(0.8)}(\{1,2\}) \approx \begin{bmatrix} 33.08 \\ 30.78 \\ 33.77 \\ 30.83 \end{bmatrix}.$$

In order to contradict the converse of Corollary 4, it is sufficient to find a specific long run allocation $\overline{\mathbf{T}}^{(0.8)}$ such that

$$[\overline{\mathbf{T}}_1^{(0.8)}(s_k)\ \overline{\mathbf{T}}_2^{(0.8)}(s_k)] \in \mathbf{Co}^{(0.8)}(\Gamma_{s_k}), \qquad k = 1, 2, 3, 4, \tag{24}$$

but for which the 4-by-1 MDP-CPDP:

$$\boldsymbol{\gamma}_j^{(\beta)} = \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \overline{\mathbf{T}}_j^{(\beta)}, \quad j = 1, 2$$

does not respect the single step cooperation maintenance property for some initial state s. In other words, we look for $(\overline{\mathbf{T}}_1^{(0.8)}, \overline{\mathbf{T}}_2^{(0.8)})$ such that

$$\begin{cases} \overline{\mathbf{T}}_1^{(0.8)} \geq \mathbf{v}^{(0.8)}(\{1\}) \\ \overline{\mathbf{T}}_2^{(0.8)} \geq \mathbf{v}^{(0.8)}(\{2\}) \\ \overline{\mathbf{T}}_1^{(0.8)} + \overline{\mathbf{T}}_2^{(0.8)} = \mathbf{v}^{(0.8)}(\{1,2\}) \end{cases} \tag{25}$$

and such that there exists at least one player i and an integer $k \in [1; 4]$ such that

$$\widetilde{\overline{\mathbf{T}}}_i^{(0.8)}(k) < \widetilde{\mathbf{v}}_k^{(0.8)}(\{i\})$$

where

$$\begin{aligned} \widetilde{\overline{\mathbf{T}}}_i^{(0.8)} &= \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \overline{\mathbf{T}}_i^{(0.8)} \\ \widetilde{\mathbf{v}}^{(0.8)}(\{i\}) &= \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \mathbf{v}^{(0.8)}(\{i\}) \quad i = 1, 2. \end{aligned} \tag{26}$$

Since the values are component-wisely superadditive by construction, then the Core $\mathbf{Co}(\Gamma_s)$ for the two-player case always exists, for all $s \in S$. Hence, there always exist $(\overline{\mathbf{T}}_1^{(0.8)}, \overline{\mathbf{T}}_2^{(0.8)}) \in \mathbb{R}^2$ satisfying (25). Let us select:

$$\overline{\mathbf{T}}_1^{(0.8)} = \mathbf{v}^{(0.8)}(\{1\}) + \begin{bmatrix} 0.7 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \left[\mathbf{v}^{(0.8)}(\{1,2\}) - [\mathbf{v}^{(0.8)}(\{1\}) + \mathbf{v}^{(0.8)}(\{2\})]\right]$$

$$\overline{\mathbf{T}}_2^{(0.8)} = \mathbf{v}^{(0.8)}(\{2\}) + \begin{bmatrix} 0.3 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \left[\mathbf{v}^{(0.8)}(\{1,2\}) - [\mathbf{v}^{(0.8)}(\{1\}) + \mathbf{v}^{(0.8)}(\{2\})]\right]$$

Substituting the values of $\mathbf{v}^{(0.8)}$, we obtain

$$\begin{aligned} \overline{\mathbf{T}}_1^{(0.8)} &\approx \begin{bmatrix} 19.07 & 14.87 & 10.44 & 19.60 \end{bmatrix}^T \\ \overline{\mathbf{T}}_2^{(0.8)} &\approx \begin{bmatrix} 14.01 & 15.91 & 23.32 & 11.23 \end{bmatrix}^T \end{aligned}$$

By computing $\widetilde{\overline{\mathbf{T}}}^{(0.8)}$ and $\widetilde{\mathbf{v}}^{(0.8)}$ we find that:

$$\begin{aligned}
\widetilde{\overline{\mathbf{T}}}_1^{(0.8)}(2) \approx 2.92 < \widetilde{\mathbf{v}}_2^{(0.8)}(\{1\}) \approx 3.65 \\
\widetilde{\overline{\mathbf{T}}}_1^{(0.8)}(3) \approx -0.75 < \widetilde{\mathbf{v}}_3^{(0.8)}(\{1\}) \approx 0.51 \\
\widetilde{\overline{\mathbf{T}}}_2^{(0.8)}(1) \approx 0.48 < \widetilde{\mathbf{v}}_1^{(0.8)}(\{2\}) \approx 1.61 \\
\widetilde{\overline{\mathbf{T}}}_2^{(0.8)}(4) \approx 0.90 < \widetilde{\mathbf{v}}_4^{(0.8)}(\{2\}) \approx 3.00
\end{aligned}$$

Therefore, the converse of Corollary 4 is not true. On the other hand, it is interesting to observe that in this example, by randomly generating vectors $(\overline{\mathbf{T}}_1^{(0.8)}, \overline{\mathbf{T}}_2^{(0.8)})$ and fulfilling the relation (24), in about the 99.45% of the trials the converse of Corollary 4 was verified.

8.3 Strictly convex single stage games

In the spirit of Section 6, we show that the strict convexity Condition 2 on the single stage games ensures the MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ to satisfy Property 5 for all discount factors small enough.

Theorem 8. *Suppose that the strict convexity Condition 2 on the single stage games $\{\Omega_s\}_{s\in S}$ is valid. Consider $\mathbf{T} \equiv \mathbf{Sh}$. Then the single step cooperation maintenance Property 5 is valid for all β close enough to 0.*

Proof. Thanks to the linearity property of coalition values (see Proposition 1) we can reformulate Property 5 as

$$\sum_{i\in\Lambda} \gamma_i^{(\beta)}(s) \geq \sum_{s'\in S} \left[\delta_{s,s'} - \beta p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\right] v^{(\beta)}(\Lambda, \Gamma_{s'}), \qquad \forall\, \Lambda \subset \mathcal{C},\ s \in S.$$

From (9), considering $\mathbf{T} \equiv \mathbf{Sh}$,

$$\sum_{i\in\Lambda} \gamma_i^{(\beta)}(s) = \sum_{s'\in S} \left[\delta_{s,s'} - \beta p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\right] \sum_{i\in\Lambda} \mathbf{Sh}_i^{(\beta)}(\Gamma_{s'}).$$

By hypothesis, for all $s \in S$ the Shapley value $\mathbf{Sh}(\Omega_s) = \mathbf{Sh}^{(0)}(\Gamma_s)$ belongs to the strict Core $\mathbf{sCo}^{(\beta)}(\Omega_s)$ for all β sufficiently close to 0. Hence, by right continuity of the Shapley value and of coalition values in $\beta = 0$ (see proof of Lemma 2), we conclude that, for all β sufficiently close to 0,

$$\sum_{s'\in S} \left[\delta_{s,s'} - \beta p(s'|s, \underline{\mathbf{f}}_{\mathcal{C}}^{*})\right] \left[\sum_{i\in\Lambda} \mathbf{Sh}_i^{(\beta)}(\Gamma_{s'}) - v^{(\beta)}(\Lambda, \Gamma_{s'})\right] \geq 0,$$

where $\underline{\mathbf{f}}_{\mathcal{C}}^{*}$ is the optimal strategy for grand coalition for all β sufficiently small. Hence, the thesis is proved. □

Acknowledgements

This reasearch was supported by "Agence Nationale de la Recherche" with reference ANR-09-VERS-001, and by the European research project SAPHYRE, which is partly funded by the European Union under its FP7 ICT Objective 1.1 - The Network of the Future. The authors would like to thank Professor Eitan Altman for fruitful discussions.

References

[1] K. Avrachenkov, L. Cottatellucci, L. Maggi, Algorithms for uniform optimal strategies in two-player zero-sum stochastic games with perfect information, INRIA Research Report No. 7355, pp. 1-25 (2010).

[2] O.N. Bondareva, Some applications of linear programming methods to the theory of cooperative games, Problemy Kybernetiki, Vol. 10, pp. 119-139 (1963).

[3] J. Filar, L.A. Petrosjan, Dynamic Cooperative Games, International Game Theory Review, Vol. 2, No. 1, pp. 47-65 (2000).

[4] J. Filar, K. Vrieze, Competitive Markov Decision Processes, Springer (1996).

[5] L. Kranich, A. Perea, H. Peters, Dynamic Cooperative Games, Unpublished Mimeo (2001).

[6] A. Hordijk, R. Dekker, L.C.M. Kallenberg, Sensitivity Analysis in Discounted Markov Decision Processes, OR Spektrum, Vol. 7, No. 3, pp. 143-151 (1985).

[7] V.V. Mazalov, A.N. Rettieva, Fish wars and cooperation maintenance, Ecological Modelling, Vol. 221, Issue 12, pp. 1545-1553 (2010).

[8] R.B. Myerson, Game Theory - Analysis of conflict, Harvard University Press (1991).

[9] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton: Princeton University Press (1944).

[10] J. Oviedo, The Core of a Repeated n-Person Cooperative Game, European Journal of Operational Research, Vol. 127, Issue 3, pp. 519-524 (2000).

[11] B. Peleg, P. Sudhölter, Introduction to the Theory of Cooperative Games, Springer, 2nd edition (2007).

[12] L.A. Petrosjan, Cooperative Stochastic Games, Proceedings of the 10th International Symposium on Dynamic Games and Applications, Vol. 2 (2002).

[13] A. Predtetchinski, The strong sequential core for stationary cooperative games, Games and Economic Behavior, Vol. 61, pp. 50-66 (2007).

[14] M.L. Puterman, Markov Decision Processes, Wiley (1994).

[15] L.S. Shapley, On balanced sets and cores, Naval Research Logistics Quarterly, Vol. 14, pp. 453-460 (1967).

[16] L.S. Shapley, Cores of Convex Games, International Journal of Game Theory, Vol. 1, No. 1, pp. 11-26 (1971).