SAPHYRE

Contract No. FP7-ICT-248001

Applied Game Theory (final)
D2.2b

Contractual date:	M30
Actual date:	M30
Authors:	Luca Anchora, Leonardo Badia, Laura Cottatellucci, David Gesbert, Zuleita Ka-Ming Ho, Eduard Jorswieck, Eleftherios Karipidis, Lorenzo Maggi, Christian Scheunert, Anne Wolf, Lei Xiao
Participants:	TUD, CFR, ECM, LiU
Work package:	WP2
Security:	Public
Nature:	Report
Version:	1.0
Number of pages:	185

Abstract

This report describes the final results from the application of game theory to the spectrum and infrastructure sharing problems considered within SAPHYRE. First, noncooperative game theory is used to design transmit beamforming vectors for systems where users access bands with different priorities. Second, the problems of beamforming and power allocation for partial channel knowledge are modelled as Bayesian games and their Nash equilibria are analysed. Third, a framework for cooperative games in dynamics systems, capable to distribute instantaneously the payoff, is developed. Fourth, game-theoretic tools are applied to develop efficient resource allocation algorithms for spectrum and infrastructure sharing problems.

Keywords

Bayesian Games, Beamforming, Game Theory, Interference Channel, MISO, MIMO, Nash Equilibrium, Pricing, Power Control, SISO, Spectrum Sharing.

Executive Summary

The technical goals of SAPHYRE are to demonstrate how physical resource sharing in wireless networks improves spectral efficiency, enhances coverage, increases user satisfaction, and maintains Quality-of-Service (QoS) performance. But, spectrum sharing in wireless systems also comes at the cost of increased interference in the physical layer and gives rise to conflict situations. The main goal of Task 2.2 was to analyse these resource conflicts. Game theory is a branch of applied mathematics that can be used for analysing optimisation problems with multiple conflicting objective functions. In Task 2.2, non-cooperative and cooperative game theory was applied to find important operating points in various achievable utility regions and to motivate distributed algorithms for resource allocation. This report summarizes the main results; it is organised in four chapters.

In Chapter 1, noncooperative game theory is used for the analysis of a multi-user network, where multiple transmitter-receiver (TX-RX) pairs coexist in the same geographical area and utilize the same spectral band. The spectrum sharing scenario is modelled by the multiple-input single-output (MISO) interference channel (IC).

Section 1.1 considers the scenario where multiple secondary TX-RX pairs coexist with multiple primary systems and the secondary TXs are considered to operate under secondary to primary user interference-limitation constraints. Two cases of particular interest are distinguished and the transmission strategies corresponding to Pareto-optimal operation points of the secondary systems are characterised. In the first case, the primary systems are assumed to tolerate a certain amount of interference from secondary systems. This amount of interference is controlled by penalising the secondary systems in proportion to the interference they produce on the primary users. This mechanism is referred to as pricing and is interpreted as introducing the effect of disturbance created from a user as a penalty measure in his utility function. In this means, the secondary TXs can be controlled to choose their transmission strategies satisfying soft interference constraints on the primary users. In the second case, the primary systems are assumed to tolerate no interference from secondary systems at all. That is null-shaping constraints are imposed on the secondary users. The solution to the first case simplifies the second case significantly. It is shown that all points on the Pareto boundary can be achieved as the outcome of a noncooperative game by imposing certain null shaping constraints.

Section 1.2 considers the scenario where two non-cooperating cells, each have a protected band to provide service to its high priority users. A shared band for the two cells is employed to deliver service to low priority users. The situation between the two cells is formulated as a non-cooperative game and its Nash equilibrium (NE) is studied. It is proven that the game belongs to a class of games called supermodular games which have several interesting properties, such as global stability of a unique NE. A sufficient condition is provided for the existence of a unique NE and its efficiency is studied by extensive simulations.

In Chapter 2, spectrum-sharing scenarios are considered with partial channel state information (CSI) at the TXs and RXs. The problems of beamforming and power allocation are modelled as Bayesian games and their NE are analysed.

Section 2.1 considers the multi-antenna spectrum-sharing scenario, which is modelled by the multiple-input multiple-output (MIMO) IC. The beamforming design is studied under two extreme criteria; namely, egoism and altruism. The former maximizes the beamforming gain at the intended RX and the latter minimizes interference created towards other RXs. These designs are motivated by previous studies, where it was shown that combining egoistic and altruistic beamforming is instrumental for optimizing the rates in a MISO IC. Using the framework of Bayesian games, more light is shed on these game-theoretic concepts, in the more general context of MIMO IC and particularly when coordinating parties only have CSI that they can be measured directly. These games are analyzed by establishing the equilibria for the egoistic and altruistic games.

Section 2.2 considers a quasi-static block fading single-input single-output (SISO) IC, with knowledge of the state of the direct links but only statistical knowledge on the interfering links. With this assumption, reliable communications are not possible and a certain level of outage has to be tolerated. Power allocation is considered for utility functions based on the real throughput accounting for the outage events. Power allocation algorithms are proposed based on both Bayesian games and optimization. In the context of Bayesian games, the two cases of power allocation for predefined transmission rates and joint power and rate allocation are investigated. The first game is a concave game and the mathematical tools by Rosen ’65 are adopted for its analysis. The second group of games is studied introducing an equivalent game. The characteristics of the game-theoretical approaches are analyzed in terms of existence and multiplicity of the NE. Special attention is devoted to the extreme regimes of high noise and interference limited regime. In the former case, a closed-form expression for the NE is provided. In the latter case, criteria for the convergence of best response algorithms, the existence and uniqueness of the NE are discussed. The optimization approach is also analyzed in the two above mentioned regimes and closed form expressions for the allocation are provided.

In Chapter 3, spectrum sharing systems which are evolving in time are considered. A framework for cooperative games in dynamics systems, capable to distribute instantaneously the payoff, is proposed. Fundamental mathematical tools are developed to deal with dynamics systems and guarantee the stability of the coalition and a fair instantaneous redistribution of the payoff over time. With this aim, the dynamic of the system is modeled as a Markov Decision Process (MDP), algorithms are developed for distributing the payoff, and the issue of complexity is analyzed. Then, the theoretical results are applied to communication systems.

Section 3.1 provides analytical tools to address the quite frequent situation where some providers share their own resources (e.g. nodes) in a common dynamic network but prefer to keep control of their own nodes. Zero-sum two-player stochastic games with perfect information are studied. Two algorithms are proposed to find the uniform optimal strategies and one method to compute the optimality range of discount factors. The convergence in finite time is proven for one algorithm. The uniform optimal strategies are also optimal for the long run average criterion and, in transient games, for the undiscounted criterion as well.

Section 3.2 extends the previous setting to the case of a system that evolves dynamically and all decision makers/operators can take decision simultaneously, eventually, influencing the evolution of the system itself. Multi-agent MDPs are studied in which cooperation among players is allowed. A cooperative payoff distribution procedure (MDP-CPDP) is found that distributes in the course of the game the payoff that players would get in the long run static game. It is shown under which conditions such a MDP-CPDP fulfills fundamental properties that guarantee the stability of the game, namely, time consistency property, contents greedy players, and strengthen the coalition cohesiveness throughout the game.

Section 3.3 considers a special kind of games known as weighted voting games to investigate suboptimal algorithms with polynomial complexity providing solutions within a controlled confidence interval of the optimum solution. The approximation of the Shapley-Shubik power index in the Markovian game (SSM) is investigated. It is proven that an exponential number of queries on coalition values is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy. Motivated by this, three randomized approaches are proposed to compute a confidence interval for SSM. They rest upon two different assumptions, static and dynamic, about the process through which the estimator agent learns the coalition values.

Section 3.4 studies a cooperative game where several providers coexist by sharing network nodes which are individually controlled. They offer a connection service to the same service point and want to minimize the cost of the service offered to their customers while maximizing the costs for the customers of their opponents by properly defining a routing strategy. Algorithms are provided to determine the network link costs in such a way that all providers have interest in cooperating. As by-product, the proposed algorithm is applied to two-player games both in networks subject to hacker attacks and in epidemic networks.

Section 3.5 considers a multiple access channel (MAC) in which the channel coefficients follow a Markov chain on a finite set of states. It is assumed that any subset of users that does not intend to cooperate can, in the worst case scenario, jam the active users. The feasibility region of the Markovian MAC is derived, under both the discounted and the average criterion. The set of allocation rates in the Markovian process is computed, that are feasible, efficient, and stable. Some fair allocations are analyzed, such as max-min fairness, proportional fairness and Nash bargaining solution. A condition is provided ensuring the consistency of such fairness criteria between each single stage game and the long run game. The situation in which no agreement is reached is investigated and the relation between the already mentioned fair allocations criteria and the Nash bargaining solution is studied, when the number of players increases.

In Chapter 4, game theory is applied to the investigation of both spectrum and infrastructure sharing.

Section 4.1 considers an orthogonal sharing mechanism among the network operators. Many algorithms can be proposed to this end, each one leading to a different result in terms of achievable data rate or allocation fairness. An upper bound on the achievable data rate is obtained by evaluating a centralised algorithm, which represents a cooperative game where the aim of the involved players (i.e. the operators) is to maximise the aggregate capacity instead of their own. A sharing gain is identified, in particular for scenarios where the involved operators have a different traffic load and unused resources can be opportunistically exploited.

Section 4.2 considers the infrastructure sharing scenario of relay sharing. The theory of coalitional games is applied to the case of a network with cross-layer interaction between routing and medium access control. The proper cooperation mechanism to be adopted by the users is investigated so that a gain is obtained by both those who have their transmission relayed to the final destination and also those who act as relays.

Section 4.3 continues the study of infrastructure (relay) sharing. A scenario is considered where the network operators have many relay nodes but are willing to share only a part of them. Then, a proper selection mechanism has to be executed in order to maximise the cell coverage and the overall system throughput. To this aim, a Bayesian Network approach is proposed to compute the correlation between local network parameters and overall performance, whereas the selection of the nodes to be shared is made by means of a cooperative game theoretic approach. It is shown that almost the entire sharing gain is achieved by putting in common just a small fraction of the relays, provided they are selected optimally.

Contents

Notations

Abbreviations

ACK	Acknowledge
ARQ	Automatic Repeat reQuest
BIC	Bayesian Information Criterion
BS	Base Station
BN	Bayesian Network
CDMA	Code Division Multiple Access
CPDP	Cooperative Payoff Distribution Procedure
CQI	Channel Quality Indicator
CSI	Channel State Information
CTS	Clear-To-Send
DAG	Directed Acyclic Graph
dB	decibel
DCI	Dynamic Confidence Interval
FIFO	First-In First-Out
HMC	Homogeneous Markov Chain
IC	Interference Channel
i.i.d.	independent identically distributed
ITC	Interference Temperature Constraints
LTE	Long Term Evolution
MAC	Multiple Access Channel
MDP	Markov Decision Process
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Maximum Likelihood
MRT	Maximum Ratio Transmission
NBS	Nash Bargaining Solution
NE	Nash Equilibrium
NTU	Non-Transferable Utility
OFDMA	Orthogonal Frequency Division Multiple Access
OLSR	Optimized Link State Routing
PSD	Power Spectral Density
QoS	Quality of Service
RRA	Radio Resource Allocator
RTS	Request-To-Send
RX	Receiver

SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SCI	Static Confidence Interval
ShM	Shapley value of cooperative Markovian game
SSM	Shapley-Shubik power index in the Markovian game
TDD	Time-Division Duplex
TDMA	Time Division Multiple Access
TU	Transferable Utility
TX	Transmitter
UE	User Equipment
ZFB	Zero Forcing Beamforming

1 Non-cooperative Games for Multi-antenna Spectrum Sharing Scenarios

1.1 Beamforming in Scenarios with Preferred Users

In this section, we consider a network of licensed primary users/systems that are capable to detect their environment and to reconfigure their operation accordingly. These capabilities are feasible due to measuring and feedback mechanisms in the network [1]. Through the detection of spectrum holes, additional users can be supported in the network to utilize these available resources. Nevertheless, an improved interference management is possible in these systems in order to increase the efficiency of the spectrum utilization.

In a setting where hierarchy exists between the systems, such detection capabilities can be well utilized. Consider a network composed of primary users, the offered radio resources might not be used up completely such that the left over resources can be made available to secondary users. Secondary users can use these resources under the condition of not imposing QoS degradation to the primary systems. This scenario corresponds to the underlay paradigm described in the overview paper [2].

A limited QoS degradation to the primary users is described by interference temperature constraints (ITC) [1]. The ITC in [3] is distinguished in soft- and peak-power-shaping constraints. These constraints refer to the maximum average power or peak average power tolerated at the primary receivers, respectively. In our case, the two types of constraints are equivalent since we consider only single stream beamforming which is shown in [4] to be optimal in the MISO IC. When no interference on the primary users is allowed, the constraint is said to be a null-shaping constraint.

Secondary transmitters equipped with multiple antennas can concentrate their radiation pattern in the direction of their designated receivers or away from primary receivers. Hence, through adaptive beamforming, a balance between the interference induced on primary receivers and spatial multiplexing can be done [5]. In [5], the authors consider the setting of a single secondary transmitter coexisting with multiple primary users and provide optimal transmission strategies for the secondary transmitter under ITC. It is shown that single stream beamforming is optimal in such a scenario. In [6], the authors characterize the Pareto boundary of the MISO IC through controlling the ITCs on the receivers. Each Pareto rate tuple is achieved in a decentralized manner, where each transmitter maximize its rate independently limited to ITCs at the other receivers.

In the following sections, we investigate the design of beamforming vectors that are Pareto optimal for the secondary systems for two settings: (i) In the first setting, the

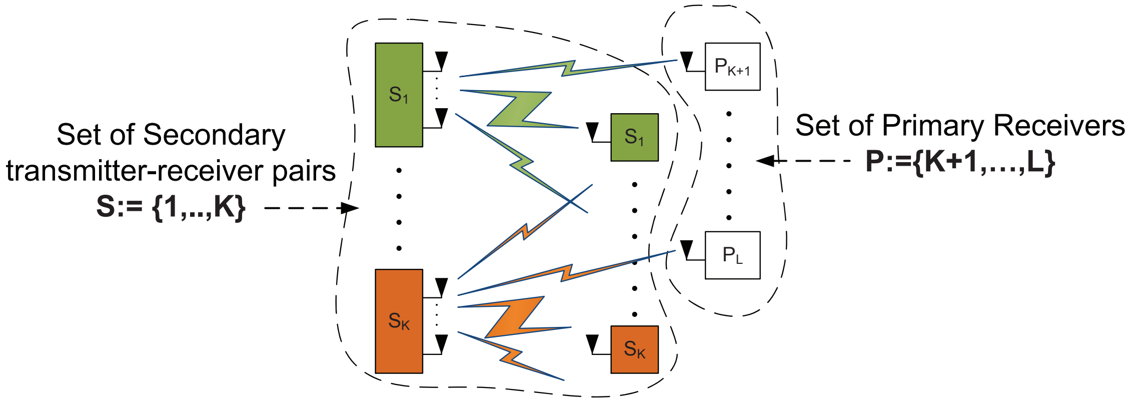

Figure 1.1: A radio network composed of primary systems (smartphones) and secondary systems (laptops).

secondary users are penalized for interfering on the primary systems. The penalty is proportional to the interference rate produced from the secondary transmitter to each primary user. This mechanism is referred to as pricing and is interpreted as introducing the effect of disturbance created from a user as a penalty measure in his utility function [7]. In this means, the secondary transmitters can be controlled to choose their transmission strategies satisfying soft interference constraints on the primary users. In [8], this model of exogenous prices is used to analyze a noncooperative game between the secondary users. Distributed algorithms are provided that iteratively modify the prices weights and eventually reach the Nash equilibrium (NE) that satisfies the ITCs. In our case, the prices weights are assumed to be fixed. (ii) In the second setting, the primary systems do not tolerate any interference from the secondary systems, i.e. null-shaping constraints are imposed on the secondary users. The solution to the first setting simplifies the second scenario significantly.

For both settings, we characterize the transmit beamforming vectors that achieve operation points on the Pareto boundary of the secondary users' utility region. Motivated by distributed operation of the secondary systems, we turn our interest to the design of null-shaping constraints that achieve Pareto-optimal operating points. For the two-user case, we characterize these constraints that are to be imposed on the noncooperative systems such that any point on the Pareto boundary of the MISO IC rate region is achieved in a distributed manner.

Notations. Column vectors and matrices are given in lowercase and uppercase boldface letters respectively. The notation $x_{k,\ell}$ describes the ℓth component of vector $\boldsymbol{x}_k$. The Euclidean norm of a vector $\boldsymbol{a} \in \mathbb{C}^{N\times 1}$ is written as $\|\boldsymbol{a}\|$ and the absolute value of $b \in \mathbb{C}$ is denoted by $|b|$. The eigenvector which corresponds to the ith eigenvalue of the matrix $\boldsymbol{Z}$ is denoted by $V_i(\boldsymbol{Z})$. The eigenvector corresponding to the largest eigenvalue of a matrix $\boldsymbol{Z}$ is specified as $V_{\max}(\boldsymbol{Z})$. We always assume that the eigenvalues are ordered in nondecreasing order such that $\mu_i(\boldsymbol{Z}) \leq \mu_{i+1}(\boldsymbol{Z})$. The orthogonal projector onto the orthogonal complement of the column space of $\boldsymbol{Z}$ is written as $\Pi_{\boldsymbol{Z}}^{\perp} = \boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}^H\boldsymbol{Z})^{-1}\boldsymbol{Z}^H$.

System Model. We consider a scenario in which K secondary transmitter-receiver pairs, also referred to as users, coexist with L primary systems as depicted in Figure 1.1. Define the set of secondary users as $\mathcal{S} \triangleq \{1, \ldots, K\}$ and that of the primary users as $\mathcal{P} \triangleq \{K+1, \ldots, K+L\}$. Each transmitter is equipped with N transmit antennas and each receiver with a single antenna. All systems share the same bandwidth such that the setting leads to a MISO IC [9].

Channel Model. The quasi-static block flat-fading vector channel from secondary transmitter k to secondary receiver $j \in \mathcal{S}$ is denoted by $\boldsymbol{h}_{kj} \in \mathbb{C}^{N\times 1}$. The channel vectors from secondary transmitter k to primary receiver $j \in \mathcal{P}$ is denoted by $\boldsymbol{z}_{kj} \in \mathbb{C}^{N\times 1}$. The beamforming vector of secondary transmitter k is represented with $\boldsymbol{w}_k \in \mathbb{C}^{N\times 1}$. Moreover, each transmitter has a total power constraint of $P \triangleq 1$. Generalization to different power constraints at each transmitter can be done without affecting the results in the next sections. This leads to the constraint: $\|\boldsymbol{w}_k\|^2 \leq 1, k \in \mathcal{S}$. The set of feasible transmit beamforming vectors of transmitter $k \in \mathcal{S}$, is defined as

$$A_k \triangleq \{\boldsymbol{w} : \|\boldsymbol{w}\|^2 \leq 1\}. \tag{1.1}$$

Noise plus interference originating from primary users is assumed independent and identically distributed (i.i.d.) complex Gaussian with zero mean and variance σ^2. We define the transmit signal-to-noise ratio (SNR) as $1/\sigma^2$. Each secondary transmitter is assumed to have perfect local CSI, i.e. it has perfect knowledge of the channel vectors only between itself and all secondary and primary receivers. The achievable rate of secondary user k is

$$R_k(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) = \mathrm{ld}\left(1 + \frac{|\boldsymbol{h}_{kk}^H \boldsymbol{w}_k|^2}{\sigma^2 + \sum_{l\in\mathcal{S}\setminus\{k\}} |\boldsymbol{h}_{lk}^H \boldsymbol{w}_l|^2}\right), \tag{1.2}$$

where the receivers are assumed to treat interference from secondary transmitters as additive noise. Thus, each transmitter needs not know the channels incident at its receiver in order to allocate the transmission rate in (1.2). The achievable rate region defined as

$$\mathcal{R} \triangleq \{(R_1, \ldots, R_K) : \boldsymbol{w}_k \in A_k, k \in \mathcal{S}\}, \tag{1.3}$$

is the set of all rate tuples achieved by feasible beamforming vectors.

Gain Region. Define the power gain achieved by transmitter k on all receivers as a function of the beamforming vector $\boldsymbol{w}_k$ as

$$x_{k,\ell}(\boldsymbol{w}_k) = |\boldsymbol{h}_{k\ell}^H \boldsymbol{w}_k|^2, \quad \ell \in \mathcal{S} \cup \mathcal{P}.$$

For user k, the achievable gain region is defined as

$$\Omega_k = \bigcup_{\|\boldsymbol{w}_k\|=1} \big(x_{k,1}(\boldsymbol{w}_k), \ldots, x_{k,K+L}(\boldsymbol{w}_k)\big).$$

Definition 1.1. A point $\boldsymbol{y} \in \mathbb{R}_+^n$ is called *upper boundary point* of a convex set $\mathcal{C}$ in direction $\boldsymbol{e}$ if $\boldsymbol{y} \in \mathcal{C}$ while the set

$$\mathcal{K}_{\boldsymbol{y}}(\boldsymbol{e}) = \{\boldsymbol{y}' \in \mathbb{R}_+^n | y'_\ell e_\ell \geq y_\ell e_\ell \; \forall\, 1 \leq \ell \leq n\} \subset \mathbb{R}_+^n \setminus \mathcal{C},$$

where the inequality has at least one strict inequality and directional vector $\boldsymbol{e} \in \{-1, +1\}^n$. We denote the set of upper boundary points in direction $\boldsymbol{e}$ as $\partial^e \mathcal{C}$.

Lemma 1.1. *The set Ω_k is compact and has a convex boundary in direction $\boldsymbol{e} \in \{-1, +1\}^N \setminus -\boldsymbol{1}$ according to Definition 1.1.*

The proof of this lemma follows from observing that the set Ω_k is the joint numerical range of the matrices $\boldsymbol{h}_{k1}\boldsymbol{h}_{k1}^H, \ldots, \boldsymbol{h}_{k(K+L)}\boldsymbol{h}_{k(K+L)}^H$. This set is not necessarily convex, however in [10] we proved that this set has a convex boundary in the specified directions. Next, we define the upper boundary of a set in direction $\boldsymbol{e}$.

1.1.1 Soft Constraints Through Pricing

A pricing scheme can limit the secondary users from choosing the strategies that induce increased costs. By defining the utility function containing a penalty term that corresponds to the amount of interference the secondary users produce on the primary users, the ITCs can be applied. For user k, we define its utility function as its achievable rate in (1.2) minus the sum of weighted interference rates this user produces at the primary users, i.e.

$$u_k(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) = R_k - \sum_{\ell \in \mathcal{P}} \eta_{k,\ell} \mathrm{ld}\left(1 + |\boldsymbol{h}_{k\ell}^H \boldsymbol{w}_k|^2\right), \tag{1.4}$$

where $\boldsymbol{\eta}_k \in \mathbb{R}_+^L$ is a weight vector that regulates the penalty respective to each primary user. Similar formulation of the pricing mechanism to reflect the interference power on primary users is given in [8]. In the latter work, the authors consider noncooperative or distributed operation of the secondary users in a cognitive radio setting with ITCs. The weights are along with the strategies of the transmitters iteratively determined such that the outcome corresponding to the NE satisfies the ITCs. In our case, we assume the weight vector $\boldsymbol{\eta}_k, k \in \mathcal{S}$ is fixed. There is another operational meaning of the utility function in (1.4): It corresponds to the weighted sum of mutual information from the k-th secondary transmitter to all primary users. Hence, it could be interpreted as the amount of information by link k eavesdropped at the primary receivers.

Given the utility function in (1.4), the utility region is the set of all utility tuples that can be achieved using beamforming vectors that satisfy the transmit power constraint, i.e.

$$\mathcal{U} \triangleq \bigcup_{\substack{\boldsymbol{w}_k : \|\boldsymbol{w}_k\| \leq 1 \\ k \in \mathcal{S}}} \{u_1(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K), \ldots, u_K(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)\}.$$

Definition 1.2. A utility tuple $(u_1, \ldots, u_K)$ is Pareto optimal if there is no other tuple $(q_1, \ldots, q_K)$ with $(q_1, \ldots, q_K) \geq (u_1, \ldots, u_K)$ and $(q_1, \ldots, q_K) \neq (u_1, \ldots, u_K)$ (the inequality is componentwise).

The Pareto boundary, denoted by $\mathcal{PB}$, is the set of all Pareto-optimal points.

Theorem 1.1. *All beamforming vectors that achieve points on the Pareto boundary of the utility region* $\mathcal{U}$ *can be written as*

$$\boldsymbol{w}_k = V_{\max}\left(\sum_{\ell=1}^{K+L} \lambda_{k,\ell} e_{k,\ell} \boldsymbol{h}_{k\ell} \boldsymbol{h}_{k\ell}^H\right), \tag{1.5}$$

where

$$e_{k,\ell} = \begin{cases} +1 & \ell = k \\ -1 & \textit{otherwise} \end{cases}, \tag{1.6}$$

and $k \in \mathcal{S}$ *and* $\boldsymbol{\lambda}_k \in \Lambda_{K+L}$. *The set* Λ_K *is defined as*

$$\boldsymbol{\lambda} \in \Lambda_K = \left\{\boldsymbol{\lambda} \in [0,1]^K : \sum_{\ell=1}^{K} \lambda_\ell = 1\right\}.$$

Note that the converse is not true. Not all beamforming vectors defined by (1.5) lie on the Pareto boundary, i.e. all beamforming vectors that achieve utility tuples on the Pareto boundary can be represented as $\boldsymbol{w}_k$ in (1.5) but not all $\boldsymbol{w}_k$ in (1.5) correspond to points on the Pareto boundary. Moreover, in the case that $N < K + L$, full power transmission does not achieve all Pareto-optimal points, hence the transmit powers of the beamforming vectors determined in Theorem 1.1 have to be varied between full and zero power allocation. In [10], it is shown that power control is only needed for specific beamforming vectors in the set determined in Theorem 1.1.

Proof. In [11, Theorem 2], it is shown for the case when no constraints exist that the beamforming vectors that achieve Pareto-optimal points are within the beamforming vectors that achieve the upper boundary of the gain region in the specified direction $\boldsymbol{e}$. We prove by contradiction that this result also holds for our case. Assume that there exist a beamforming vector $\boldsymbol{w}_k$ for user k which is not at the upper boundary of the kth user gain region in direction $\boldsymbol{e}_k$, i.e.

$$\boldsymbol{x}_k(\boldsymbol{w}_k) \notin \partial^{\boldsymbol{e}_k} \Omega_k, \tag{1.7}$$

and achieves a point on the Pareto boundary of $\mathcal{U}$, i.e.

$$(u_1(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K), \ldots, u_K(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)) \in \mathcal{PB}. \tag{1.8}$$

By assuming (1.7), we can find another beamforming vector $\boldsymbol{v}_k$ with $\|\boldsymbol{v}_k\| = 1$ and the property

$$x_{k,\ell}(\boldsymbol{w}_k)e_{k,\ell} \leq x_{k,\ell}(\boldsymbol{v}_k)e_{k,l}, \quad \text{for all } \ell \in \mathcal{S} \cup \mathcal{P}. \tag{1.9}$$

Next we distinguish two case for the inequality in (1.9). The first case is when $\ell = k$. In this case, user k can increase the gain in the direction of his receiver. The second case is when $\ell \neq k$ corresponding to the case when user k can reduce the gain on the other receivers.

1. Assume the inequality is strict for $\ell = k$ with $e_{k,k} = 1$ as given in (1.6), then $x_{k,k}(\boldsymbol{w}_k) < x_{k,k}(\boldsymbol{v}_k)$. The gains to all other receivers are to stay unchanged such that $x_{k,\ell}(\boldsymbol{w}_k) = x_{k,\ell}(\boldsymbol{v}_k)$ for $\ell \neq k$. Then, $u_k(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) < u_k(\boldsymbol{w}_1, \ldots, \boldsymbol{v}_k, \ldots, \boldsymbol{w}_K)$ holds because of the following. In (1.2), the rate for user k changes when using $\boldsymbol{v}_k$ as $x_{k,k}(\boldsymbol{w}_k) < x_{k,k}(\boldsymbol{v}_k)$ and hence

$$\mathrm{ld}\left(1 + \frac{x_{k,k}(\boldsymbol{w}_k)}{\sigma^2 + \sum\limits_{l \in \mathcal{S}\setminus\{k\}} x_{\ell,k}(\boldsymbol{w}_\ell)}\right) < \mathrm{ld}\left(1 + \frac{x_{k,k}(\boldsymbol{v}_k)}{\sigma^2 + \sum\limits_{l \in \mathcal{S}\setminus\{k\}} x_{\ell,k}(\boldsymbol{w}_\ell)}\right),$$

   i.e. $R_k(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) < R_k(\boldsymbol{w}_1, \ldots, \boldsymbol{v}_k, \ldots, \boldsymbol{w}_K)$, which leads to a strict increase in the kth user utility given in (1.4), since the penalty term remains unchanged. This result contradicts (1.8).
2. Assume the inequality is strict for $j \neq k$ with $e_{k,j} = -1$, i.e. $x_{k,j}(\boldsymbol{w}_k) > x_{k,j}(\boldsymbol{v}_k)$. The gains to all other receivers stay the same such that $x_{k,\ell}(\boldsymbol{w}_k) = x_{k,\ell}(\boldsymbol{v}_k)$ for $\ell \neq j$. Then, $u_j(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) < u_j(\boldsymbol{w}_1, \ldots, \boldsymbol{v}_k, \ldots, \boldsymbol{w}_K)$ holds since the following. We distinguish two cases:
   a) If $j \in \mathcal{S}$, the rate for user j changes as $x_{k,j}(\boldsymbol{v}_k) < x_{k,j}(\boldsymbol{w}_k)$ and hence

$$\begin{aligned} &\mathrm{ld}\left(1 + \frac{x_{j,j}(\boldsymbol{w}_j)}{\sigma^2 + x_{k,j}(\boldsymbol{w}_k) + \sum\limits_{l \in \mathcal{S}\setminus\{k\}} x_{\ell,j}(\boldsymbol{w}_\ell)}\right) \\ &< \mathrm{ld}\left(1 + \frac{x_{j,j}(\boldsymbol{w}_j)}{\sigma^2 + x_{k,j}(\boldsymbol{v}_k) + \sum\limits_{l \in \mathcal{S}\setminus\{k\}} x_{\ell,j}(\boldsymbol{w}_\ell)}\right), \end{aligned}$$

   i.e. $R_k(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K) < R_k(\boldsymbol{w}_1, \ldots, \boldsymbol{v}_k, \ldots, \boldsymbol{w}_K)$. Hence, the utility of the jth user is strictly increased which contradicts assumption (1.8).
   b) If $j \in \mathcal{P}$, the rate for user k does not change, however his penalty term strictly decreases as $x_{k,j}(\boldsymbol{w}_k) > x_{k,j}(\boldsymbol{v}_k)$ and hence

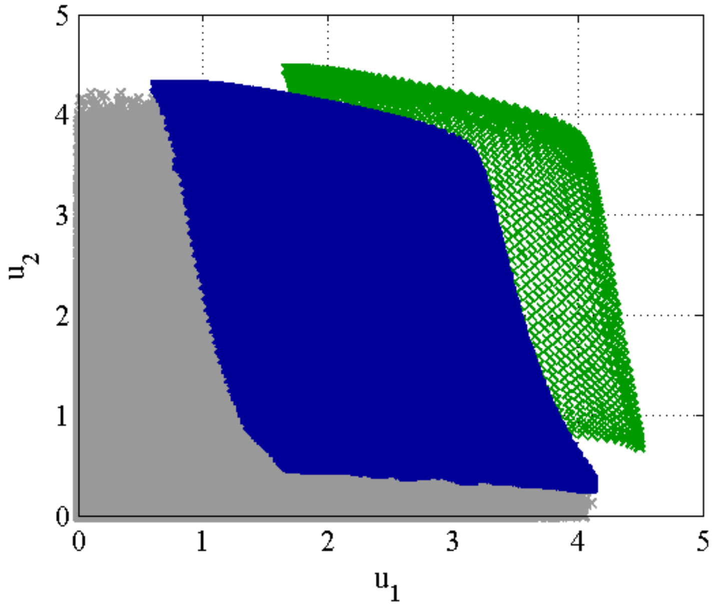

Figure 1.2: In green is the Pareto boundary without constraints for the transmission of the secondary users. Blue contains the Pareto boundary with pricing. Grey points are within the utility region under the same constraints. In this setting, three transmit antennas are used at the transmitters and the SNR defined as $1/\sigma^2$ is 10 dB.

$$\begin{aligned}
&\eta_{k,j}\mathrm{ld}\big(1 + x_{k,j}(\boldsymbol{w}_k)\big) + \sum_{\ell \in \mathcal{P}\setminus\{j\}} \eta_{k,\ell}\mathrm{ld}\big(1 + x_{k,\ell}(\boldsymbol{w}_k)\big) \\
&> \eta_{k,j}\mathrm{ld}\big(1 + x_{k,j}(\boldsymbol{v}_k)\big) + \sum_{\ell \in \mathcal{P}\setminus\{j\}} \eta_{k,\ell}\mathrm{ld}\big(1 + x_{k,\ell}(\boldsymbol{w}_k)\big).
\end{aligned}$$

which implies that the utility of user k increases contradicting (1.8).

Above, we have proven that a point which is not on the boundary of the gain region in the direction $\boldsymbol{e}_k$ given in (1.6) does not achieve Pareto-optimal points. In other words, any point in the set $\mathcal{PB}$ has a corresponding beamforming vector that achieves gains on the boundary of the gain region. □

In Theorem 1.1, the Pareto boundary is characterized for the utility region of the secondary users where fixed prices weights are given. This characterization does not hold for the rate region under soft- or peak-power-shaping constraints. Under such conditions, each Pareto rate tuple is achieved under different prices weights. Acquiring these weights requires an iterative process [6].

A plot of the utility region for two secondary users coexisting with a single primary user is given in Figure 1.2. Under no interference constraints, i.e. zero prices weights, the setting is equivalent to one without the primary user. The Pareto boundary in this case is characterized requiring the summation in (1.5) over the secondary users only, i.e. the same characterization as in [11, Theorem 2]. The corresponding Pareto boundary is plotted in green. For the case that the prices weights are fixed such that $\eta_{1,3} = \eta_{2,3} = 0.5$, the utility region, corresponding to two million randomly generated beamforming vectors that satisfy the transmit power constraint, is represented in grey. This grey region hence illustrates the achievable utility region $\mathcal{U}$. The utilities achieved through beamforming vectors from Theorem 1.1 for the same prices weights are plotted in blue. These beamforming vectors are shown to lead to Pareto-optimal utility tuples of the utility region represented in grey.

1.1.2 Null-Shaping Constraints and Greedy User Selection

We now assume that the primary systems do not tolerate any interference from the secondary systems, i.e. the ITCs are fixed to be zero. In this case, the constraints are called null-shaping constraints [3]. The outline and main results of this section are as follows: First, given a set of existing primary users, the Pareto boundary of the secondary users' achievable MISO IC rate region is characterized under the null-shaping constraints. The Pareto-optimal strategies can be performed by the secondary users if they cooperate. Then, motivated by distributed (noncooperative) operation of the secondary systems, we turn our interest to the design of null-shaping constraints that improve the efficiency of the noncooperative systems. We characterize the null-shaping constraints, corresponding to virtual primary users, such that all points on the Pareto boundary of the MISO IC rate region without constraints are achieved by the noncooperative secondary systems. This result shows that imposing null-shaping constraints on the secondary users can be sufficient to improve their noncooperative outcome. Last, following the previous result, we investigate the noncooperative secondary user selection problem to increase the achievable sum rate. The selection should activate only a subset of the existing secondary systems for operation. Assuming that the secondary users are noncooperative and null-shaping constraints corresponding to existing primary users exist, we provide a secondary user selection algorithm that improves the sum performance of the systems. The algorithm is greedy such that in each iteration step, activating a secondary system has to increase the sum rate of the selected secondary systems set.

Pareto Boundary for Given Constraints

In this section, we characterize the beamforming vectors that achieve Pareto-optimal points under null-shaping constraints. In other words, given the channels to primary users, the secondary transmitters are to form a null in the direction of these channels, i.e.

$$\left|\boldsymbol{z}_{k\ell}^H \boldsymbol{w}_k\right| = 0, \quad \text{for all } k \in \mathcal{S}, \ell \in \mathcal{P}. \tag{1.10}$$

Hence, the set of feasible transmission strategies for secondary transmitter $k \in \mathcal{S}$ is

$$\widetilde{A}_k \triangleq \{\boldsymbol{w} : \|\boldsymbol{w}\|^2 \leq 1, \left|\boldsymbol{z}_{k\ell}^H \boldsymbol{w}\right| = 0 \text{ for all } \ell \in \mathcal{P}\}. \tag{1.11}$$

The achievable rate region under null-shaping constraint is defined as

$$\widetilde{\mathcal{R}} \triangleq \Big\{(R_1,\dots,R_K) : \boldsymbol{w}_k \in \widetilde{A}_k, k \in \mathcal{S}\Big\}. \tag{1.12}$$

In [12], null-shaping constraints on secondary users are considered in a noncooperative MIMO cognitive radio game. The orthogonal projector onto the null space of the primary user's channel is used to transform the secondary user's rate maximization problem to one that is convenient to study. Next, we use a similar application of this projection. Define the matrix containing the channels from secondary transmitter k to all primary users as $\boldsymbol{Z}_k \triangleq [\boldsymbol{z}_{k1},\dots,\boldsymbol{z}_{kL}]$.

Proposition 1.1. ([13, Corollary 1]) Assume $N \geq K + L$. All beamforming vectors which fulfill the null-shaping constraints in (1.10) and achieve the Pareto boundary of $\widetilde{\mathcal{R}}$ in (1.12) are

$$\boldsymbol{w}_k(\boldsymbol{\lambda_k}) = V_{\text{max}}\left(\sum_{\ell=1}^{K} \lambda_{k,\ell} e_{k,\ell} \Pi^{\perp}_{\boldsymbol{Z}_k} \boldsymbol{h}_{k\ell}\boldsymbol{h}_{k\ell}^H \Pi^{\perp}_{\boldsymbol{Z}_k}\right), \quad k \in \mathcal{S}, \tag{1.13}$$

where

$$e_{k,\ell} = \begin{cases} +1 & \ell = k \\ -1 & \text{otherwise} \end{cases}, \tag{1.14}$$

and $\boldsymbol{\lambda}_k \in \Lambda_K$. The set Λ_K is defined as

$$\Lambda_K \triangleq \left\{\boldsymbol{\lambda} \in [0,1]^K : \sum_{\ell=1}^{K} \lambda_\ell = 1\right\}. \tag{1.15}$$

We assume $N \geq K + L$ in Proposition 1.1 for two reasons: First, the constraints in (1.10) can be satisfied by an active secondary transmitter when $N \geq L$. Second, for $N \geq K$ full power transmission is optimal to achieve Pareto-optimal points [10, Section III.A]. For $N < K$, a transmitter has to vary its transmission power for specific beamforming vectors in order to achieve Pareto-optimal operating points [10, Section III.B].

In Figure 1.3, we show a comparison of the rate regions with and without null-shaping constraints. These regions correspond to $\mathcal{R}$ and $\widetilde{\mathcal{R}}$ defined in (1.3) and (1.12), respectively. The setting has two secondary transmitters equipped with three antennas and a single primary user. The Pareto boundary of $\mathcal{R}$ is marked with crosses which correspond to beamforming vectors characterized in [11, Theorem 2].

For the case in which null-shaping constraints are imposed on the secondary users, the Pareto boundary of $\widetilde{\mathcal{R}}$ is achieved by beamforming vectors characterized in Proposition 1.1. By varying the parameters λ_1 and λ_2 in (1.13) between zero and one in a 0.02 step-length, the Pareto boundary of $\widetilde{\mathcal{R}}$ is plotted in Figure 1.3 with circle markers. The region $\widetilde{\mathcal{R}}$ is

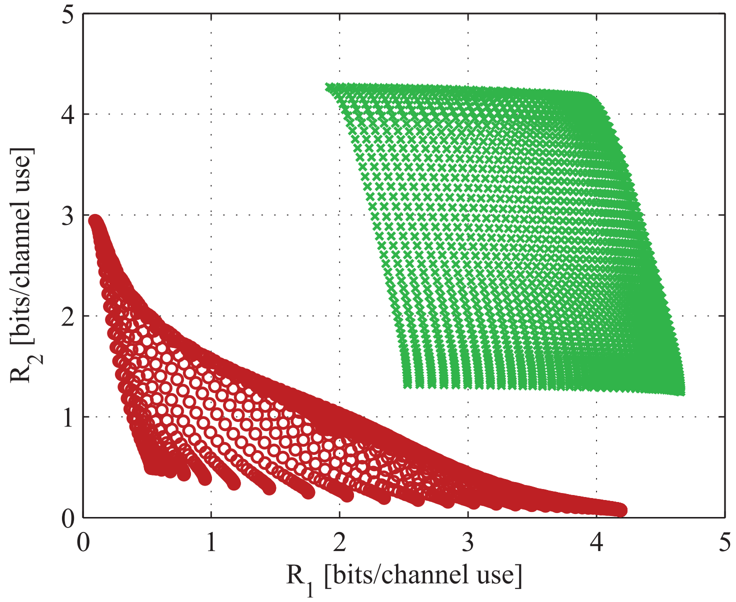

Figure 1.3: The Pareto boundary of $\mathcal{R}$ in (1.3) is marked with crosses. The Pareto boundary of $\widetilde{\mathcal{R}}$ in (1.12) is marked with circles. The plots are made for single random channel realization with $N = 3$ and SNR = 10 dB.

certainly always smaller and within the region $\mathcal{R}$, because the feasible transmission strategies set is smaller, $\widetilde{A}_k \subseteq A_k$ for all $k \in \mathcal{S}$. Notice that each secondary transmitter uses three antennas. Due to the null-shaping constraint, one spacial transmit dimension is reduced for each secondary transmitter such that two spacial dimensions are left available to operate in.

Given certain null-shaping constraints, beamforming vectors described in Proposition Proposition 1.1 achieve Pareto-optimal rate tuples for the secondary users. In the next section, we study whether through the choice of virtual null-shaping constraints the secondary users could operate on the Pareto boundary of $\mathcal{R}$.

Constraints Achieving Pareto Boundary

In this section, we investigate the design of null-shaping constraints such that the secondary users in a noncooperative game achieve the Pareto boundary of $\mathcal{R}$ defined in (1.3). We denote this game in strategic form [14, Part I] by

$$\langle \mathcal{S}, (A_k)_{k\in\mathcal{S}}, (R_k)_{k\in\mathcal{S}} \rangle, \tag{1.16}$$

where A_k is defined in (1.1) and R_k is defined in (1.2). The players of this game are assumed to be rational, i.e. seek to maximize their utility. The described game is a game of complete information such that its outcome is a NE. In a setting where null-shaping constraints on the secondary users do not exist, the unique NE strategy of each system is maximum-ratio combining [15]

$$\boldsymbol{w}_k^{\mathrm{NE}} = \frac{\boldsymbol{h}_{kk}}{\|\boldsymbol{h}_{kk}\|}, \quad k \in \mathcal{S}.$$

The rate tuple corresponding to the NE strategy profile is not necessarily Pareto optimal [15]. Of interest is to design constraints on the noncooperative systems such that NE rate tuples lie on the Pareto boundary of the rate region achieved with no constraints.

Proposition 1.2. Assume $N \geq K$. Define the matrix

$$\widetilde{\boldsymbol{Z}}_k(\boldsymbol{\lambda}_k) = \left[\tilde{\boldsymbol{z}}_1(\boldsymbol{\lambda}_k), \ldots, \tilde{\boldsymbol{z}}_{K-1}(\boldsymbol{\lambda}_k)\right], \tag{1.17}$$

where

$$\tilde{\boldsymbol{z}}_i(\boldsymbol{\lambda}_k) = V_i\left(\sum_{\ell=1}^{K} \lambda_{k,\ell} e_{k,\ell} \boldsymbol{h}_{k\ell} \boldsymbol{h}_{k\ell}^H\right),$$

with $\boldsymbol{\lambda}_k \in \Lambda_K$ defined in (1.15) and $e_{k,\ell}$ defined in (1.14). All points on the Pareto boundary of the rate region $\mathcal{R}$ defined in (1.3) can be reached by the beamforming vectors

$$\boldsymbol{w}_k(\boldsymbol{\lambda}) = \frac{\Pi^{\perp}_{\widetilde{\boldsymbol{Z}}_k(\boldsymbol{\lambda}_k)} \boldsymbol{h}_{kk}}{\left\|\Pi^{\perp}_{\widetilde{\boldsymbol{Z}}_k(\boldsymbol{\lambda}_k)} \boldsymbol{h}_{kk}\right\|}. \tag{1.18}$$

This result follows directly from [10, Corollary 1] and generalizes the result in [13, Proposition 1] where only the two-user case has been considered. Note that the beamforming vector in (1.18) is the unique NE for transmitter k which abides by the null-shaping constraints. The noncooperative game in (1.16) is only between the secondary users. The null-shaping constraints in (1.17) are imposed on the secondary users by an authority which is not included as a player in the game. This authority in game theoretic terms is represented by an arbitrator [16].

Interestingly, the null-shaping constraints are sufficient to characterize the Pareto boundary of the rate region $\mathcal{R}$ without constraints. Thus, Pareto-optimal points can be obtained in with noncooperative strategies described in (1.18). This result is related to the result in [6] in that both methods achieve the same Pareto boundary in NE, and both methods utilize interference constraints on the transmitters. However, ITCs for each receiver are considered in [6], and we consider null-shaping constraints that correspond to $K-1$ virtual primary receivers. In [6], each point on the boundary is determined iteratively while here we provide the constraints and the corresponding strategies in closed form.

Next, we continue to consider the noncooperative operation of the secondary users. However, the null-shaping constraints imposed on these users correspond to the existing primary users as described in (1.10). Our objective is to increase the sum rate for the noncooperative secondary users by secondary user subset-selection.

Result: Greedy User Selection
Input: set of Secondary Users, $\mathcal{K}$
initialize $\mathcal{U} = \{\}$, $\mathcal{T} = \mathcal{K}$, $n = 0$, $SR(\{\}) = 0$;
while $n \leq |\mathcal{K}|$ **do**
    $n = n + 1$;
    $k_n = \arg\max_{u \in \mathcal{T}} SR(\{u\} \cup \mathcal{U})$;
    if $SR(\mathcal{U}) > SR(\{k_n\} \cup \mathcal{U})$ **then**
        break;
    else
        $\mathcal{T} = \mathcal{T} \backslash \{k_n\}$;
        $\mathcal{U} = \mathcal{U} \cup \{k_n\}$;
    end
end
Output: Set of selected users, $\mathcal{U}$

Algorithm 1: Greedy User Selection.

Greedy Noncooperative Users Selection

In this section, each transmitter chooses its NE transmission strategy, such that the null-shaping constraints in (1.10) are satisfied. The noncooperative game is described by

$$\langle \mathcal{S}, (\widetilde{A}_k)_{k \in \mathcal{S}}, \widetilde{\mathcal{R}} \rangle, \tag{1.19}$$

where $\widetilde{A}_k$ is defined in (1.11) and $\widetilde{\mathcal{R}}$ is defined in (1.12). The unique NE strategy of a secondary transmitter is

$$\boldsymbol{w}_k^{\text{NE}} = \frac{\Pi_{\boldsymbol{Z}_k}^{\perp} \boldsymbol{h}_{kk}}{\left\| \Pi_{\boldsymbol{Z}_k}^{\perp} \boldsymbol{h}_{kk} \right\|}, \quad k \in \mathcal{S}, \tag{1.20}$$

where $\boldsymbol{Z}_k \triangleq [\boldsymbol{z}_{k1}, \ldots, \boldsymbol{z}_{kL}]$. We address the issue of selecting a subset of the users $\mathcal{U} \subseteq \mathcal{S}$ to operate in a noncooperative game. This scheme is motivated by the fact that when interference dominates in the network, turning some links off could increase the achievable sum rate of the remaining systems through interference reduction. Note that the remaining users $\mathcal{S} \backslash \mathcal{U}$ are to be inactive when not selected. The selection mechanism could be done by an authority or an arbitrator. For a set of secondary users $\mathcal{U}$, $\mathcal{U} \subseteq \mathcal{K}$, the achievable sum rate is written as

$$SR(\mathcal{U}) = \sum_{k \in \mathcal{U}} \log_2 \left(1 + \frac{|\boldsymbol{h}_{kk}^H \boldsymbol{w}_k^{\text{NE}}|^2}{\sigma^2 + \sum\limits_{\ell \in \mathcal{U} \backslash \{k\}} |\boldsymbol{h}_{\ell k}^H \boldsymbol{w}_\ell^{\text{NE}}|^2} \right),$$

where $\boldsymbol{w}_k^{\text{NE}}, k \in \mathcal{U}$, is given in (1.20). Finding the optimal user selection has very high complexity and could be done by an exhaustive search over all possible combinations

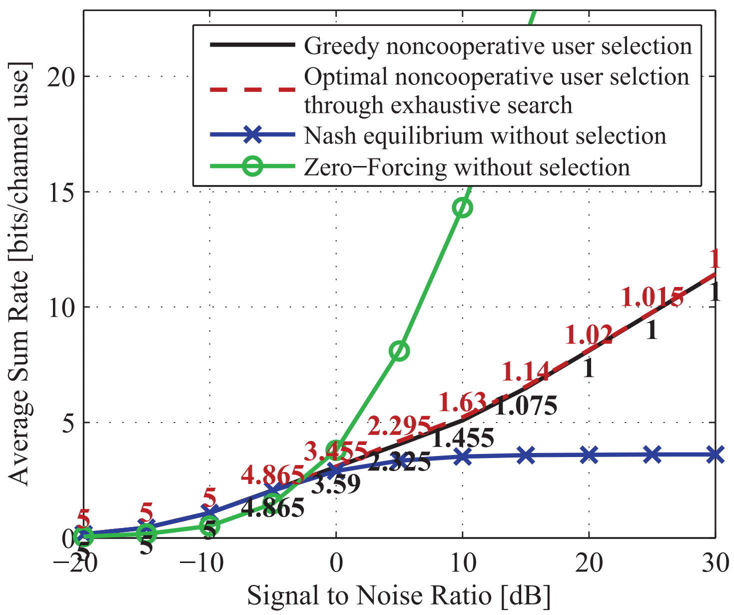

Figure 1.4: Achievable average sum rates for 5 secondary users with 6 transmit antennas and a single primary user.

of the user subsets. We propose a low complexity suboptimal user selection algorithm, described in Algorithm 1. This algorithm is influenced by the algorithm proposed in [17] for the user selection in a downlink scenario with zero-forcing beamforming at the transmitter. See [17] for details on the complexity of the algorithm. Starting with an empty secondary user set, in each iteration step of Algorithm 1, a secondary user that contributes highest sum rates to the set is added. The algorithm terminates when the sum rate decreases with an additional user or when all users have been selected. The number of selected users and their performance depends on the SNR.

In Figure 1.4, the achievable average sum rate of the secondary systems in different operation schemes is compared for increasing SNR. The simulations are performed for a single existing primary user and 5 secondary users with each transmitter using 6 transmit antennas. The average sum rate is taken over 200 random samples of each channel. The performance of Algorithm 1 is given under Greedy noncooperative user selection, and the average number of selected users is written under the curve. The optimal noncooperative user selection is obtained through exhaustive search, and the corresponding performance is plotted with dashed line. The average number of supported users for this scheme is written over the curve. The exhaustive search algorithm searches between K factorial user combinations. Note that both algorithms select a subset of the secondary users that are noncooperative. The performance of the suboptimal algorithm is very close to the optimal one. The gap between the optimal and suboptimal algorithm curves is largest in the intermediate SNR regime between 0 and 20 dB. Moreover, it is noticed that the gap increases as the number of transmit antennas increases. In the low SNR regime, the number of supported users is highest and decreases for increasing SNR values. In the high SNR

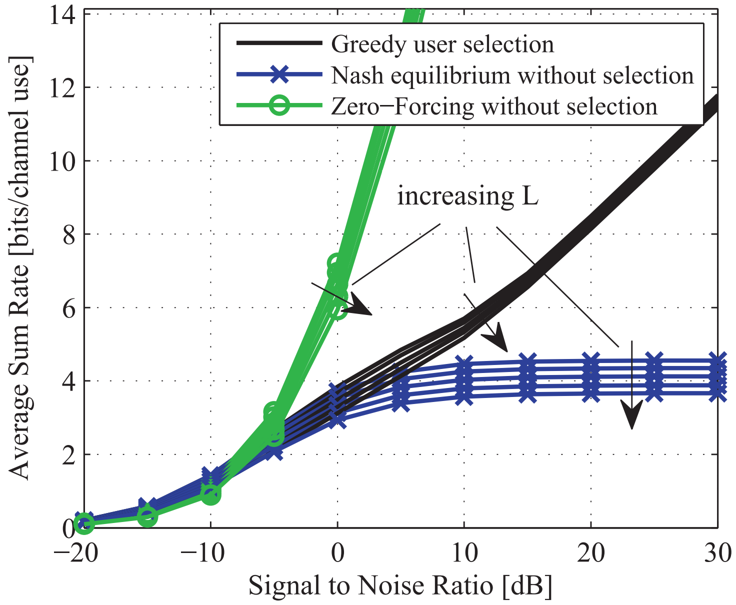

Figure 1.5: Achievable average sum rates for 5 secondary users with 10 transmit antennas. Increasing the number of primary users L reduces the average sum rate of the secondary users.

regime only a single transmitter is eventually selected. Similar curves are given in [18], where opportunistic beamforming in the MISO broadcast channel is analyzed. There, for small SNR, all spacial dimensions are exploited and users are scheduled, whereas for high SNR, only one user on one beam is scheduled due to interference.

The curve with cross markers in Figure 1.4 describes the performance when all noncooperative secondary transmitters are always chosen. This operation scheme has good performance in the low SNR regime. At high SNR however, the average sum rate saturates and approaches a constant value due to multiuser interference. In selecting fewer users to operate as in Algorithm 1, better performance is achieved. The curve with circle markers describes the average sum rate achieved when all secondary users perform zero-forcing transmission strategies. This transmission strategy performs best in the high SNR regime and achieves maximum multiplexing gain. A single transmitter is capable of performing this strategy if its number of transmit antennas is greater or equal to the number of primary receivers plus the number of secondary receivers. This transmission scheme is not denoted as optimal because it is not achieved in the game described for this section in (1.19).

In Figure 1.5, the achievable average sum rate of the secondary systems is compared for increasing SNR and increasing number of existing primary users. Again, the average sum rate is taken over 200 random channel vector samples. The number of primary users is varied between 1 and 5 and the number of transmit antennas is 10. Including additional primary users, the average sum rate of the secondary systems always decreases. It is observed, however, that for a single channel sample the performance of the systems could increase with increasing the number of primary users. This observation supports the result given in Proposition 1.2 that null shaping constraints could improve the performance of the noncooperative systems.

1.2 Supermodular Games for Systems with Protected and Shared Bands

In this section, two non-cooperating cells are considered, each with its protected band to provide service to its high priority users. A shared band for the two cells is employed to deliver service to low priority users. We formulate the situation between the two cells as a non-cooperative game and study its NE. We prove that the game belongs to a class of games called supermodular games which have several interesting properties, such as the global stability of a unique NE. We provide a sufficient condition for the existence of a unique NE and study its efficiency by extensive simulations.

The ever increasing demands on the wireless data networks has pushed more advanced technologies to be integrated into commercial systems, where multiple-input multiple-output (MIMO) is already a reality with its inclusion in several commercial standards [19]. The MIMO technology enables several techniques for communications, where the manufacturer/operator can choose the most interesting option on the basis of the scenario and requirements. One of the most attractive options is the zero forcing beamforming (ZFB) [20] that guarantees no interference between the serviced users, showing interesting performance and thus, motivating its inclusion in several commercial standards.

We consider a setup which includes two base stations (BS). Each BS has a *protected band* which it uses exclusively, and also a band which is shared with the other BS, named the *shared band*. This setup has been considered in [21, 22] with single antennas at the transmitters and receivers. In [21], the considered noncooperative game is shown to be a supermodular game. Necessary and sufficient conditions for the uniqueness of NE are provided. In [22], the manipulability of the NE is studied, and a cheat proof mechanism is provided to suppress untruthful feedback from the mobile users to the BSs. In this work, MIMO is applied over the protected bands through the ZFB strategy, so that each BS will guarantee zero interference to its customers, while the two BSs will interfere on each other in the shared band.

In order to avoid the large amount of overhead that cooperation induces, we assume the two BSs are non-cooperative. We formulate the situation between the two BSs as a non-cooperative game in which the objective of each BS is to maximize the system data rate. The outcome of the non-cooperative game is a NE [14]. We prove the existence of an NE and provide sufficient conditions for its uniqueness. By proving that the non-cooperative game is a supermodular game [23], the uniqueness of NE implies the global convergence of best response dynamics (iterative waterfilling). The efficiency of the NE is compared to the cooperative maximum sum rate by extensive simulations. It is shown that the setup of protected and shared bands provides significant efficiency gains to the non-cooperative systems.

Notations

Column vectors and matrices are given in lowercase and uppercase boldface letters, respectively. $\|\boldsymbol{a}\|$ is the Euclidean norm of $\boldsymbol{a} \in \mathbb{C}^N$. $|b|$ denotes the absolute value of $b \in \mathbb{C}$. $(\cdot)^H$ denotes Hermitian transpose. $[\boldsymbol{A}]_{i,j}$ is the entry in $\boldsymbol{A}$ corresponding to the ith row and jth column. Define $[x]_a^b := \max\{\min\{x, b\}, a\}$. ld is logarithm base two.

1.2.1 System and Channel Model

A wireless multi-antenna downlink system is considered where two BSs are active: BS_1 and BS_2. The BSs are equipped with multiple antennas while several users are in the scenario, each one with a single antenna. The total available spectrum is divided in three different bands. The first and second bands are exclusively granted to BS_1 and BS_2, respectively; while the third band is shared between the two BSs. Therefore, BS_1 has two sub-bands to operate: a protected band where it can provide service without any interference from the other BS, so that it will be allocated to high priority users (even running demanding applications and/or customers paying high price for an “excellent" communication). The second band is a shared band where serviced users will suffer uncontrolled interference from the neighbor BS, so that low priority users will be assigned to this sub-band.

We assume that each BS is equipped with two antennas. Thus, a BS can serve two users in the protected band simultaneously. The shared band, we assume that a single user is served by each BS. The channel vector in the protected band from BS_i to user j is denoted by $\boldsymbol{h}_{ij}$. In the shared band, the channel vector from BS_i to user k is $\tilde{\boldsymbol{h}}_{ik}$. The channel characterization within each of the sub-bands follows a quasi static block fading model, where the channel from each transmit antenna to a user is characterized by independent and identically distributed (i.i.d.) complex Gaussian entries $\sim \mathcal{CN}(0, 1)$.

The availability of multiple antennas at the transmitter side enables the application of MIMO processing at the transmitter side (i.e., precoding). The system is assumed to operate time division duplexing (TDD), so that the channel information is available at the transmitter, by reciprocity with the uplink channel, before the transmission starts. Several options are presented in literature for the precoding at the transmitter side in MIMO, where this section will consider the zero forcing beamforming (ZFB) [20] strategy due to its high performance and low complexity.

Define the matrix $\boldsymbol{H}_i = [\boldsymbol{h}_{i1}, \boldsymbol{h}_{i2}]$. Its pseudo inverse is

$$\boldsymbol{W}_i = \boldsymbol{H}_i^H (\boldsymbol{H}_i \boldsymbol{H}_i^H)^{-1} = [\boldsymbol{w}_{i1}, \boldsymbol{w}_{i2}]^H, \tag{1.21}$$

with

$$\|\boldsymbol{w}_{ij}\|^2 = g_{ij} := [(\boldsymbol{H}_i \boldsymbol{H}_i^H)^{-1}]_{j,j} \tag{1.22}$$

where the vectors $\boldsymbol{w}_{ij}$ are orthogonal. Using ZFB in the dedicated band, the interference free signal at a receiver j is written as

$$y_{ij} = \sqrt{\frac{p_{ij}}{g_{ij}}}\boldsymbol{h}_{ij}^H \boldsymbol{w}_{ij} s_{ij} + n_j, \tag{1.23}$$

where $s_{ij} \sim \mathcal{CN}(0,1)$ is the symbol transmitted by BS_i to receiver j, p_{ij} is the power allocation to user j, and n_j is the noise term with $n_j \sim \mathcal{CN}(0,\sigma^2)$.

In the shared band, the received signal at user k served by BS_k is

$$\tilde{y}_k = \sqrt{\tilde{p}_k}\tilde{\boldsymbol{h}}_{kk}^H \tilde{\boldsymbol{w}}_k \tilde{s}_k + \sqrt{\tilde{p}_\ell}\tilde{\boldsymbol{h}}_{\ell k}^H \tilde{\boldsymbol{w}}_\ell \tilde{s}_\ell + n_k, \quad k \neq \ell, \tag{1.24}$$

where $\tilde{s}_k \sim \mathcal{CN}(0,1)$ is the symbol transmitted by BS_k in the shared band and the beamforming vector used at BS_k is $\tilde{\boldsymbol{w}}_k \in \mathcal{W}$ with

$$\mathcal{W} := \{\tilde{\boldsymbol{w}} \in \mathbb{C}^2 : \|\tilde{\boldsymbol{w}}\|^2 \leq 1\}. \tag{1.25}$$

Each BS i has a total power constraint P_i such that $(p_{i1}, p_{i2}, \tilde{p}_i) \in \mathcal{P}_i$ with

$$\mathcal{P}_i := \{(p_1, p_2, p_3) \in \mathbb{R}_+^3 : p_1 + p_2 + p_3 \leq P_i\}. \tag{1.26}$$

As the beamforming within the protected band is known to be through ZBF, then the transmission strategy of a BS is a choice of beamforming vector in the shared band and the power allocation to its three users. Next, we will analyze the transmission strategies of the BSs assuming their distributed operation, i.e., a BS cannot cooperate with the other BS. The setting is described by a strategic game and the outcome, NE, determines the noncooperative operation of the BSs.

1.2.2 Strategic Game Formulation

A strategic game [14] between the two BSs is defined as

$$G = \langle \{1,2\}, (\mathcal{A}_1, \mathcal{A}_2), (R_1, R_2) \rangle. \tag{1.27}$$

where $\{1,2\}$ is the set of players corresponding to the two BSs. The strategy set $\mathcal{A}_i$ of a player $i \in \{1,2\}$ is defined as

$$(p_{i1}, p_{i2}, \tilde{p}_i, \tilde{\boldsymbol{w}}_i) \in \mathcal{A}_i := \mathcal{P}_i \times \mathcal{W}. \tag{1.28}$$

where $\mathcal{P}_i$ is defined in (1.26) and $\mathcal{W}$ is defined in (1.25). The utility of player 1 (analogously player 2) is the achievable sum rate

$$R_1(p_{11}, p_{12}, \tilde{p}_1, \tilde{\boldsymbol{w}}_1, \tilde{p}_2, \tilde{\boldsymbol{w}}_2) = \sum_{k=1}^{2} \mathrm{ld}\left(1 + \frac{p_{1k}}{g_{1k}\sigma^2}\right) + \mathrm{ld}\left(1 + \frac{\tilde{p}_1 |\tilde{\boldsymbol{h}}_{11}^H \tilde{\boldsymbol{w}}_1|^2}{\sigma^2 + \tilde{p}_2 |\tilde{\boldsymbol{h}}_{21}^H \tilde{\boldsymbol{w}}_2|^2}\right), \tag{1.29}$$

where single-user decoding is assumed in the shared band.

The outcome of a strategic game is a NE. A NE of the game G in (1.27) is a strategy profile, $(\boldsymbol{a}_1^{\text{NE}}, \boldsymbol{a}_2^{\text{NE}}) \in \mathcal{A}_1 \times \mathcal{A}_2$, in which no player can increase his utility by choosing another strategy unilaterally, i.e.,

$$R_1(\boldsymbol{a}_1^{\text{NE}}, \boldsymbol{a}_2^{\text{NE}}) \geq R_1(\boldsymbol{a}_1, \boldsymbol{a}_2^{\text{NE}}), \quad \text{for all } \boldsymbol{a}_1 \in \mathcal{A}_1, \tag{1.30}$$
$$R_2(\boldsymbol{a}_1^{\text{NE}}, \boldsymbol{a}_2^{\text{NE}}) \geq R_2(\boldsymbol{a}_1^{\text{NE}}, \boldsymbol{a}_2), \quad \text{for all } \boldsymbol{a}_2 \in \mathcal{A}_2. \tag{1.31}$$

In a strategic game, each player maximizes his utility taking the strategies of the other players as given. Given a strategy $\boldsymbol{a}_2 \in \mathcal{A}_2$ of player 2, the *best response* of player 1 to $\boldsymbol{a}_2$ is the strategy $\boldsymbol{a}_1$ that maximizes his utility. The best response of player 1 and player 2 solve the following problems

$$\max_{(p_{11}, p_{12}, \tilde{p}_1, \tilde{\boldsymbol{w}}_1) \in \mathcal{A}_1} R_1(p_{11}, p_{12}, \tilde{p}_1, \tilde{\boldsymbol{w}}_1, \tilde{p}_2, \tilde{\boldsymbol{w}}_2), \tag{1.32}$$
$$\max_{(p_{21}, p_{22}, \tilde{p}_2, \tilde{\boldsymbol{w}}_2) \in \mathcal{A}_1} R_2(p_{21}, p_{22}, \tilde{p}_2, \tilde{\boldsymbol{w}}_2, \tilde{p}_1, \tilde{\boldsymbol{w}}_1), \tag{1.33}$$

respectively. From the optimization problem in (1.32), we have the following two observations. First, maximum ratio transmission (MRT), written as $\tilde{\boldsymbol{w}}_1^{\text{MRT}} = \tilde{\boldsymbol{h}}_{11} / \|\tilde{\boldsymbol{h}}_{11}\|$, is always optimal in the shared band independent of the power allocation and choice of beamforming vectors of player 2. Second, player 1 (i.e., BS_1) will choose a power allocation that will satisfy the total power constraint with equality, i.e., $p_{11} + p_{12} + \tilde{p}_1 = P_1$. The same holds for player 2.

Since a BS will allocate all available power to the users, we perform a change of variables in order to reduce the dimension of the strategy spaces of the players. Define $\pi_i \in [0, 1]$. Player 1 allocates $(1 - \pi_1)P_1$ in his protected band and $\pi_1 P_1$ in the shared band. Define $\beta_i \in [0, 1]$ and substitute the power allocation in the protected band as $p_{11} = (1 - \pi_1)P_1\beta_1$ and $p_{12} = (1 - \pi_1)P_1(1 - \beta_1)$. Analogously, player 2 allocates $p_{21} = \pi_2 P_2 \beta_2$ and $p_{22} = \pi_2 P_2 (1 - \beta_2)$ for his two users in his protected band and $(1 - \pi_2)P_2$ for the user in his shared band. The sum rate of player 1 is given on the top of the next page in (1.35) and similarly for player 2 in (1.36), where

$$\tilde{g}_{kj} := |\tilde{\boldsymbol{h}}_{kj}^H \tilde{\boldsymbol{w}}_k^{\text{MRT}}|^2, \quad k, j = 1, 2. \tag{1.34}$$

The problems in (1.32) and (1.33) for BS_1 and BS_2, are

$$\max_{\substack{\beta_1 \in [0,1] \\ \pi_1 \in [0,1]}} R_1(\beta_1, \pi_1, \pi_2) \quad \text{and} \quad \max_{\substack{\beta_2 \in [0,1] \\ \pi_2 \in [0,1]}} R_2(\beta_2, \pi_2, \pi_1), \tag{1.37}$$

respectively.

Lemma 1.2. *The objectives in* (1.37) *can be written depending only on the variables* π_1 *and* π_2 *as in* (1.41) *and* (1.42) *shown on the top of the page after the next page with*

$$\beta_1^*(\pi_1) = \left[1/2 + \left(g_{12}\sigma^2 - g_{11}\sigma^2\right)/(2(1 - \pi_1)P_1)\right]_0^1, \tag{1.38}$$
$$\beta_2^*(\pi_2) = \left[1/2 + \left(g_{22}\sigma^2 - g_{21}\sigma^2\right)/(2\pi_2 P_2)\right]_0^1. \tag{1.39}$$

$$R_1(\beta_1,\pi_1,\pi_2) = \mathrm{ld}\left(1+\frac{(1-\pi_1)P_1\beta_1}{g_{11}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{(1-\pi_1)P_1(1-\beta_1)}{g_{12}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{\pi_1 P_1 \tilde{g}_{11}}{\sigma^2+(1-\pi_2)P_2\tilde{g}_{21}}\right) \tag{1.35}$$

$$R_2(\beta_2,\pi_2,\pi_1) = \mathrm{ld}\left(1+\frac{\pi_2 P_2\beta_2}{g_{21}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{\pi_2 P_2(1-\beta_2)}{g_{22}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{(1-\pi_2) P_2 \tilde{g}_{22}}{\sigma^2+\pi_1 P_1\tilde{g}_{12}}\right) \tag{1.36}$$

$$R_1(\pi_1,\pi_2) = \mathrm{ld}\left(1+\frac{(1-\pi_1)P_1\beta_1^*(\pi_1)}{g_{11}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{(1-\pi_1)P_1(1-\beta_1^*(\pi_1))}{g_{12}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{\pi_1 P_1 \tilde{g}_{11}}{\sigma^2+(1-\pi_2)P_2\tilde{g}_{21}}\right) \tag{1.41}$$

$$R_2(\pi_1,\pi_2) = \mathrm{ld}\left(1+\frac{\pi_2 P_2\beta_2^*(\pi_2)}{g_{22}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{\pi_2 P_2(1-\beta_2^*(\pi_2))}{g_{22}\sigma^2}\right) + \mathrm{ld}\left(1+\frac{(1-\pi_2) P_2 \tilde{g}_{22}}{\sigma^2+\pi_1 P_1\tilde{g}_{12}}\right) \tag{1.42}$$

Proof. The first derivative of R_1 in (1.35) w.r.t. β_1 is

$$\frac{\partial R_1(\beta_1,\pi_1,\pi_2)}{\partial \beta_1} = (1-\pi_1)P_1/\big(g_{11}\sigma^2+(1-\pi_1)P_1\beta_1\big) - (1-\pi_1)P_1/\big(g_{12}\sigma^2+(1-\pi_1)P_1(1-\beta_1)\big). \tag{1.40}$$

Setting (1.40) to zero and solving for β_1, we get the expression in (1.38). Including $\beta_1^*(\pi_1)$ in (1.35), we get (1.41). The analysis is analogous for the second player. □

According to Lemma Lemma 1.2, the optimization problems in (1.37) can be written as

$$\max_{\pi_1\in[0,1]} R_1(\pi_1,\pi_2) \quad \text{and} \quad \max_{\pi_2\in[0,1]} R_2(\pi_1,\pi_2). \tag{1.43}$$

We state the new game in which the strategy space of each player is a single dimensional set as

$$G' = \langle\{1,2\},([0,1],[0,1]),(R_1(\pi_1,\pi_2),R_2(\pi_1,\pi_2))\rangle. \tag{1.44}$$

Next, we will show that the game G' is a supermodular game.

Supermodular Games

The non-cooperative game G' in (1.44) is supermodular [23] if the following conditions are satisfied for each player: (C1) The strategy set of single dimensional feasible strategies is a compact set. (C2) The utility function $R_i(\pi_1, \pi_2)$ is upper semi continuous and has increasing differences in (π_1, π_2) on $[0, 1] \times [0, 1]$. For a more general definition of supermodular games in multi-dimensional strategies, refer to [24].

The first condition (C1) is satisfied because the strategy space $[0, 1]$ is compact. Second, the utility functions in (1.41) and (1.42) are continuous which satisfies the first part of property (C2). The second part of the property can be proven by showing

$$\frac{\partial^2 R_i(\pi_1, \pi_2)}{\partial \pi_1 \partial \pi_2} \geq 0, \quad i \in \{1, 2\}. \tag{1.45}$$

For player 1 (analogously player 2), (1.45) is fulfilled since

$$\frac{\partial^2 R_1(\pi_1, \pi_2)}{\partial \pi_1 \partial \pi_2} = \frac{P_1 \tilde{g}_{11} P_2 \tilde{g}_{21}}{(\sigma^2 + (1 - \pi_2) P_2 \tilde{g}_{21} + \pi_1 P_1 \tilde{g}_{11})^2} \geq 0, \tag{1.46}$$

Hence, the game G' is a supermodular game. Supermodular games have several interesting properties. A few properties from [24, Theorem 4.2.1] and [25, Result 4] are: (A1) There exists at least one pure strategy NE. (A2) The set of NEs is a complete lattice and there exist a largest and a smallest element. (A3) A unique NE is globally stable. Due to their interesting properties, supermodular games have been recognized in a few wireless network settings such as in [26, 27].

According to (A1), the game G' has at least one NE. From (A3) follows that the uniqueness on NE implies convergence of best response dynamics from all strategy points. Next, we will deliver a sufficient condition under which a unique NE exists in the game G'.

Best Response and Uniqueness of NE

In order to characterize the conditions for the existence of a unique NE and also how to reach it, we must formulate the best response of each player.

Proposition 1.3. The solution of the power allocation problem in (1.43) is characterized for BS_1 by the following cases: <u>Case 1-1:</u> If $g_{12} > \frac{1}{\sigma^2} + g_{11} - \frac{P_1}{\sigma^2} \pi_{11}^*(\pi_2)$ then

$$\pi_{11}^*(\pi_2) = \left[\frac{1}{2} + \frac{g_{11}\sigma^2}{2P_1} - \frac{\sigma^2 + (1 - \pi_2)\tilde{g}_{21} P_2}{2\tilde{g}_{11} P_1} \right]_0^1. \tag{1.47}$$

<u>Case 1-2:</u> If $g_{11} > \frac{1}{\sigma^2} + g_{12} - \frac{P_1}{\sigma^2} \pi_{12}^*(\pi_2)$ then

$$\pi_{12}^*(\pi_2) = \left[\frac{1}{2} + \frac{g_{12}\sigma^2}{2P_1} - \frac{\sigma^2 + (1 - \pi_2)\tilde{g}_{21} P_2}{2\tilde{g}_{11} P_1} \right]_0^1. \tag{1.48}$$

<u>Case 1-3:</u> If Cases 1-1 and 1-2 do not hold then

$$\pi_{13}^*(\pi_2) = \left[\frac{1}{3} + \frac{(g_{11} + g_{12})\sigma^2}{3P_1} - \frac{2\sigma^2 + 2(1 - \pi_2)\tilde{g}_{21} P_2}{3\tilde{g}_{11} P_1} \right]_0^1 \tag{1.49}$$

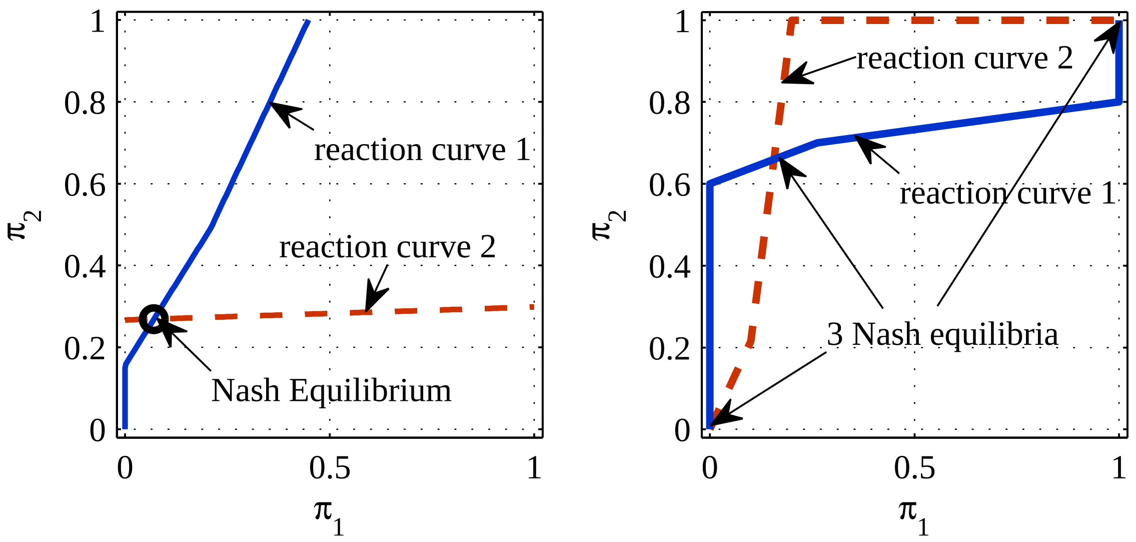

Figure 1.6: Plot of the reaction curves of the players. The intersection point of the two reaction curves is a NE.

In Cases 1-1 and 1-2, only one user is supported in the protected band of BS_1. For BS_2, the solution to its optimization problem in (1.43) is Case 2-1: If $g_{22} > \frac{1}{\sigma^2} + g_{21} - \frac{P_2}{\sigma^2}\pi^*_{21}(\pi_1)$ then

$$\pi^*_{21}(\pi_1) = 1 - \left[\frac{1}{2} + \frac{g_{21}\sigma^2}{2P_2} - \frac{\sigma^2 + \pi_1 \tilde{g}_{12} P_1}{2\tilde{g}_{22}P_2}\right]_0^1. \tag{1.50}$$

Case 2-2: If $g_{21} > \frac{1}{\sigma^2} + g_{22} - \frac{P_2}{\sigma^2}\pi^*_{22}(\pi_1)$ then

$$\pi^*_{22}(\pi_1) = 1 - \left[\frac{1}{2} + \frac{g_{22}\sigma^2}{2P_2} - \frac{\sigma^2 + \pi_1 \tilde{g}_{12} P_1}{2\tilde{g}_{22}P_2}\right]_0^1. \tag{1.51}$$

Case 2-3: If Case 2-1 and 2-2 do not hold then

$$\pi^*_{23}(\pi_1) = 1 - \left[\frac{1}{3} + \frac{(g_{21} + g_{22})\sigma^2}{3P_2} - \frac{2\sigma^2 + 2\pi_1 \tilde{g}_{12} P_1}{3\tilde{g}_{22}P_2}\right]_0^1. \tag{1.52}$$

Proof. The proof is provided in [28]. □

In Figure 1.6 the strategy spaces of the transmitters are plotted choosing $P_1 = P_2 = 1$. The *reaction curves* of the players represent the best response functions characterized in Proposition 1.3. A NE is a state of mutual best responses of the players. The NE can be found in Figure 1.6 as the intersection point of the reaction curves, where two cases are presented. In Figure 1.6 (left), a single NE exists. In Figure 1.6 (right) three NEs exist. A work that investigates the uniqueness of the NE using the constellation of the reaction curves is reported in [29]. In [29], two system pairs are considered that operate on two parallel channels, an interference channel and an interference relay channel. Two relaying strategies are investigated for which the uniqueness conditions of the NE are derived.

In Figure 1.6 (left), it can be seen that the reaction curve of player 1 consists of three parts. When π_1 is zero for small π_2, then the user in the shared band of BS_1 is not supported. When π_2 is between 0.2 and 0.5, all three users are supported for BS_1. For π_2 greater than 0.5, the reaction curve changes slope, corresponding to the case where one user is supported in the protected band, along with the user in the shared band.

Proposition 1.4. Sufficient condition for the existence of a unique NE is

$$\frac{\tilde{g}_{12}}{\tilde{g}_{11}} \frac{\tilde{g}_{21}}{\tilde{g}_{22}} < \frac{9}{4}. \tag{1.53}$$

Proof. The proof is provided in [28]. □

It can be observed that the sufficient condition on the uniqueness of NE is satisfied, when the product of the interference gains are small, compared to the product of the direct channel gains. Since in our case, uniqueness of NE implies global convergence of best response dynamics, a unique NE can be reached in a distributed manner when each BS iteratively applies a best response to the strategy of the other BS. According to simulations, a few iterations are sufficient to achieve a stable state in NE. In [30], a unified framework for parallel interference channels is given, for the analysis of the convergence of the best response dynamics. The condition in [30, Theorem 3] calculated for our setting is

$$\max\left\{a\frac{\tilde{g}_{12}}{\tilde{g}_{11}}, \frac{1}{a}\frac{\tilde{g}_{21}}{\tilde{g}_{22}}\right\} < 1, \quad \text{for } a > 0, \tag{1.54}$$

which leads to the condition

$$\frac{\tilde{g}_{12}}{\tilde{g}_{11}} \frac{\tilde{g}_{21}}{\tilde{g}_{22}} < 1. \tag{1.55}$$

Comparing (1.53) and (1.55), it can be observed that the condition in (1.53) is less restrictive than (1.55), for indicating whether a unique NE exists. Next, we will analyze the efficiency of the NE of our system.

1.2.3 Simulation Results

We compare the sum rate achieved in NE to the maximum achievable sum rate of the system. Finding the maximum sum rate of the two cells requires the search over all strategies of the BSs. The strategy space of each BS is given in (1.28). Since, each BS has a protected band in which no interference is caused to the other BS, then the maximum sum rate must be achieved with full power transmission in that subband. Accordingly, the strategy space $\mathcal{P}_i$ in (1.26) is parameterized using two parameters $\pi_i \in [0, 1]$ and $\beta_i \in [0, 1]$. In the shared band, the beamforming vectors that are necessary to obtain the maximum sum rate point are parameterized in [31], requiring two real-valued parameters,

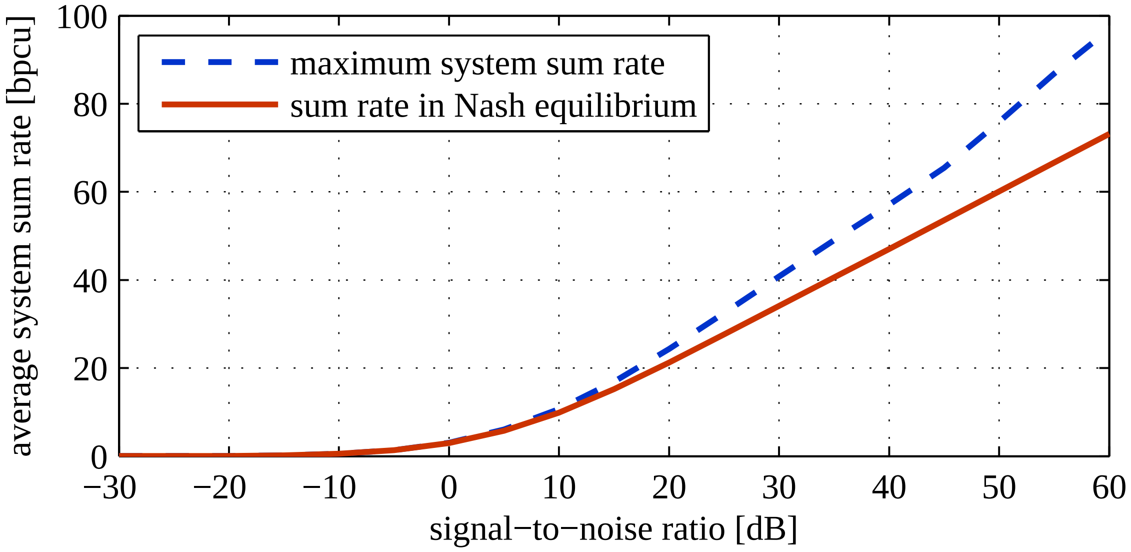

Figure 1.7: Average sum rate versus signal to noise ratio.

each ranging between zero and one. Accordingly, the achievable sum rate region can be produced using eight parameters, each between zero and one. We find the maximum sum rate using grid search.

In Figure 1.7 the average maximum sum rate is plotted averaging over 500 channel realizations. For the simulation we choose the maximum transmission power as $P_1 = P_2 = 1$. Signal-to-noise ratio is defined as SNR= $1/\sigma^2$. It can be observed that the NE is efficient compared to the maximum achievable sum rate. In the low SNR regime, the rates in NE are sum rate optimal. In mid and high SNR, slight performance loss in NE can be observed with respect to the optimal sum rate curve. Note, that the performance loss in NE is due to noncooperation between the BSs, which means that no signaling overhead is required between them.

2 Bayesian Games for Spectrum Sharing Systems with Partial Information

2.1 Bayesian Games for MIMO Interference Channels

This section considers the MIMO IC which has relevance in applications such as multi-cell coordination in cellular networks as well as spectrum sharing in wireless networks. We study the beamforming design under two extreme criteria; namely, egoism and altruism. The former maximizes the beamforming gain at the intended receiver and the latter minimizes interference created towards other receivers. The motivation is that combining egoistic and altruistic beamforming has been shown previously to be instrumental when optimizing the rates in a MISO IC, where the receivers have no interference canceling capability. Here, by using the framework of Bayesian games, we shed more light on these game-theoretic concepts in the more general context of MIMO channels and more particularly when coordinating parties only have CSI of the links that they can measure directly.

2.1.1 MIMO Interference Channel

Let $\mathcal{N} = \{1, \dots, N\}$ be a set containing a finite set $\mathcal{N}_c$, with cardinality $N_c \leq N$, of cooperating TXs, also termed as players. From now on, we use players and TXs interchangeably. We call the set $\mathcal{N}_c$ a coordination cluster and TXs outside the cluster will contribute to uncontrolled interference. The provided model has general applications in which the TXs can be base stations in cellular downlink where typically coordination is restricted to a subset of neighbouring cell sites while more distant sites cannot be coordinated over [32] ; nodes in ad-hoc network and spectrum sharing scenarios.

Each TX is equipped with N_t antennas and the RX with N_r antennas. Each TX communicates with a unique RX at a time. TXs are not allowed or able to exchange users' packet (message) information, giving rise to an interference channel over which we seek some form of beamforming-based coordination. The channel from TX i to RX j $\boldsymbol{H}_{ji} \in \mathbb{C}^{N_r \times N_t}$ is given by:

$$\boldsymbol{H}_{ji} = \sqrt{\alpha_{ji}}\bar{\boldsymbol{H}}_{ji}, \quad i, j = 1, \dots, N_c \tag{2.1}$$

Each element in channel matrix $\bar{\boldsymbol{H}}_{ji}$ is an i.i.d. complex Gaussian random variable with zero mean and unit variance and α_{ji} denotes the slow-varying shadowing and path-loss attenuation. $\bar{\boldsymbol{H}}_{ji}$ is circularly symmetric complex Gaussian and the probability density is

$$f_{\bar{\boldsymbol{H}}_{ji}}(\boldsymbol{H}) = \frac{1}{\pi^{N_t N_r}} exp(-Tr\left(\boldsymbol{H}\boldsymbol{H}^H\right)). \tag{2.2}$$

Limited Channel Knowledge

Although there may exist various ranges and definitions of local CSI, we assume a standard definition of a quasi-distributed CSI scenario where the devices (TX and RX alike) are able to gain knowledge of those local channel coefficients *directly connected* to them, as illustrated in Figure 2.1, possibly complemented with some limited non local information (to be defined later).

The set of CSI locally available (resp. not available) at TX i denoted by $\mathbb{B}_i$ (resp. $\mathbb{B}_i^{\perp}$) is denoted by:

$$\mathbb{B}_i = \{\boldsymbol{H}_{ji}\}_{j=1,\dots,N_c} \; ; \; \mathbb{B}_i^{\perp} = \{\boldsymbol{H}_{kl}\}_{k,l=1\dots N_c} \setminus \mathbb{B}_i \tag{2.3}$$

Similarly, define the set of channels known (resp. unknown) at RX i denoted by $\mathbb{M}_i$ (resp. $\mathbb{M}_i^{\perp}$) as: $\mathbb{M}_i = \{\boldsymbol{H}_{ij}\}_{j=1,\dots,N_c}$; $\mathbb{M}_i^{\perp} = \{\boldsymbol{H}_{kl}\}_{k,l=1\dots N_c} \setminus \mathbb{M}_i$. By construction here, locally available channel knowledge, $\mathbb{B}_i$, is only known to TX i but *not* other TXs. We call this knowledge $\mathbb{B}_i$ the *type* of player (TX) i, in the game-theoretic terminology [33].

In the view of TX i, the decision to be made shall be based on its type $\mathbb{B}_i$ and its *beliefs* on other TXs types. Since TX i does not know other TXs types, we assume that TX i has a probability density over the possible values of other players channel knowledge $\mathbb{B}_j$. For simplicity, we assume that these *beliefs* are symmetric: the probability density of the *Gaussian channels* available at TX i regarding $\mathbb{B}_j$ is the same as the probability density of TX j over $\mathbb{B}_i$. The asymmetric path loss attenuations α_{ji} are assumed to be long-term statistics and known to the TXs. And we assume that the channel coefficients in the network are statistically independent from each other. We define here the joint beliefs (probability density) at TX i:

$$\mu_i = p(\mathbb{B}_i^{\perp}) = f_{\bar{\boldsymbol{H}}_{ji}}(\boldsymbol{H})^{N_c(N_c-1)} = \mu. \tag{2.4}$$

The TX index i is dropped because the beliefs are symmetric among TXs, given the asymmetric path loss coefficients α_{ji}. $p(.)$ is a probability measure and $f_{\bar{\boldsymbol{H}}_{ji}}(\boldsymbol{H})$ is the probability density of a complex Gaussian channel defined in (2.2). The second equality relies on the assumptions that the channel coefficients from any TX to any RX are independent.

Based on its *belief*, TX i designs the transmit beamforming vector, $\boldsymbol{w}_i \in \mathbb{C}^{N_t \times 1}$. As in several important contributions dealing with coordination on the interference channel [31, 34–39], we assume linear beamforming. We call the transmit beamforming vector $\boldsymbol{w}_i$ an action of TX i and denote the set of all possible actions by $\mathcal{A}$ at any TX.

$$\mathcal{A} = \left\{\boldsymbol{w} \in \mathbb{C}^{N_t \times 1} : \|\boldsymbol{w}\|^2 \leq 1\right\} \tag{2.5}$$

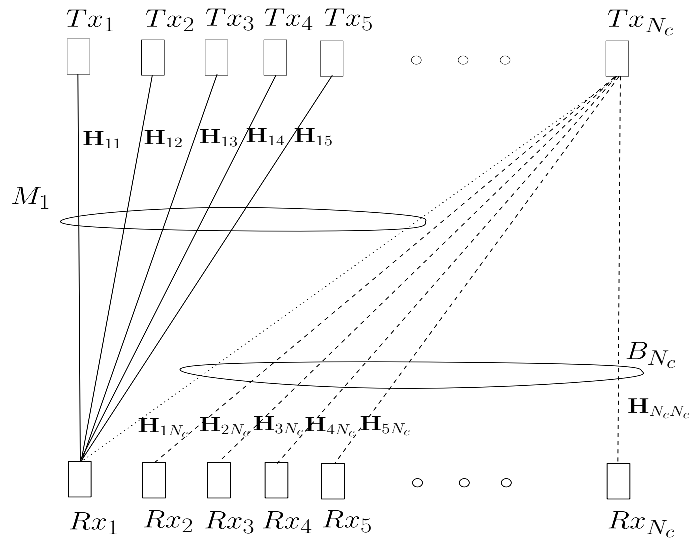

Figure 2.1: Limited channel knowledge model: as an illustration, the local CSI available at TX N_c is shown in dashed lines. The local CSI available at RX 1 is shown in solid lines.

The received signal at RX i is therefore

$$y_i = \boldsymbol{v}_i^H \boldsymbol{H}_{ii} \boldsymbol{w}_i + \sum_{j \neq i}^{N_c} \boldsymbol{v}_i^H \boldsymbol{H}_{ij} \boldsymbol{w}_j + n_i \tag{2.6}$$

where n_i is a Gaussian noise with power σ_i^2. Note that the noise levels σ_i^2 depend on the link index which was not considered in previous work on transmitter coordination. The RXs are assumed to employ maximum SINR (Max-SINR) beamforming throughout the section so as to also maximize the link rates [40]. The receive beamformer $\boldsymbol{v}_i$ is classically given by:

$$\boldsymbol{v}_i = \frac{\boldsymbol{C}_{Ri}{}^{-1} \boldsymbol{H}_{ii} \boldsymbol{w}_i}{\|\boldsymbol{C}_{Ri}{}^{-1} \boldsymbol{H}_{ii} \boldsymbol{w}_i\|} \tag{2.7}$$

where $\boldsymbol{C}_{Ri}$ is the covariance matrix of received interference and noise

$$\boldsymbol{C}_{Ri} = \sum_{j \neq i} \boldsymbol{H}_{ij} \boldsymbol{w}_j \boldsymbol{w}_j^H \boldsymbol{H}_{ij}^H P + \sigma_i^2 \boldsymbol{I}. \tag{2.8}$$

P is the transmit power. Note that the receive beamformer $\boldsymbol{v}_i$ is a function of all transmit beamforming vectors $\boldsymbol{w}_i$. When the transmit beamforming vector $\boldsymbol{w}_i$ is optimized, the received beamforming vector is modified accordingly.

Importantly, the noise will in practice capture thermal noise effects but also any interference originating from the rest of the network, i.e. coming from transmitters located beyond the coordination cluster. Thus, depending on path loss and shadowing effects, the $\{\sigma_i^2\}$ may be quite different from each other [41]. Figure 2.2 illustrates a system of

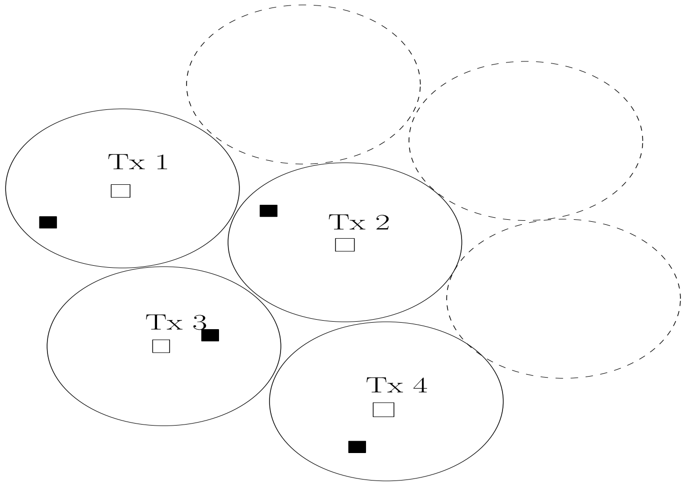

Figure 2.2: This figure illustrates a system of $N = 7$ cells where $N_c = 4$ form a coordination cluster. Empty squares represent transmitters whereas filled squares represent receivers. The noise power (which includes out of cluster interference) undergone in each cell varies from link to link.

$N = 7$ cells where $N_c = 4$ form a coordination cluster. Note that we consider the sum of uncoordinated source of interference and thermal noise to be spatially white. The non-colored interference assumption is justified in the scenario where receivers cannot obtain specific knowledge of the interference covariance and can be interpreted as a worst case scenario, since the receivers cannot use their spatial degrees of freedom to further cancel uncontrolled interference.

Receiver feedback v.s. Reciprocal Channel: In the case of reciprocal channels, e.g. time-division-duplex systems (TDD), the feedback requirement to obtain $\mathbb{B}_i$ can be replaced by a channel estimation step based on uplink pilot sequences. Additionally, it will be classically assumed that the receivers are able to estimate the covariance matrix of their interference signal, based on, say, transmit pilot sequences.

2.1.2 Bayesian Games with Receiver Beamformer Feedback

We assume that TX i has the local channel state information $\mathbb{B}_i$ and the added knowledge of receive beamformers through a feedback channel. Note that in the case of reciprocal channels, the receive beamformer feedback is *not* required.

We can now define the Bayesian game on interference channel.

Definition 2.1. The Bayesian game on interference channel can be described by a 5-tuple:

$$G =< \mathcal{N}_c, \mathcal{A}, \{\mathbb{B}_i\}, \mu, \{u_i\} > \qquad (2.9)$$

where μ denotes the *beliefs* of the players and $\{u_i\}$ denotes the utility functions of the players, which can be either egoistic or altruistic.

Specific definitions of u_i will be given in the following sections. The players are assumed to be *rational* as they maximize their own utility based on their *types* and *beliefs*.

Definition 2.2. A pure-strategy of player i, $s_i : \mathbb{B}_i \rightarrow \mathcal{A}_i$ is a deterministic choice of action given information $\mathbb{B}_i$ of player i.

Definition 2.3. A strategy profile $\boldsymbol{s}^* = (s_i^*, s_{-i}^*)$ achieves the Bayesian Equilibrium if s_i^* is the best response of player i given strategy tuple s_{-i}^* for all other players and is characterized by

$$\forall i \;\; s_i^* = \arg\max \mathcal{E}_{\mathbb{B}_i^\perp} \left\{ u_i(s_i, s_{-i}^*) \right\}. \tag{2.10}$$

Note that, intuitively, the player's strategy is optimized by averaging over the *beliefs* (the distribution of all missing state information) while in a standard game, such expectation is not required.

In the following sections, we derive the equilibria for egoistic and altruistic Bayesian games respectively.

Egoistic Bayesian Game

Definition 2.4. Denote the set of transmit beamforming vectors of players $j, j \neq i$, by $\boldsymbol{w}_{-i}$. The egoistic utility function for TX i is defined as its received SINR

$$u_i(\boldsymbol{w}_i, \boldsymbol{w}_{-i}) = \frac{|\boldsymbol{v}_i^H \boldsymbol{H}_{ii} \boldsymbol{w}_i|^2 P}{\sum_{j \neq i}^{N_c} |\boldsymbol{v}_i^H \boldsymbol{H}_{ij} \boldsymbol{w}_j|^2 P + \sigma_i^2}. \tag{2.11}$$

Based on TX i's belief, TX i maximizes the utility function in (2.11) where $\boldsymbol{v}_i$ is a known quantity.

Lemma 2.1. *There exist at least one Bayesian Equilibrium in the egoistic Bayesian Game* G (2.9) *with utility function defined in* (2.11).

Proof. $\mathcal{A}_i$ is convex, closed and bounded for all players i and the egoistic utility function $u_i(\boldsymbol{w}_i, \boldsymbol{w}_{-i})$ is continuous in both $\boldsymbol{w}_i$ and $\boldsymbol{w}_{-i}$. The utility function is convex in $\boldsymbol{w}_i$ for any set $\boldsymbol{w}_{-i}$. Thus, at least one Bayesian Equilibrium exists [14, 42]. □

Theorem 2.1. *The best-response strategy of player i in the egoistic Bayesian Game G* (2.9) *with utility function* (2.11) *is to maximize the utility function based on its belief:*

$$\boldsymbol{w}_i^{Ego} = \arg\max \mathcal{E}_{\mathbb{B}_i^\perp} \left\{ u_i(\boldsymbol{w}_i, \boldsymbol{w}_{-i}) \right\}. \tag{2.12}$$

The best-response strategy of player i is

$$\boldsymbol{w}_i^{Ego} = V_{\max}(\boldsymbol{E}_i) \tag{2.13}$$

where $\boldsymbol{E}_i$ *denotes the* egoistic equilibrium matrix *for TX* i*, given by*

$$\boldsymbol{E}_i = \boldsymbol{H}_{ii}^H \boldsymbol{v}_i \boldsymbol{v}_i^H \boldsymbol{H}_{ii}.$$

Proof. The knowledge of receive beamformers decorrelates the maximization problem which can be written as

$$\begin{aligned} \boldsymbol{w}_i^{Ego} &= \arg \max_{\|\boldsymbol{w}_i\| \leq 1} \mathcal{E}_{\mathbb{B}_i^{\perp}} \left\{ \frac{1}{\sum_{j \neq i}^{N_c} |\boldsymbol{v}_i^H \boldsymbol{H}_{ij} \boldsymbol{w}_j|^2 P + \sigma_i^2} \right\} \\ &= \arg \max_{\|\boldsymbol{w}_i\| \leq 1} \boldsymbol{w}_i^H \boldsymbol{H}_{ii}^H \boldsymbol{v}_i \boldsymbol{v}_i^H \boldsymbol{H}_{ii} \boldsymbol{w}_i \end{aligned} \quad (2.14)$$

The egoistic-optimal transmit beamformer is therefore the dominant eigenvector of $\boldsymbol{H}_{ii}^H \boldsymbol{v}_i \boldsymbol{v}_i^H \boldsymbol{H}_{ii}$. □

Altruistic Bayesian Game

Definition 2.5. The utility of the altruistic game is defined here so as to minimize the sum of interference powers caused to other receivers.

$$u_i(\boldsymbol{w}_i, \boldsymbol{w}_{-i}) = -\sum_{j \neq i} |\boldsymbol{v}_j^H \boldsymbol{H}_{ji} \boldsymbol{w}_i|^2 \quad (2.15)$$

Note that the receive beamforming vectors $\boldsymbol{v}_j$ is a Max-SINR receiver which depends on the transmit beamforming vectors $\boldsymbol{w}_j$ and cause conflicts between TXs.

Lemma 2.2. *There exist at least one Bayesian Equilibrium in the altruistic Bayesian Game* G (2.9) *with utility function defined in* (2.15)*.*

Proof. $\mathcal{A}_i$ is convex, closed and bounded for all players i and the altruistic utility function $u_i(\boldsymbol{w}_i, \boldsymbol{w}_{-i})$ is continuous in both $\boldsymbol{w}_i$ and $\boldsymbol{w}_{-i}$. The utility function is concave in $\boldsymbol{w}_i$ for any set $\boldsymbol{w}_{-i}$. Thus, at least one Bayesian Equilibrium exists [14, 42] . □

Theorem 2.2. *Based on belief* μ*, TX* i *seeks to maximize the utility function defined in* (2.15)*. The best-response strategy is*

$$\boldsymbol{w}_i^{Alt} = V_{\min} \left(\sum_{j \neq i} \boldsymbol{A}_{ji} \right) \quad (2.16)$$

where $\boldsymbol{A}_{ji}$ *denotes the* altruistic equilibrium matrix for TX i towards RX j*, defined by* $\boldsymbol{A}_{ji} = \boldsymbol{H}_{ji}^H \boldsymbol{v}_j \boldsymbol{v}_j^H \boldsymbol{H}_{ji}$.

Proof. Recall the utility function to be $-\sum_{j \neq i} |\boldsymbol{v}_j^H \boldsymbol{H}_{ji} \boldsymbol{w}_i|^2 = -\sum_{j \neq i} \boldsymbol{w}_i^H \boldsymbol{A}_{ji} \boldsymbol{w}_i$. Since $\boldsymbol{v}_j$ are known from feedback or estimation in reciprocal channels, the optimal $\boldsymbol{w}_i$ is the least dominant eigenvector of the matrix $\sum_{j \neq i} \boldsymbol{A}_{ji}$. □

2.2 Bayesian Games for Uncoordinated SISO Spectrum Sharing Systems

A large gain in spectral efficiency is promised to wireless systems if spectrum is shared. A complete sharing of data and control information (information data and CSI) would lead to a multiple-access channel in the uplink and to a broadcast channel in the downlink, but would imply a high capacity backhaul, scalability issues, high complexity. SAPHYRE aims at investigating the tradeoff between performance and level of data and control information sharing in a network with complete sharing of the spectrum. In this section, we investigate the performance of a system intrinsically characterized by (i) a limited level of cooperation among communication entities which are rather competing for the same resources and (ii) a decentralized resource management. We assume here that data information sharing is not possible and we study the impact of a limited sharing of control information (e.g. CSI, coordination, etc.).

In literature, many contributions focus on the channels with complete channel state information at the transmitters. Alternatively, iterative algorithms are proposed whose convergence to an equilibrium point is based on the feedbacks from the receivers. A well known and thoroughly studied example of this class of algorithms is the iterative waterfilling algorithm suitable for frequency selective interference channels (see [43] and references therein). Each receiver feedbacks the transmitter of interest with overall power spectral density (PSD) of the interference plus noise and the transmitter adapts its transmit PSD consequently. The convergence speed of these algorithms limits their applicability. Additionally, the required feedbacks reduce the system spectral efficiency. In [44] slow fading channels are considered with slow fading and initial partial channel state information. By using the approach of repeated games, information about the channel and the interactions is acquired. When the constraints of a communication system do not allow for the convergence of iterative algorithms (e.g. systems whose channels can be considered constant during the transmission of a codeword with constrained length but still varying from codeword to codeword or channels with constrained delay capacities) or do not support the intensive feedbacks required by iterative algorithms, Bayesian games provide a convenient theoretical framework.

In this section, we consider a quasi-static block fading interference channel with knowledge of the state of the direct links but only statistical knowledge on the interfering links. With this assumption, reliable communications are not possible and a certain level of outage has to be tolerated. We consider the resource allocation for utility functions based on the real throughput accounting for the outage events. We propose resource allocation algorithms based on both Bayesian games and optimization. In the context of Bayesian games, we investigate the two cases of power allocation for predefined transmission rates and joint power and rate allocation. The first game is a concave game and the mathematical tools by Rosen '65 are adopted for its analysis. The second group of games is studied introducing an equivalent game. The characteristics of the game theoretical approaches are analyzed in terms of existence and multiplicity of the NE. Special attention is devoted to the extreme regimes of high noise and interference limited regime. In the former case, a closed form expression for the NE is provided. In the latter case, criteria for the convergence of best response algorithms, the existence and uniqueness of the NE are discussed. The optimization approach is also analyzed in the two above mentioned regimes and closed form expressions for the resource allocation are provided.

System Model

Let us consider an interference channel with two sources $\mathcal{S}_1, \mathcal{S}_2$ and two destinations $\mathcal{D}_1, \mathcal{D}_2$. The two sources transmit independent information and source $\mathcal{S}_i$ aims at communicating with destination $\mathcal{D}_i$, for $i = 1, 2$. We assume that the channel is block fading, i.e. the channel gains of all the links are constant in the timeframe of a codeword but are independent and identically distributed from codeword to codeword. Note that these channels are often referred to as quasistatic channels or as channels with delay-limited capacity [6]. We denote by $g_i,\ i = 1, 2$ the channel power gains of the direct links $\mathcal{S}_1 - \mathcal{D}_1$ and $\mathcal{S}_2 - \mathcal{D}_2$ and by h_{12} and h_{21} the channel power gains of the interfering links $\mathcal{S}_1 - \mathcal{D}_2$ and $\mathcal{S}_2 - \mathcal{D}_1$. All the channel gains fade independently such that the channel power gain statistics are completely determined by the marginal distributions. Each source transmits only private information that can be decoded only by its targeted destination, or equivalently, each receiver performs single user decoding. Additionally, each source knows the realizations of both direct links g_1 and g_2 but not the realizations of the power gains h_{12} and h_{21} for the interfering links. This corresponds to a typical situation (e.g. in cellular systems) where the receivers estimate only the channel gains of the direct links and feed them back to the transmitter but neglect the interfering links. Throughout this work we make the additional assumption that the power gains of the interfering links are Rayleigh distributed, i.e. their probability density function is given by $\gamma_{H_{ij}}(h_{ij}) = \frac{1}{\sigma_{ij}^2}\mathrm{e}^{-\frac{h_{ij}}{\sigma_{ij}^2}}$. Furthermore, these statistics are known to both sources. At the receiver the channel is impaired by additive Gaussian noise with variance N_0.

Problem Statement

Because of the partial knowledge of the channel by the sources and the assumption of block fading, reliable communications, i.e. with error probability arbitrarily small, are not feasible and outage events may happen. If the source i transmits at a certain rate, expressed in nat/sec, with constant transmitted power P_i, an outage event happens if[1]

$$R_i > \log\left(1 + \frac{P_i g_i}{N_0 + P_j h_{ji}}\right), \qquad i, j = 1, 2 \text{ with } i \neq j, \tag{2.17}$$

and the outage probability of source i depends on the choice of R_i, P_i and P_j. We define the throughput as the average information that can be correctly received by the destina-

[1] We adopt the notation $\log$ for natural logarithms and rates are expressed in nat/sec.

tion. The throughput is given by

$$T_i(P_i, R_i, P_j) = R_i \Pr\left\{ R_i \leq \log\left(1 + \frac{P_i g_i}{N_0 + P_j h_{ji}}\right)\right\} \tag{2.18}$$

where $i, j = 1, 2$ with $i \neq j$, and $\Pr\{\mathcal{E}\}$ denotes the probability of the event $\mathcal{E}$.

The two sources need to determine autonomously and in a decentralized manner the transmitting power P_i and, eventually also the rate R_i. A natural criterion is to allocate such resources in order to maximize the throughput while keeping power consumption moderate. Then, we define the objective function for source $\mathcal{S}_i$ as

$$u_i((P_i, R_i), (P_j, R_j)) = T_i(P_i, R_i, P_j) - C_i P_i \tag{2.19}$$

where C_i is the cost for unit power.

By making use of the assumption on the power gain distributions of the interfering links, the utility of $\mathcal{S}_i$ is given by

$$\begin{aligned}
& u_i((R_i, P_i), (R_j, P_j)) \\
& = R_i \Pr\left\{ R_i \leq \log\left(1 + \frac{P_i g_i}{N_0 + P_j h_{ji}}\right)\right\} - C_i P_i \\
& = \begin{cases} R_i\left(1 - \exp\left(-\frac{t_i}{P_j \sigma_{ij}^2}\right)\right) & -C_i P_i, \{P_j > 0,\ P_i, R_i \geq 0\}; \\ 0, & \{P_j > 0,\ P_i = R_i = 0\}; \\ R_i - C_i P_i, & \{P_j = 0, R_i, P_i \geq 0, P_i \geq \frac{(\mathrm{e}^{R_i}-1)N_0}{g_i}\}; \\ -C_i P_i, & \{P_j = 0, R_i, P_i \geq 0, P_i \leq \frac{(\mathrm{e}^{R_i}-1)N_0}{g_i}\}; \end{cases} \\
& = \begin{cases} R_i F_i(t_i) - C_i P_i, & \{P_j > 0,\ P_i, R_i \geq 0\} \setminus \{P_i = R_i = 0\}; \\ 0, & \{P_j > 0,\ P_i = R_i = 0\}; \\ R_i - C_i P_i, & \{P_j = 0, R_i, P_i \geq 0, P_i \geq \frac{(\mathrm{e}^{R_i}-1)N_0}{g_i}\}; \\ -C_i P_i, & \{P_j = 0, R_i, P_i \geq 0, P_i \leq \frac{(\mathrm{e}^{R_i}-1)N_0}{g_i}\}; \end{cases}
\end{aligned} \tag{2.20}$$

where $t_i = \frac{P_i g_i}{\mathrm{e}^{R_i}-1} - N_0$ and $F_i(t_i) = 1 - \exp\left(-\frac{t_i}{P_j \sigma_{ij}^2}\right)$ and C_i is the cost of unit power by user i.

Since the objective function of $\mathcal{S}_i$ depends also on the power allocated by $\mathcal{S}_j$ the problem falls naturally in the framework of strategic games. Then, the objective of source $\mathcal{S}_i$ is to determine the transmit power P_i, and eventually the rate R_i, that selfishly maximizes its utility function $u_i((P_i, R_i), (P_j, R_j))$ under the assumption that a similar strategy is adopted by the other source. In the following Sections 2.2.1 and 2.2.2 we investigate the previous problem for two cases. More specifically, in Section 2.2.1 we consider the practical case as the rate is fixed[2] and each source has to determine the power which

[2] In practical systems rates are typically allocated at higher layer and defined in a discrete set eventually singleton.

maximizes its utility (2.19). In Section 2.2.2 we consider the general case where each source has to select its strategy defined in terms of the power and the transmitting rate, jointly. In Section 2.2.3 we also consider the optimization approach in the two asymptotic regimes of interference limited and noise limited systems. Closed form expressions for the resource allocation are provided.

2.2.1 Interference Game for Power Allocation

In this section we assume that the rates R_1 and R_2 are assigned to the source and define the power allocation problem as a strategic game $\mathcal{G}_{\mathcal{P}}$ defined by the triplet $\left\{\mathrm{S}, \mathcal{P}, (u_i)_{i\in\mathrm{S}}\right\}$, where $\mathrm{S} = \{\mathcal{S}_1, \mathcal{S}_2\}$ is the set of players or sources, $\mathcal{P}$ represents the set of strategies, and u_i is the utility function of source $\mathcal{S}_i$. The set of strategies is $\mathcal{P} = \mathbb{R}^+ \times \mathbb{R}^+$, where $\mathbb{R}^+$ is the set of nonnegative real numbers. Note that in this case R_i and R_j need be interpreted as parameters of the utility function (2.19) rather than as variables. Furthermore, we assume that $R_i \neq 0$ for $i = 1, 2$ otherwise the game becomes trivial. The definition of the game depends on the costs C_1 and C_2 via the utilities in (2.19). When convenient for the sake of comprehension we express this dependency explicitly via the notation $\mathcal{G}_{\mathcal{P}}(C_1, C_2)$.

We shall look for a NE, that is, a strategy $(P_1^*, P_2^*) \in \mathcal{P}$ such that for any $(P_1, P_2) \in \mathcal{P}$,

$$u_1((R_1, P_1), (R_2, P_2^*)) \leq u_1((R_1, P_1^*), (R_2, P_2^*)),$$
$$u_2((R_2, P_2), (R_1, P_1^*)) \leq u_2((R_2, P_2^*), (R_1, P_1^*)).$$

The existence of NE for game $\mathcal{G}_P$ is established in the following proposition.

Proposition 2.1. A NE of the game $\mathcal{G}_P$ exists in any closed interval and it is a fixed point of the equation

$$\rho((P_1^*, P_2^*), (P_1^*, P_2^*); R_1, R_2) = \max_{(\pi_1,\pi_2)\in\mathcal{P}} \rho((P_1^*, P_2^*), (\pi_1, \pi_2); R_1, R_2) \tag{2.21}$$

being

$$\rho((P_1, P_2), (\pi_1, \pi_2); R_1, R_2) = u_1((R_1, \pi_1), (R_2, P_2)) + u_2((R_2, \pi_2), (R_1, P_1)). \tag{2.22}$$

The Nash equilibriums of game $\mathcal{G}_{\mathcal{P}}$ need to satisfy the system of equations

$$\frac{\partial u_i}{\partial P_i} = -\frac{R_i}{P_j \sigma_{ji}^2} F'(t_i) - C_i = 0 \qquad i, j \in \{1, 2\},\ i \neq j. \tag{2.23}$$

From (2.23) it is straightforward to express P_i as a function of the strategy of the other competing player P_j

$$P_i = \frac{(\mathrm{e}^{R_i} - 1)}{g_i} \left(N_0 - \sigma_{ji}^2 P_j \log\left(\frac{C_i (\mathrm{e}^{R_i} - 1) \sigma_{ji}^2 P_j}{R_i g_i} \right) \right) \tag{2.24}$$

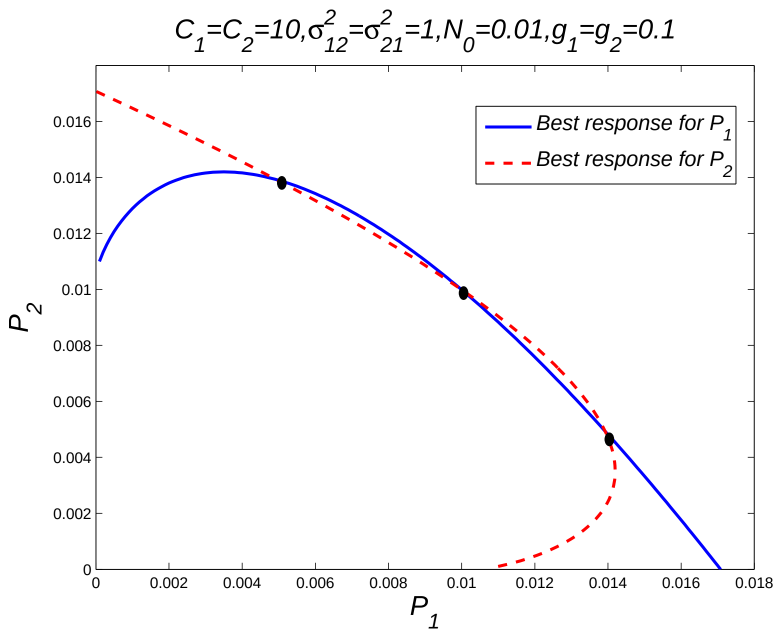

Figure 2.3: Example of a system with three NE

with $i, j = 1, 2$ and $i \neq j$. This function provides the best strategy transmitter i can apply when transmitter j adopts the strategy P_i, and it is shortly referred to as best response of user P_i. The plot of these curves on the plane $P_1 - P_2$ admit interesting interpretation. The points where the two curves intersect correspond to Nash equilibriums. Additionally, it provides information on the convergence of a best response algorithm.

The analysis of the best response algorithm yields to the following proposition characterizing the set of NE.

Proposition 2.2. The strategic game $\mathcal{G}_{\mathcal{P}}$ might have at most three NE in the interval $[0, P_1^{\blacktriangle}] \times [0, P_2^{\blacktriangle}]$, being $P_i^{\blacktriangle} = \frac{N_0}{g_i}(\mathrm{e}^{R_i} - 1) + \frac{R_i}{C_i \mathrm{e}}$.

Figure 2.3 shows the best responses of a system with more than a NE and the following setting: $R_1 = 0.1$, $R_2 = 0.1$, $C_1 = 10$ $C_2 = 10$, $g_1 = 0.1$, $g_2 = 0.1$, $\sigma_{12}^2 = 1$, $\sigma_{21}^2 = 1$, and $N_0 = 0.01$.

The characteristics of the NE set are of primary relevance to predict the output of a game. The uniqueness of a NE is an appealing property for an uncoordinated system. In the following proposition we provide sufficient conditions for the uniqueness of the NE

Proposition 2.3. Let $P_i = P_i(P_j)$ denotes the best responses of user i to power allocation

P_j of user j as defined in (2.24). If

$$P_i \left(\frac{N_0}{g_j}(\mathrm{e}^{R_j} - 1) + \frac{R_j}{C_j \mathrm{e}} \right) > \frac{g_j R_j}{\mathrm{e} C_j \sigma_{ij}^2 (\mathrm{e}^{R_j} - 1)} \tag{2.25}$$

$$P_j \left(\frac{\mathrm{e}^{R_i} - 1}{g_i} N_0 \right) > 0 \qquad i, j = 1, 2 \text{ and } i \neq j \tag{2.26}$$

game $\mathcal{G}_{\mathcal{P}}$ has a unique NE.

In the rest of this section we modify the game $\mathcal{G}_{\mathcal{P}}$ to account for the relevant practical issue of power constraints. The constrained game $\mathcal{G}_{\mathcal{P}}^c$ is defined by $\mathcal{G}_{\mathcal{P}}^c = \{\mathcal{S}, \mathcal{P}_c, (T_i)_{i \in \mathcal{S}}\}$, where $\mathcal{S}$ is as for the game $\mathcal{G}_{\mathcal{P}}$, $T_i \equiv T_i(P_I, R_i, P_j)$ is the throughput defined in (2.18), and the strategy set $\mathcal{P}_c = [0, P_1^{\mathrm{MAX}}] \times [0, P_2^{\mathrm{MAX}}]$, being P_1^{MAX} and P_2^{MAX} the maximum transmit powers. Games $\mathcal{G}_{\mathcal{P}}$ and $\mathcal{G}_{\mathcal{P}}^c$ are closely related as illustrated in the following proposition.

Proposition 2.4. The NE of the game $\mathcal{G}_{\mathcal{P}}^c$ exist and correspond to the NE (P_1^*, P_2^*) of games $\mathcal{G}_{\mathcal{P}}(C_1, C_2)$ such that $C_1(P_1^{\mathrm{MAX}} - P_1^*) = C_2(P_2^{\mathrm{MAX}} - P_2^*) = 0$ for $C_1, C_2 \geq 0$.

2.2.2 Interference Games for Joint Power and Rate Allocation

In this section we consider a communication system where the transmitters need to allocate both power and rate jointly with the aim of maximizing the utility function (2.19). The problem is defined as a strategic game $\mathcal{G} = \left\{\mathcal{S}, \overline{\mathcal{P}}, \{u_i\}_{i \in \{1,2\}}\right\}$, where $\mathcal{S}$ is the set of players (the two transmitters), $\mathcal{P}$ is the strategy set defined by $\overline{\mathcal{P}} \equiv \{((P_1, R_1), (P_2, R_2)) | P_1, P_2, R_1, R_2 \geq 0\}$, and u_i being the utility function defined in (2.19). Power and rate allocation is obtained as an equilibrium point of the system. When both transmitters aim at maximizing their utility function, a NE is the allocation strategy $(P_1^*, R_1^*, P_2^*, R_2^*)$ such that

$$u_1 \left(P_1^*, R_1^*, P_2^*, R_2^*\right) \geq u_1 \left(P_1, R_1, P_2^*, R_2^*\right) \quad \text{for} \quad \forall P_1, R_1 \in \mathbb{R}_+$$
$$u_2 \left(P_1^*, R_1^*, P_2^*, R_2^*\right) \geq u_2 \left(P_1^*, R_1^*, P_2, R_2\right) \quad \text{for} \quad \forall P_2, R_2 \in \mathbb{R}_+.$$

It is straightforward to verify that the utility function is not concave in R_i. Then, the classical results on N-concave games in [45] cannot be applied. The analysis of the general case results is very complex. A preliminary characterization of NE for game $\mathcal{G}$ is provided in the following proposition. This proposition provides closed form expressions for the NE at the boundary of the strategy set jointly with explicit conditions for the points being NE. Possible NE internal to the strategy set are provided in an implicit form and they will be further analyzed in additional propositions.

Proposition 2.5. A boundary point of the strategy set $\overline{\mathcal{P}}$ is a NE if and only if

$$P_i = R_i = 0 \tag{2.27}$$

$$P_j = \frac{1}{C_i} - \frac{N_0}{g_j} \qquad\qquad R_j = \log\left(1 + \frac{g_j - N_0 C_j}{N_0 C_j}\right) \tag{2.28}$$

and the following conditions are satisfied

$$g_j - N_0 C_j \geq 0 \tag{2.29}$$

$$\frac{g_i \alpha_j}{C_i \sigma_{ji}^2} \exp\left(-\frac{g_i \alpha_j}{C_i \sigma_{ji}^2} + \frac{1}{N_0 \alpha_j} + 1\right) \geq 1, \tag{2.30}$$

being $\alpha_j = \frac{C_j g_j}{g_j - N_0 C_j}$.

An internal point of the strategy set $\overline{\mathcal{P}}$ is a NE if and only if it is solution of the system of equations

$$\frac{1}{P_j \sigma_{ji}^2} \exp\left(-\frac{t_i}{P_j \sigma_{ji}^2}\right) = \frac{C_i(\mathrm{e}^{R_i} - 1)}{R_i g_i} \qquad i, j = 1, 2. \tag{2.31}$$

where $t_i = \frac{P_i g_i}{\mathrm{e}^{R_i} - 1} - N_0$ and P_1 and P_2 are given as functions of R_1 and R_2 by

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} C_1 \frac{\mathrm{e}^{R_1}}{\mathrm{e}^{R_1}-1} & \frac{C_1 \sigma_{21}^2}{R_1 g_1}(\mathrm{e}^{R_1} - 1) \\ \frac{C_2 \sigma_{12}^2}{R_2 g_2}(\mathrm{e}^{R_2} - 1) & C_2 \frac{\mathrm{e}^{R_2}}{\mathrm{e}^{R_2}-1} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{2.32}$$

and it satisfies the following inequalities

$$1 + R_i + \frac{g_i R_i}{C_i P_j \sigma_{ji}^2 (\mathrm{e}^{R_i} - 1)} - \frac{2 R_i \mathrm{e}^{R_i}}{\mathrm{e}^{R_i} - 1} > 0 \tag{2.33}$$

$$\frac{R_i^2 g_i}{C_i P_j \sigma_{ji}^2 (\mathrm{e}^{R_i} - 1)} - R_i - \left(1 - \frac{R_i \mathrm{e}^{R_i}}{\mathrm{e}^{R_i} - 1}\right)^2 > 0. \tag{2.34}$$

In order to get additional insights into the system behavior and in particular into the Nash equilibriums internal to the strategy set $\overline{\mathcal{P}}$, we consider firstly the following extreme cases before discussing the general case: (1) the noise tends to zero, (*interference limited regime*), (2) the noise is much higher than the transmitted power (*high noise regime*).

Interference Limited Regime

When the noise variance N_0 is negligible compared to the interference power level, the payoff function is efficiently approximated by (2.20), with $t_i = \frac{P_i g_i}{\mathrm{e}^{R_i}-1}$. Note that in the interference limited regime, the payoff (2.20) of user i is defined for $0 \geq N_0 \ll P_j$, i.e. $P_j > 0$. In the following proposition equilibriums of game $\mathcal{G}$ are obtained as equilibriums of an equivalent game in a single decision variable x_i for user i.

Proposition 2.6. When the the noise variance tends to zero, the NE of game $\mathcal{G}$ and internal to $\overline{\mathcal{P}}$ satisfy the system of equations

$$\begin{aligned} x_1 &= \kappa_2 f(x_2) \\ x_2 &= \kappa_1 f(x_1) \end{aligned} \tag{2.35}$$

where $x_i = \frac{g_i}{C_i P_j \sigma_{ji}^2}$, $\kappa_i = \frac{C_i g_j}{C_j \sigma_{ij}^2}$, $i, j \in 1, 2$, $i \neq j$ and

$$f(x) = \left(1 - \frac{\mathrm{e}^{R(x)} - 1}{xR(x)}\right)^{-1} \left(1 - \mathrm{e}^{-R(x)}\right)^{-1} \tag{2.36}$$

for $1 < x < \infty$. In (2.36), $R(x)$ is the unique **positive** solution of the equation

$$1 - \frac{xR}{\mathrm{e}^R - 1} \exp\left(-\frac{x}{\mathrm{e}^R} + \frac{\mathrm{e}^R - 1}{R\mathrm{e}^R}\right) = 0 \tag{2.37}$$

such that

$$-x + \frac{\mathrm{e}^R - 1}{R} \neq 0. \tag{2.38}$$

Let (x_1^0, x_2^0) be solutions of system (2.35). The corresponding NE is given by

$$P_1 = \frac{g_2}{C_2 x_2^0 \sigma_{12}^2}, \quad R_1 = R(x_1^0),$$
$$P_2 = \frac{g_1}{C_1 x_1^0 \sigma_{21}^2}, \quad R_2 = R(x_2^0).$$

Remarks

- The solution $\overline{R}(x)$ to the equation $\frac{\mathrm{e}^R - 1}{R} = x$ is also a solution to (2.37). Such a solution corresponds to a minimizer of the utility function.
- The solution $R(x_j)$ to (2.37) is the rate which maximizes the utility function corresponding to the transmit power of the other transmitter $P_i = \frac{g_j}{C_j x_j \sigma_{ij}^2}$. It lies in the interval $\left(0, \overline{R}(x_j)\right)$ and we refer to it as the *best response in terms of rate* of player j to strategy P_i of player i. Similarly, $\kappa_j f(x_j)$ is is inverse proportional to the *best response in terms of power* of user j to the strategy P_i of its opponent.
- Interestingly, the solution (x_1^0, x_2^0) to system (2.35) depends on the system parameters only through the constants κ_1 and κ_2.
- The existence and uniqueness of NE for the class of systems considered in Proposition 2.6 reduces to the analysis of the solution of system (2.35) and depends on the system via x_1 and x_2.
- The solution to equation (2.37) can be effectively approximated by $R(x) \approx 0.8 \log(x)$. Then, the function $f(x)$ is approximated by

$$f(x) \approx \left(1 - \frac{\mathrm{e}^{0.8 \log(x)} - 1}{x 0.8 \log(x)}\right)^{-1} \left(1 - \mathrm{e}^{-0.8 \log(x)}\right)^{-1}. \tag{2.39}$$

In Figure 2.5, $f(x_i)$ is plotted and compared to its approximation $\widetilde{f}(x_i)$. The approximation $\widetilde{f}(x_i)$ matches perfectly $f(x)$ such that can be utilized efficiently for practical and analytical objectives.

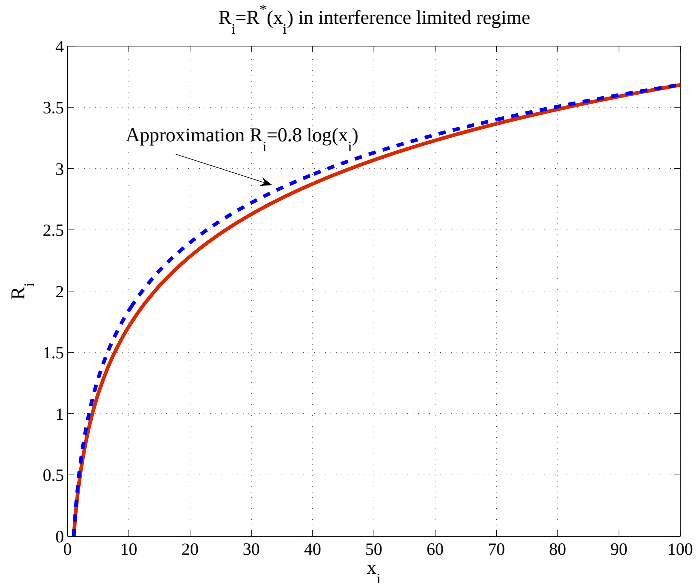

Figure 2.4: Best response $R^*(x_i)$ of user i to the transmitted power $P_j = \frac{g_i}{x_i \sigma_{ji} C_i}$ in solid line and its approximation $0.8 \log x_i$ in dashed line.

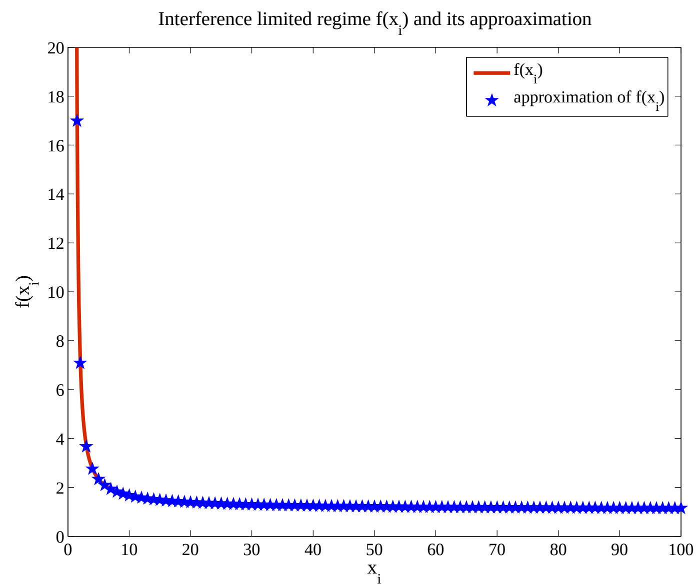

Figure 2.5: $f(x_i)$ in solid line and its approximation $\widetilde{f}(x_i)$

The following proposition provides sufficient conditions for the existence of a NE.

Proposition 2.7. When the interference is negligible compared to the noise variance, a NE of the game $\mathcal{G}$ exists if

$$(\kappa_1 - 1)(\kappa_2 - 1) > 0$$

with κ_i defined in Proposition 2.6.

General conditions for the uniqueness of the NE are difficult to determine analytically. Let us observe that in general a system with noise and interference from the primary source that tend to zero may have more than one NE. Let us consider the two systems corresponding to the two pairs of coefficients $\kappa_1^{(1)} = \kappa_2^{(1)} = 1.05$ and $\kappa_1^{(2)} = \kappa_2^{(2)} = 2$. The two curves $x_2 = \kappa_1^{(1)} f(x_1)$, for $i = \{1, 2\}$ cross each other in $x_1 = x_2$. Additionally, the curve $x_2 = \kappa_1^{(1)} f(x_1)$ has two asymptotes in $x_1 = 1$ and $x_2 = 1$. Then, by observing Figure 2.6 , it becomes apparent that the curves with $\kappa_1^{(1)} = \kappa_2^{(1)} = 1.05$ will cross again for high x_1 and x_2 values. In contrast, the curves with $\kappa_1^{(1)} = \kappa_2^{(1)} = 2$ will diverge from each other, and these crossing points correspond to NE. It is worth noticing that for $x_1 \gg 1$, $x_2 \approx 1$, (and for $x_2 \gg 1$, $x_1 \approx 1$). Then, from a telecommunication point of view, it is necessary to question whether the model for $N_0 \ll P_j g_j$ is still applicable. In fact, in such a case, $P_i \ll \frac{g_i}{C_i \sigma_{ij}^2}$, but also $P_i \gg N_0$ has to be satisfied because of the system model assumptions. Typically, the additional NE with some $x_i \approx 1$ are not interesting from a physical point of view since the system model assumptions are not satisfied.

By numerical simulations, we could observe that games with multiple NE exist for a very restricted range of system parameters, more specifically for $1 \leq \kappa_i \leq 1.1$.

Proposition 2.6 suggests also an iterative algorithm for computing NE based on the best response. Choose an arbitrary point $x_1^{(0)}$ and compute the corresponding value $x_2^{(0)} = \kappa_1 f(x_1^{(0)})$. From a practical point of view, this is equivalent to choose arbitrarily the transmitted power $P_2^{(0)} = \frac{g_1}{\sigma_{21}^2 x_1^{(0)} C_1}$ for transmitter 2 and determine the power allocation for user 1 which maximizes its utility function. The optimum power allocation for user 1 is $P_1^{(0)} = \frac{g_2}{\sigma_{12}^2 x_2^{(0)} C_2}$. We shortly refer to $P_1^{(0)}$ as the best response of user 1 to user 2. Then, by using $x_2^{(0)}$ it is possible to compute $x_1^{(1)} = \kappa_2 f\left(x_2^{(0)}\right)$, the best response of user 2 to user 1. By iterating on the computation of the best responses of user 1 and user 2 we can obtain resource allocations closer and closer to the NE and converge to the NE. We refer to this algorithm as the best response algorithm.

The best response algorithm is very appealing for its simplicity. Nevertheless, its convergence is not guaranteed. This issue is illustrated in Figure 2.6. Let us consider the interference channel with $\kappa_1 = \kappa_2 = 1.05$ and the corresponding solid and dashed curves $x_2 = \kappa_1 f(x_1)$ and $x_1 = \kappa_2 f(x_2)$. The NE exists and is unique but the best response algorithm diverges from the NE even for choices of the initial point arbitrarily close to the NE but different from it. Numerical results show that if κ_1 and κ_2 are both greater than 1.1, the best response algorithm always converges to a NE.

The following analytical result holds.

Proposition 2.8. For sufficiently large κ_1 and κ_2, the fixed point iterations

$$\begin{cases} x_1^{(k+1)} = \kappa_2 f(x_2^{(k)}), \\ x_2^{(k+1)} = \kappa_1 f(x_1^{(k)}), \end{cases} \tag{2.40}$$

converge.

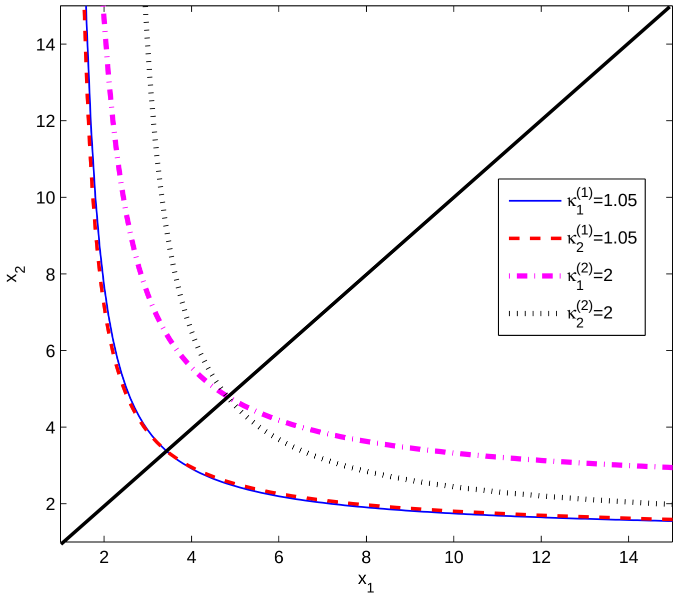

Figure 2.6: Graphical investigation of convergence of the best response algorithm in the interference limited regime

In fact, large values of κ_1 and κ_2 correspond to a realistic situation for system where the noise is negligible compared to the transmitted powers of the users.

High Noise Regime

Let us turn to the case when noise is much higher than the useful received power, $P_i g_i \ll N_0$. The throughput can be approximated by

$$\begin{aligned} \overline{T}_i\left(P_i, R_i, P_j, P_*\right) &= R_i \mathrm{Pr}\left\{R_i \leq \frac{P_i g_i}{N_0 + P_j h_{ji}}\right\} \\ &= R_i \mathrm{Pr}\left\{h_{ji} \leq \frac{1}{P_j}\left(P_i \frac{g_i}{R_i} - N_0\right)\right\} \end{aligned} \tag{2.41}$$

Interestingly, the throughput in (2.41) is nonzero for $\frac{P_k}{R_k} > \frac{N_0}{g_k}$. Since Proposition 2.5 defines completely the NE on the boundary of the strategy set in the general case, in this section we focus on only internal points of $\overline{\mathcal{P}}$. Then, the utility function is given by

$$v_i = R_i\left(1 - \exp\left(-\frac{\left(P_i\frac{g_i}{R_i} - N_0\right)}{P_j\sigma_{ji}^2}\right)\right) - C_iP_i \tag{2.42}$$

for $i = 1,2$. Correspondingly, we consider the game $\overline{\mathcal{G}} = \left\{\mathcal{S}, \mathcal{V}, \overset{o}{\overline{\mathcal{P}}}\right\}$, where the set of players coincides with the corresponding set in $\mathcal{G}$ while the utility function set $\mathcal{V}$ consists of the functions (2.42) and $\overset{o}{\overline{\mathcal{P}}}$ is the open interval obtained from $\overline{\mathcal{P}}$. The joint rate and power allocation is given by NE of game $\overline{\mathcal{G}}$.

The following proposition states the conditions for the existence and uniqueness of a NE in the strategy set and provides the equilibrium point.

Proposition 2.9. Game $\overline{\mathcal{G}}$ admits a NE if and only if

$$\frac{g_i}{C_i} > N_0, \quad i = 1,2.$$

If the above conditions are satisfied, $\overline{\mathcal{G}}$ has a unique equilibrium $\left((R_i^*, P_i^*), (R_j^*, P_j^*)\right)$ where P_i^* and P_j^* are the unique roots of the equations

$$\left(1 - \ln\left(\frac{C_jP_i\sigma_{ij}^2}{g_j}\right)\right)P_i\sigma_{ij}^2 = \frac{g_j}{C_j} - N_0 \tag{2.43}$$

and

$$\left(1 - \ln\left(\frac{C_iP_j\sigma_{ji}^2}{g_i}\right)\right)P_j\sigma_{ji}^2 = \frac{g_i}{C_i} - N_0 \tag{2.44}$$

in the intervals $\left(0, \frac{g_j}{C_j\sigma_{ij}^2}\right)$ and $(0, \frac{g_i}{C_i\sigma_{ji}^2})$ respectively. Also,

$$R_i = \frac{P_ig_iC_i}{g_i - P_j\sigma_{ji}^2C_i} \quad \text{and} \quad R_j = \frac{P_jg_jC_j}{g_j - P_i\sigma_{ij}^2C_j}.$$

Interestingly, the power allocation of user i decouples from the one of user j and P_i depends on its opponent only via the system parameter ratio $\frac{C_j}{g_j}$.

General Case

Let us consider now the general case, when the noise, the powers of interferences and the transmitted powers are of the same order of magnitude. A NE necessarily satisfies the system of equations (2.31) and (2.32). Substituting (2.32) in (2.31) yields

$$1 - \frac{x_iR_i}{\mathrm{e}^{R_i} - 1}\exp\left(-\frac{x_i}{\mathrm{e}^{R_i}} + \frac{\mathrm{e}^{R_i} - 1}{R_i\mathrm{e}^{R_i}} + n_i\right) = 0 \quad i = 1,2 \tag{2.45}$$

with $n_i = \frac{N_0}{P_j \sigma_{ji}^2}$. Equations (2.32) and (2.45) provide an equivalent system to be satisfied by NE. In order to determine a NE we can proceed as in the case of the interference limited regime. Observe that, in this case, (2.45) depends on the system parameters and the other player strategy not only via x_i but also via n_i. Then, the general analysis feasible for any communication system in the interference limited regime is no longer possible and the existence and multiplicity of NE should be studied independently for each communication system. In the following, we detail guidelines for this analysis.

From (2.45), it is possible to determine the best response in terms of rate of transmitter i to policy P_j of transmitter j. Conditions for the existence of such best response are detailed in the following statement.

Proposition 2.10. Equation (2.45) admits positive roots if and only if

$$1 - x_i \mathrm{e}^{-x_i+1+n_i} > 0. \tag{2.46}$$

If (2.46) is satisfied, (2.45) admits a single positive root in the interval $(0, \log x_i),$ which corresponds to the best response in terms of rate to policy P_j of user j.

From the best responses in terms of rate, it is straightforward to determine the best response in terms of powers for the two players.

2.2.3 Optimum Joint Rate and Power Allocation

In this section, we study the joint rate and power allocation when both users cooperate to maximize the utility function in the same strategy set $\overline{\mathcal{P}}$ of game $\mathcal{G}$.

The objective function is defined as

$$u\left(P_1, P_2, R_1, R_2\right) = \sum_{i=1, i\neq j}^{2} \left(T_i\left(P_i, R_i, P_j, R_j\right) - C_i P_i\right) \tag{2.47}$$

$$= \sum_{i=1}^{2} \left(R_i F_i(t_i) - C_i P_i\right). \tag{2.48}$$

We consider again the two extreme regimes when the noise is very high and when it is negligible compared to the interference power level. In both cases we show that the optimum resource allocation privileges a single user transmission. The following two propositions state the results.

Proposition 2.11. Let us assume that the noise is very high compared to the power transmitted by the transmitter, or equivalently, $\frac{g_i}{C_i} > N_0$ and $\frac{g_i}{C_i} \approx N_0$, $i = 1, 2$. Then, if

$$\log \frac{g_i}{C_i N_0} > \log \frac{g_j}{C_j N_0} \quad i, j = 1, 2 \quad i \neq j \tag{2.49}$$

transmitter i transmits at power $P_i = \frac{1}{g_i}\left(\frac{g_i}{C_i} - N_0\right)$ and rate $R_i = \log\left(\frac{g_i}{C_i N_0}\right) \approx \frac{g_i}{C_i N_0}$, and the transmitter j is silent, i.e. $P_j = R_j = 0$.

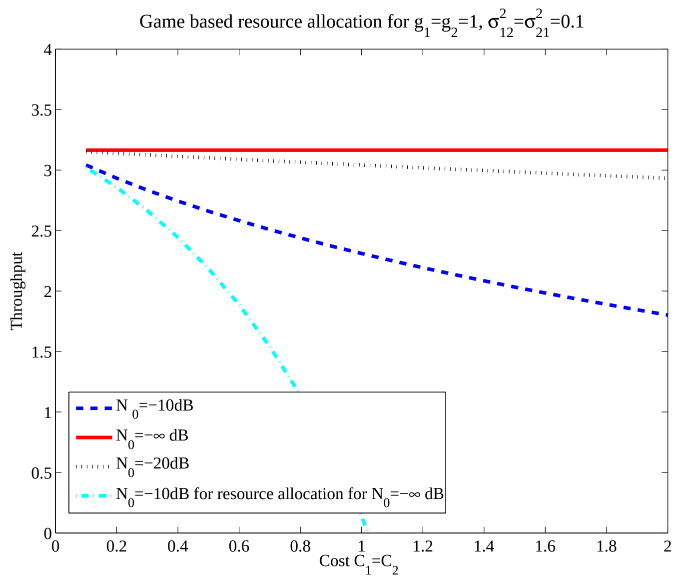

Figure 2.7: Throughput attained by NE versus costs $C_1 = C_2$ for different values of the noise.

Similarly, when the noise is negligible compared to the interference from the other user the following result holds.

Proposition 2.12. Let us assume that the noise variance is very low while the potential interference from the source could be substantially higher, i.e, $N_0 \to 0$ and $\frac{\sigma_{21}^2}{C_2} \gg 0$ for transmitter 1 and $N_0 \to 0$ and $\frac{\sigma_{12}^2}{C_1} \gg 0$ for transmitter 2. There does not exist an optimum allocation strategy for both $P_1, P_2 > 0$. If (2.49) is satisfied, transmitter i transmits at power and rate

$$P_i = \frac{1}{g_i}\left(\frac{g_i}{C_i} - N_0\right) \approx \frac{1}{C_i} \text{ and } R_i = \log\left(\frac{g_i}{C_i N_0},\right)$$

respectively, while transmitter j stays silent.

Closed form resource allocation strategies for the general case are not available and numerical constrained optimization is necessary.

2.2.4 Numerical Results

In this section, we assess the performance of the proposed algorithms and compare them. The resource allocation has a complex dependency on several system parameters, e.g. noise, channel gains, costs. We have seen from the proposed best response algorithm, in the reasonable case, the algorithm converges to the NE in the interval, not on the boundary. This gives us two benefits, firstly, this implies that a centralized communication is

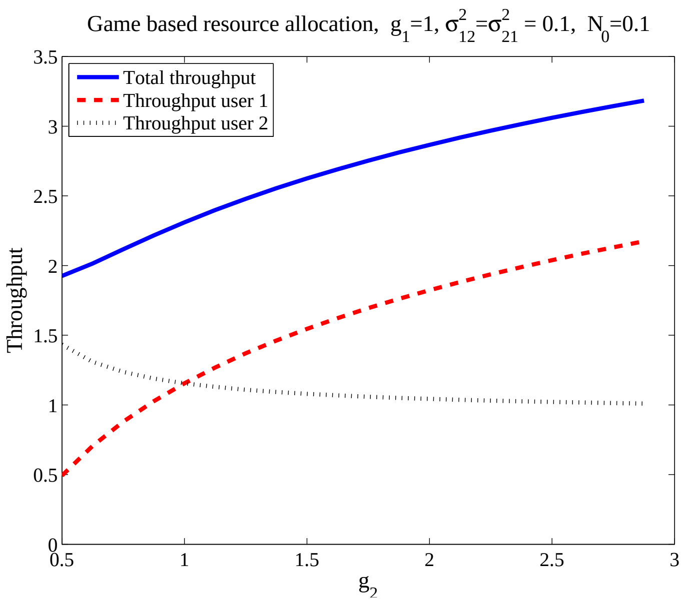

Figure 2.8: Throughput of the two users and total throughput attained by NE versus user 2 channel attenuation g_2.

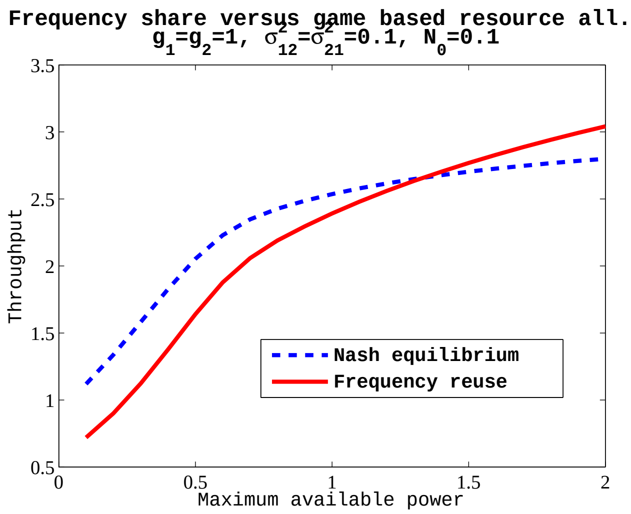

Figure 2.9: Throughput attained by NE versus Throughput by frequency sharing.

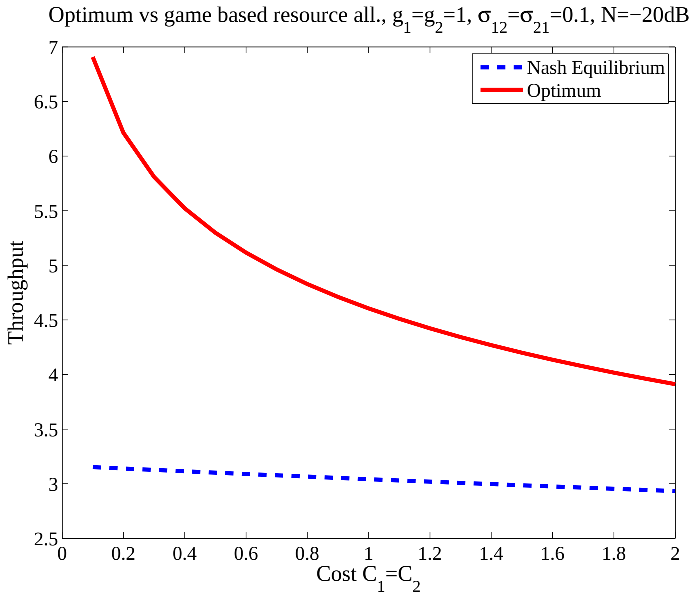

Figure 2.10: Throughput versus costs $C_1 = C_2$. Comparison between the throughput attained by NE or by optimum resource allocation.

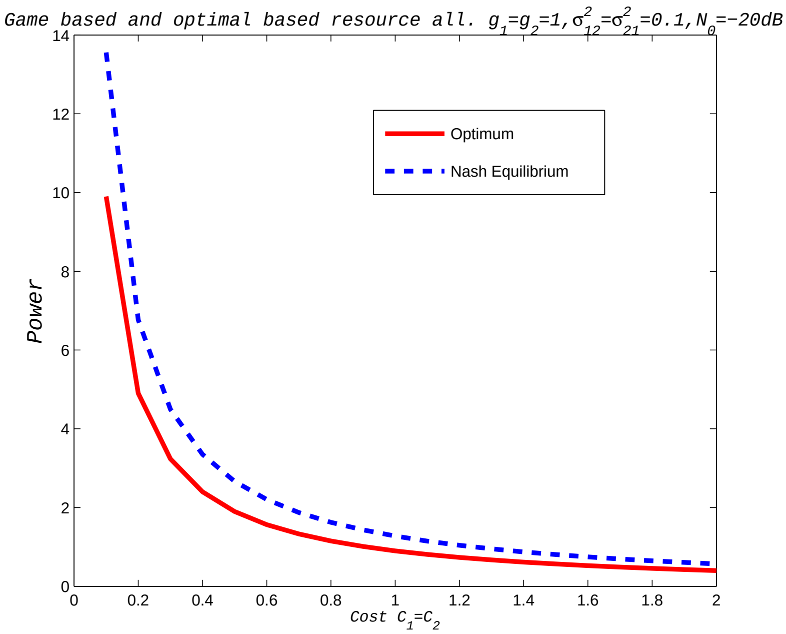

Figure 2.11: Transmitted power versus costs $C_1 = C_2$. Comparison between the resources allocated by NE or by optimum resource allocation.

unnecessary, secondly, for the NE in the interval, both users are transmitting, it guarantees the fairness of the system. We mainly consider the NE in the interval. We first investigate the performance of the game based resource allocation on the system parameters. We consider a system with parameters $\sigma_{12}^2 = \sigma_{21}^2 = 0.1$ and $g_1 = g_2 = 1$. Figure 2.7 shows the throughput attained by the game based algorithm for increasing costs $C_i = C_j$. As expected, in the general case, an increase of the costs implies a decrease of the achievable throughput. The solid line in Figure 2.7 shows the throughput in the interference limited regime. In this case the system performance is completely independent of the channel cost. At first glance, this behavior could appear surprising. However, it is a straightforward consequence of Proposition 2.6 when we observe that the best responses depend on the costs only via the ratio C_1/C_2. The dependency of the throughput on the costs becomes more and more relevant when the noise increases. Finally, the dashed dotted line in Figure 2.7 shows the degradation in terms of throughput, when the presence of noise is neglected in the resource allocation but $N_0 = -10dB$. Figure 2.8 illustrates the dependency of the throughput on the channel attenuation g_2 of user 2 for the following set of parameters: $\sigma_{12}^2 = \sigma_{21}^2 = 0.1$, $N_0 = -10dB$, $C_1 = C_2 = 1$. For increasing values of g_2, the total throughput increases. In contrast, the throughput of user 1 decreases because of the increased interference of user 2 on user 1. Note that for game based resource allocation the users access simultaneously to the channel while the optimum resource allocation privileges a time sharing policy.

In order to evaluate the SAPHYRE gain achievable by the proposed resource allocation, in Figure 2.9 we compare the game based resource allocation to the performance of the conventional system where the frequency band is divided into two sub-bands and statically assigned to each of the communications. We refer to this case as orthogonal frequency band allocation. For low power constraints, the throughput of the game based resource allocation outperforms the orthogonal frequency band allocation. However, when the available transmitting power grows large, the orthogonal frequency band strategy gains larger throughput than the game based resource allocation.

Figures 2.10 and 2.11 compare the game based resource allocation to the optimum one. They show the throughput and the power, respectively, as function of the costs. For very low values of N_0 and low costs, the optimum resource allocation outperforms significantly the game based approach at the expenses of fairness. In fact, the former assigns the spectrum to a single user. The performance loss at the NE decreases as the costs increases.

3 Cooperative Games in Spectrum Sharing Networks

The issue of sharing resources among competitive operators is a fundamental issue in next generation of communications. Cooperation enables a better exploitation of the resources and promises higher revenues for network providers. However, cooperation among competitive entities is complicated by the sensitive issue of conflicting interests. Thus, a critical point is to motivate these entities to cooperate. This can be achieved by a careful distribution of the incremental revenues obtained by the cooperation. Cooperative games offer a suitable theoretical framework to address this problem. More specifically, they offer tools to let the system work in an optimum point while distributing the optimized payoff in such a way that the grand coalition is stable and no operator or group of operators prefer to withdraw for it.

It is apparent that cooperation is appealing when the global system is characterized by parameters (e.g. traffic load, channel state, node availability) that evolve in time. Although, to have systems evolving in time is the most common situation, a framework for cooperative games in dynamics systems, capable to distribute instantaneously the payoff, was not available. In SAPHYRE, we developed some fundamental mathematical tools to deal with dynamics systems and guarantee the stability of the coalition and a fair instantaneous redistribution of the payoff over time. With this aim, we model the dynamic of the system as a Markov Decision Process (MDP) (see e.g. [46]) and developed algorithms for distributing the payoff and analyzed some complexity issue. Then, the theoretical results are applied to communication systems.

SAPHYRE work on dynamic cooperative game is presented in five sections. In the first three sections, we present the theoretical tools while the last two are dedicated to applications to communication systems. It is worth to note that the results presented here are just preliminary results and a lot has to be done to address the challenging task of motivating and guarantying competitive operators to cooperate in a stable manner.

Section 3.1 provides analytical tools to address the quite frequent situation where some providers share their own resources (e.g. nodes) in a common dynamic network but prefer to keep control of their own nodes. From a mathematical point of view, the solution of the related cooperative game to share fairly costs or revenues boils down to the solution of zero-sum games with perfect information, i.e. the solution of zero-game over MDPs where each state of the chain is controlled by a single decision maker. Thus, in Section 3.1 we deal with zero-sum two-player stochastic games with perfect information. We propose two algorithms to find the uniform optimal strategies and one method to compute the optimality range of discount factors. We prove the convergence in finite time for one algorithm. The uniform optimal strategies are also optimal for the long run average criterion and, in transient games, for the undiscounted criterion as well. Applications of these results to cooperative games for communications systems are presented in the following sections.

In Section 3.2, we aim to extend the previous setting to the case of a system that evolves dynamically and all decision makers/operators can take decision simultaneously, eventually, influencing the evolution of the system itself. A relevant constraint of the theory developed here is that utilities/costs are transferable. That implies that the operators costs or revenues are somehow shared in a monetary way. Although, this is a strong constraint that does not allow to apply the proposed dynamic cooperative game framework to non-transferable resources, such as power or frequency bands, in Section 3.5 we readapt the framework proposed in Section 3.2 to a communication system and apply it to the case of non-transferable utilities to solve a rate allocation problem. More precisely, in Section 3.2, we deal with multi-agent 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 fundamental properties that guarantee the stability of the game, namely, time consistency property, contents greedy players, and strengthen the coalition cohesiveness throughout the game.

Dynamic cooperative games consider infinite observation windows and their complexity is intrinsically quite high. Therefore, it is of primary relevance to investigate suboptimal algorithms with polynomial complexity providing solutions within a controlled confidence interval of the optimum solution. This is a very challenging task and we attacked it in Section 3.3 by considering a special kind of games known as weighted voting games. In a weighted voting Markovian game, several states succeed each other over time, following a discrete-time Markov chain. In each state, a different weighted voting static game is played by the same set of players. We investigate the approximation of the *S*hapley-*S*hubik power index in the *M*arkovian game (SSM). We prove that an exponential number of queries on coalition values is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy. Motivated by this, we propose and study three randomized approaches to compute a confidence interval for SSM. They rest upon two different assumptions, static and dynamic, about the process through which the estimator agent learns the coalition values. Such approaches can also be utilized to compute confidence intervals for the Shapley value in any Markovian game. The proposed methods require a number of queries which is polynomial in the number of players in order to achieve a polynomial accuracy.

In Section 3.4 and 3.5 we provide applications of the analytical tools presented in Section 3.1 to 3.3 to systems with shared resources.

In Section 3.4, we study a cooperative game where several providers coexist by sharing network nodes individually controlled. They offer a connection service to the same service point and want to minimize the cost of the service offered to their customers while maximizing the costs for the customers of their opponents by properly defining a routing strategy. The algorithm proposed here to find the optimal routing strategies for the network providers is based on algorithms for zero-sum stochastic games presented in

Section 3.1. Based on this approach, we determine the conditions under which the coalition is stable and no player prefer to withdraw. More specifically, we provide algorithms to determine the network link costs in such a way that all providers have interest in cooperating. As by-product, we apply the proposed algorithm to two-player games both in networks subject to hacker attacks and in epidemic networks.

In Section 3.5, we consider a multiple access channel (MAC) in which the channel coefficients follow a Markov chain on a finite set of states. We assume that any subset of users that does not intend to cooperate can, in the worst case scenario, jam the active users. This scenario can be modeled as a cooperative Markovian game with non-transferable utility. A rate allocation is fair, and belongs to the Core of the game, whenever no subsets of players can attain a better allocation when the remaining users jam. We derive the feasibility region of the Markovian MAC, under both the discounted and the average criterion. We compute the set of allocation rates in the Markovian process that are feasible, efficient, and stable. A method to derive the corresponding single stage allocations is provided. We analyze some fair allocations, such as max-min fairness, proportional fairness and Nash bargaining solution. We provide a condition ensuring the consistency of such fairness criteria between each single stage game and the long run game. We investigate the situation in which no agreement is reached and we study the relation between the already mentioned fair allocations criteria and the Nash bargaining solution, when the number of players increases.

3.1 Uniform Optimal Strategies in Two-player Zero-sum Stochastic Games

Stochastic games, also called multi-agent MDPs, are multi-stage interactions among several participants in an environment whose conditions change stochastically, influenced by the decisions of the players. A detailed survey on this topic can be found in the book [46] by Filar and Vrieze. In this section, we deal with zero-sum stochastic games with two players and with perfect information. Under the perfect information assumption, the reward and the transition probabilities in each state are controlled at most by one player. Our results are grounded on the following references. Filar proved in [47] an ordered field property for the value of switching control stochastic games; the games with perfect information are a specific case of them. Raghavan and Syed provided in [48] a policy improvement algorithm to determine the optimal strategies for two-player zero-sum perfect information games under the discounted criterion, for a fixed discount factor. Inspired by the work of Jeroslow [49], Hordijk, Dekker, and Kallenberg proposed in [50] to find the optimal discount strategies for MDPs for all discount factors close enough to 1 by utilizing the simplex method in the ordered field of rational functions with real coefficients. Filar, Altman, and Avrachenkov presented in [51] some algorithms for the computation of uniform optimal strategies in the context of perturbed MDPs; in [52], the same authors proposed an efficient asymptotic simplex method based on Laurent series expansion.
Our contribution is organized as follows. We first introduce our stochastic game model

in Section 3.1.1. In Section 3.1.2 we prove that, for all discounted factors close enough to 1, the discounted value belongs to the field of rational functions with real coefficients. Moreover, we summarize the main results of [50]. In Section 3.1.3 we present some useful results on uniform optimality in perfect information games. Then, we propose two algorithms which compute a pair of uniform discount optimal strategies $(\mathbf{f}^*, \mathbf{g}^*)$, which are optimal in the long run average criterion as well. The convergence in a finite time of the first algorithm, based on policy improvement, is proven in Section 3.1.4. A simple method to find the range of discount factors in which $(\mathbf{f}^*, \mathbf{g}^*)$ are discount optimal is shown in Section 3.1.5. We present our second algorithm, which is a best response algorithm, in Section 3.1.6. In Section 3.1.7 we show by simulations that the second algorithm has a lower complexity than the first one, in terms of number of pivot operations. In Section 3.1.8 we finally prove that, for transient stochastic games, $(\mathbf{f}^*, \mathbf{g}^*)$ are optimal under the undiscounted criterion as well.

Some notation remarks: the ordering relation between vectors of the same length $\mathbf{a} \geq (\leq)\mathbf{b}$ means that for every component $\mathbf{a}(i)$ and $\mathbf{b}(i)$, $\mathbf{a}(i) \geq (\leq)\mathbf{b}(i)$. The indicator function is referred to as $\mathbb{1}$. The symbol δ stands for Kronecker delta. The discount factor and the interest rate are barred, i.e. $(\overline{\beta}, \overline{\rho})$, if they represent a fixed real value; the symbols (β, ρ) represent the related real variables.

3.1.1 The model

In a two-player stochastic game Γ we have a set of states $S = \{s_1, s_2, \ldots, s_N\}$. For each state s, the set of actions available to Player i is called $A^{(i)}(s) = \{a_1^{(i)}(s), \ldots, a_{m_i(s)}^{(i)}(s)\}$, $i = 1, 2$. In zero-sum games, for each triple (s, a_1, a_2) with $a_1 \in A^{(1)}(s)$, $a_2 \in A^{(2)}(s)$ we assign an immediate reward $\mathbf{r}(s, a_1, a_2)$ to Player 1, $-\mathbf{r}(s, a_1, a_2)$ to Player 2 and a transition probability distribution $p(.|s, a_1, a_2)$ on S.

A stationary strategy $\mathbf{u} \in \mathbf{U}_S$ for Player i determines the probability $\mathbf{u}(a|s)$ that in state s Player i chooses the action $a \in A^{(i)}(s)$. We assume that both the number of states and the overall number of available actions are finite. Let $p(s'|s, \mathbf{f}, \mathbf{g})$ and $\mathbf{r}(s, \mathbf{f}, \mathbf{g})$ be the expectation with respect to the stationary strategies $(\mathbf{f}, \mathbf{g})$ of $p(s'|s)$ and of $\mathbf{r}(s)$, respectively.

Let $\overline{\beta} \in [0; 1)$ be the discount factor and $\overline{\rho}$ be the interest rate such that $\overline{\beta}(1 + \overline{\rho}) = 1$. Note that when $\overline{\beta} \uparrow 1$, then $\overline{\rho} \downarrow 0$. We define $\mathbf{\Phi}_{\overline{\beta}}(\mathbf{f}, \mathbf{g})$ as the N-by-1 vector whose i-th component $\mathbf{\Phi}_{\overline{\beta}}(s_i, \mathbf{f}, \mathbf{g})$ equals the expected $\overline{\beta}$-discounted reward when the initial state of the stochastic game is s_i:

$$\mathbf{\Phi}_{\overline{\beta}}(\mathbf{f}, \mathbf{g}) = \sum_{t=0}^{\infty} \overline{\beta}^t \mathbf{P}^t(\mathbf{f}, \mathbf{g}) \mathbf{r}(\mathbf{f}, \mathbf{g}),$$

where $\mathbf{P}(\mathbf{f}, \mathbf{g})$ and $\mathbf{r}(\mathbf{f}, \mathbf{g})$ are the *N*-by-*N* transition probability matrix and the *N*-by-1 state-wise expected reward vector associated to the pair of strategies $(\mathbf{f}, \mathbf{g})$, respectively.

Definition 1. The $\overline{\beta}$-discounted value of the game Γ is such that

$$\mathbf{\Phi}_{\overline{\beta}}(\Gamma) = \sup_{\mathbf{f}} \inf_{\mathbf{g}} \mathbf{\Phi}_{\overline{\beta}}(\mathbf{f}, \mathbf{g}) = \inf_{\mathbf{g}} \sup_{\mathbf{f}} \mathbf{\Phi}_{\overline{\beta}}(\mathbf{f}, \mathbf{g}). \tag{3.1}$$

An optimal strategy $\mathbf{f}^*_{\overline{\beta}}$ ($\mathbf{g}^*_{\overline{\beta}}$) for Player 1 (2) assures to him a reward which is at least (at most) $\mathbf{\Phi}_{\overline{\beta}}(\Gamma)$.

Let $\mathbf{\Phi}(\mathbf{f}, \mathbf{g})$ be the long run average reward of the game Γ associated to the pair of strategies $(\mathbf{f}, \mathbf{g})$:

$$\mathbf{\Phi}(\mathbf{f}, \mathbf{g}) = \lim_{T \to \infty} \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{P}^t(\mathbf{f}, \mathbf{g}) \mathbf{r}(\mathbf{f}, \mathbf{g})$$

and let $\mathbf{\Phi}(\Gamma)$ be the value vector for the long run average criterion of the game Γ, defined in an analogous way to expression (3.1). The existence of optimal strategies in discounted stochastic games is guaranteed by the following Theorem.

Theorem 3.1 ([46]). *Under the hypothesis of discounted reward criterion, stochastic games possess a value, the optimal strategies* $(\mathbf{f}^*_{\overline{\beta}}, \mathbf{g}^*_{\overline{\beta}})$ *exist among stationary strategies and, moreover,* $\mathbf{\Phi}_{\overline{\beta}}(\Gamma) = \mathbf{\Phi}_{\overline{\beta}}(\mathbf{f}^*_{\overline{\beta}}, \mathbf{g}^*_{\overline{\beta}})$.

Definition 2. A stationary strategy $\mathbf{h}$ is said to be uniform discount optimal (or equivalently uniform optimal) for Player $i = 1, 2$ if $\mathbf{h}$ is optimal for Player i for every $\overline{\beta}$ close enough to 1 (or, equivalently, for all $\overline{\rho}$ close enough to 0).

In this section, we deal with stochastic games with perfect information.

Definition 3. Under the hypothesis of *perfect information*, in each state at most one player has more than one action available.

Let $S_1 = \{s_1, \ldots, s_{t_1}\}$ be the set of states controlled by Player 1 and $S_2 = \{s_{t_1+1}, \ldots, s_{t_1+t_2}\}$ be the set controlled by Player 2, with $t_1 + t_2 \leq N$.

3.1.2 The ordered field of rational functions with real coefficients

Let $P(\mathbb{R})$ be the ring of the polynomials with real coefficients.

Definition 4. The dominating coefficient of a polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is the coefficient a_k, where $k = \min\{i : a_i \neq 0\}$ and we denote it with $\mathcal{D}(f)$.

Let $F(\mathbb{R})$ be the non-Archimedean ordered field of fractions of polynomials with coefficients in $\mathbb{R}$:

$$f(x) = \frac{c_0 + c_1 x + \cdots + c_n x^n}{d_0 + d_1 x + \cdots + d_m x^m} \qquad f \in F(\mathbb{R}),$$

where the operations of sum and product are defined in the usual way (see [50]). Two rational functions $h/g,\ p/q$ are identical (and we say $h/g =_l p/q$) if and only if $h(x)q(x) = p(x)g(x),\ \forall\, x \in \mathbb{R}$.

Lemma 3.1 ([50]). *A complete ordering in $F(\mathbb{R})$ is obtained by the rule: $p/q >_l 0$ if and only if $\mathcal{D}(p)\mathcal{D}(q) > 0$, where $p, q \in P(\mathbb{R})$.*

In the same way, we also define the operations of maximum ($\max_l$) and minimum ($\min_l$) in $F(\mathbb{R})$.

Lemma 3.2 ([50]). *The rational function p/q is positive $(p/q >_l 0)$ if and only if there exists $x_0 > 0$ such that $p(x)/q(x) > 0$ for every $x \in (0; x_0]$.*

Computation of Blackwell optimum policy in MDPs

Let us consider a MDP, which can be seen as a two-player stochastic game in which one of the two players fixes its own strategy. Let $A(s)$ be the finite action space available in state $s \in S$. Let $m(s) = |A(s)|$.

Definition 5. The strategy $\mathbf{f}^*$ is Blackwell optimal if and only if there exists $\bar{\rho}^* > 0$ such that $\mathbf{f}^*$ is optimal in the $(\bar{\rho}+1)^{-1}$-discounted MDP for all the interest rates $\bar{\rho} \in (0; \bar{\rho}^*]$.

In [50] the authors provide an algorithm to compute the Blackwell optimal policy in MDPs. It consists in solving the following parametric linear programming model:

$$\begin{cases} \max_{\mathbf{x}}{}_l \sum_{s=1}^{N} \sum_{a=1}^{m(s)} x_{s,a}(\rho)\mathbf{r}(s,a) \\ \sum_{s=1}^{N} \sum_{a=1}^{m(s)} [(1+\rho)\delta_{s,s'} - p(s'|s,a)]\, x_{s,a}(\rho) =_l 1,\ s' \in S \\ x_{s,a}(\rho) \geq_l 0, \quad s \in S,\ a \in A(s) \end{cases} \tag{3.2}$$

in the ordered field of rational functions with real coefficients $F(\mathbb{R})$. The Blackwell optimal strategy is computed as $\mathbf{v}^*(a|s) := \mathbb{1}\left(x^*_{s,a}(\rho) >_l 0\right)$ for all $s \in S,\ a \in A(s)$, where $\{x^*_{s,a}(\rho),\ \forall\, s, a\}$ is the solution of (3.2).

Application to stochastic games

In this section we will introduce the ordered field $F(\mathbb{R})$ in stochastic games, not necessarily with perfect information.

Theorem 3.2. *Let $\mathbf{f}, \mathbf{g}$ be two stationary strategies for Players 1 and 2, respectively. Let $\mathbf{\Phi}_\rho(\mathbf{f},\mathbf{g}) : \mathbb{R} \to \mathbb{R}^N$ be the discounted reward associated to $(\mathbf{f},\mathbf{g})$, expressed as a function of ρ. Then, $\mathbf{\Phi}_\rho(\mathbf{f},\mathbf{g}) \in F(\mathbb{R})$.*

Proof. For any pair of stationary strategies $(\mathbf{f},\mathbf{g})$, we can write $\forall\, s \in [1; N]$:

$$\sum_{s'=1}^{N} [(1+\rho)\delta_{s,s'} - p(s'|s,\mathbf{f},\mathbf{g})]\mathbf{\Phi}_\rho(s',\mathbf{f},\mathbf{g}) = (1+\rho)\mathbf{r}(s,\mathbf{f},\mathbf{g}),$$

where ρ is a variable. By solving the above system of equations in the unknown $\mathbf{\Phi}_\rho$ by Cramer rule, it is evident that $\mathbf{\Phi}_\rho(\mathbf{f},\mathbf{g}) \in F(\mathbb{R})$. □

From Theorem 3.1 and Theorem 3.2 we obtain the following Lemma, which ensures that, for discounted factors close enough to 1, the discounted value exists and belongs to the field of rational functions with real coefficients.

Lemma 3.3. *Let Γ be a zero-sum stochastic game possessing uniform discount optimal strategies $\mathbf{f}^*$ and $\mathbf{g}^*$ for Players 1 and 2, respectively. Then, there exists $\mathbf{\Phi}^*_\rho(\Gamma) \in F(\mathbb{R})$ such that:*

$$\mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g}^*) \leq_l \mathbf{\Phi}_\rho(\mathbf{f}^*, \mathbf{g}^*) =_l \mathbf{\Phi}^*_\rho(\Gamma) \leq_l \mathbf{\Phi}_\rho(\mathbf{f}^*, \mathbf{g}), \quad \forall\, \mathbf{f}, \mathbf{g}. \tag{3.3}$$

Proof. By hypothesis, there exists $\overline{\rho}^* > 0$ such that $(\mathbf{f}^*, \mathbf{g}^*)$ are discounted optimal for all the interest rates $\overline{\rho} \in (0; \overline{\rho}^*]$. For Theorem 3.2, $\mathbf{\Phi}_\rho(\mathbf{f}^*, \mathbf{g}^*) \in F(\mathbb{R})$ and, from Theorem 3.1, the uniform optimum value $\mathbf{\Phi}_{\overline{\rho}}(\Gamma) = \mathbf{\Phi}_{\overline{\rho}}(\mathbf{f}^*, \mathbf{g}^*)\ \forall\, \overline{\rho} \in (0; \overline{\rho}^*]$. Hence, the saddle point relation in (3.3) holds. □

Definition 6. $\mathbf{\Phi}^*_\rho(\Gamma)$, defined as in (3.3), is the uniform discount value of the stochastic game Γ.

3.1.3 Uniform optimality in perfect information games

In a perfect information game, a pure stationary strategy for Player i is a probability distribution $\mathbf{f}(.|s)$ on the action space $A_i(s)$ such that there exists $a \in A_i(s)$ for which $\mathbf{f}(a|s) = 1$, for every $s \in S$.

Theorem 3.3 ([46]). *For a stochastic game with perfect information, both players possess uniform discount optimal pure stationary strategies, which are optimal for the average criterion as well.*

Definition 7. We call two pure stationary strategies adjacent if and only if they differ only in one state.

The following property holds, whose proof is analogous to the one in the field of real numbers in [48].

Lemma 3.4. *Let $\mathbf{g}$ be a strategy for Player 2 and $\mathbf{f}, \mathbf{f}_1$ be two adjacent strategies for Player 1. Then, either $\mathbf{\Phi}_\rho(\mathbf{f}_1, \mathbf{g}) \geq_l \mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g})$ or $\mathbf{\Phi}_\rho(\mathbf{f}_1, \mathbf{g}) \leq_l \mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g})$, which means that the two vectors are partially ordered.*

The Lemma 3.4 allows us to give the following definition.

Definition 8. Let $(\mathbf{f}, \mathbf{g})$ be a pair of pure stationary strategies for Player 1 and 2, respectively. We call $\mathbf{f}_1$ $(\mathbf{g}_1)$ a uniform adjacent improvement for Player 1 (2) in state s_t if and only if $\mathbf{f}_1$ $(\mathbf{g}_1)$ is a pure stationary strategy which differs from $\mathbf{f}$ $(\mathbf{g})$ only in state s_t and $\mathbf{\Phi}_\rho(\mathbf{f}_1, \mathbf{g}) \geq_l \mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g})$ $(\mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g}_1) \leq_l \mathbf{\Phi}_\rho(\mathbf{f}, \mathbf{g}))$, where the strict inequality holds in at least one component.

As in the case in which the discount interest rate is fixed, we achieve the following result. Its proof directly stems from the Bellman optimality equations in the ordered field $F(\mathbb{R})$.

Lemma 3.5. *Let Γ be a stochastic game with perfect information. A pair of pure stationary strategies $(\mathbf{f}^*, \mathbf{g}^*)$ is uniform discount optimal if and only if no uniform adjacent improvement is possible for both players.*

In perfect information games, the following result holds.

Lemma 3.6 ([48]). *In a zero-sum, perfect information, two-player discounted stochastic game Γ with interest rate $\overline{\rho} > 0$, a pair of pure stationary strategies $(\mathbf{f}^*, \mathbf{g}^*)$ is optimal if and only if $\mathbf{\Phi}_{\overline{\rho}}(\mathbf{f}^*, \mathbf{g}^*) = \mathbf{\Phi}_{\overline{\rho}}(\Gamma)$, the value of the discounted stochastic game Γ.*

From the above result we can easily derive the analogous property in the ordered field $F(\mathbb{R})$.

Lemma 3.7. *In a zero-sum, two-player stochastic game Γ with perfect information, a pair of pure stationary strategies $(\mathbf{f}^*, \mathbf{g}^*)$ are uniform discount optimal if and only if $\mathbf{\Phi}_\rho(\mathbf{f}^*, \mathbf{g}^*) =_l \mathbf{\Phi}^*_\rho(\Gamma) \in F(\mathbf{R})$, where $\mathbf{\Phi}^*_\rho(\Gamma)$ is the uniform discount value of Γ.*

Proof. The *only if* statement coincides with the assertion of Theorem 3.1. Conversely, if a pair of strategies $(\mathbf{f}^*, \mathbf{g}^*)$ has the property $\mathbf{\Phi}_\rho(\mathbf{f}^*, \mathbf{g}^*) =_l \mathbf{\Phi}^*_\rho(\Gamma)$, then there exists $\rho^* > 0$ such that $\forall \overline{\rho} \in (0; \rho^*]$, $\mathbf{\Phi}_{\overline{\rho}}(\mathbf{f}^*, \mathbf{g}^*)$ coincides with the value of the game Γ, $\forall \overline{\rho} \in (0; \rho^*]$. Then, thanks to Lemma 3.6, we can say that the strategies $\mathbf{f}^*, \mathbf{g}^*$ are optimal in the discounted game Γ for any $\overline{\rho} \in (0; \rho^*]$, which means that they are uniform optimal. □

Remark 3.1.1. Generally, the discounted value of a stochastic game for all the interest rates close enough to 0 belongs to the field of real Puiseux series (see [46]).

Let s_t be a state controlled by Player $i = 1, 2$ and $X \subset A_i(s_t)$. Let us call Γ^t_X the stochastic game which is equivalent to Γ except in state s_t, where Player i has only the actions X available. We present the following Lemma, whose proof is analogous to the one in the real field (see [48]).

Lemma 3.8. *Let $i = 1, 2$ and $s_t \in S_i$, $X \subset A_i(s_t)$, $Y \subset A_i(s_t)$, $X \cap Y = \emptyset$. Then $\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) \in F(\mathbb{R})$, the uniform value of the game $\Gamma^t_{X\cup Y}$, equals*

$$\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) =_l \max{}_l \{\mathbf{\Phi}^*_\rho(\Gamma^t_X), \mathbf{\Phi}^*_\rho(\Gamma^t_Y)\} \qquad \text{if } \ i = 1,$$
$$\mathbf{\Phi}^*_\rho(\Gamma^t_{X\cup Y}) =_l \min{}_l \{\mathbf{\Phi}^*_\rho(\Gamma^t_X), \mathbf{\Phi}^*_\rho(\Gamma^t_Y)\} \qquad \text{if } \ i = 2.$$

3.1.4 Policy improvement algorithm

In this section we find a policy improvement algorithm which allows to find the uniform discount optimal strategies for both players in a stochastic game with perfect information. Such strategies coincide with the optimal strategies for the long run average criterion, for Theorem 3.3. Following the lines of Raghavan and Syed's algorithm [48] for a fixed discount factor, we propose an algorithm suitable for the ordered field $F(\mathbb{R})$.

Let Γ be a zero-sum two-player stochastic game with perfect information. Let $\Gamma_i(\mathbf{f})$ be the MDP faced by Player i when the opponent fixes its own strategy g.

Algorithm 3.1.1.

Step 1 Select randomly a stationary deterministic pure strategy g for Player 2.

Step 2 Find the Blackwell optimal strategy for Player 1 in the MDP $\Gamma_1(\mathbf{g})$ by solving within the field $F(\mathbb{R})$ the following linear programming model:

$$\begin{cases} \max_{\mathbf{x}}{}_l \sum_{s=1}^{N} \sum_{a=1}^{m_1(s)} x_{s,a}(\rho)\mathbf{r}(s,a,\mathbf{g}) \\ \sum_{s=1}^{N} \sum_{a=1}^{m_1(s)} [(1+\rho)\delta_{s,s'} - p(s'|s,a,\mathbf{g})]\, x_{s,a}(\rho) =_l 1,\ s' \in S \\ x_{s,a}(\rho) \geq_l 0, \quad s \in S,\ a \in A_1(s) \end{cases} \tag{3.4}$$

and compute the pure strategy $\mathbf{f}$ as

$$\mathbf{f}(a|s) := \mathbb{1}\left(x^*_{s,a}(\rho) >_l 0\right) \qquad \forall\, s \in S,\ a \in A_1(s), \tag{3.5}$$

where $\{x^*_{s,a}(\rho),\ \forall\, s,a\}$ is the solution of (3.4).

Step 3 Find the minimum k such that in $s_{t_1+k} \in \{s_{t_1+1}, \ldots, s_{t_1+t_2}\}$ there exists an adjacent improvement g′ for Player 2, with the help of the simplex tableau associated to the following linear programming model:

$$\begin{cases} \min_{\mathbf{x}}{}_l \sum_{s=1}^{N} \sum_{a=1}^{m_2(s)} x_{s,a}(\rho)\mathbf{r}(s,\mathbf{f},a) \\ \sum_{s=1}^{N} \sum_{a=1}^{m_2(s)} [(1+\rho)\delta_{s,s'} - p(s'|s,\mathbf{f},a)]\, x_{s,a}(\rho) =_l 1,\ s' \in S \\ x_{s,a}(\rho) \geq_l 0, \quad s \in S,\ a \in A_2(s) \end{cases} \tag{3.6}$$

where the entering variables are $\{x_{s,a} :\ \mathbf{g}(a|s) = 1,\ \forall\, s\}$.
If no such improvement for Player 2 is possible, then go to step 4, otherwise set g := g′ and go to step 2.

Step 4 Set $(\mathbf{f}^*, \mathbf{g}^*) := (\mathbf{f}, \mathbf{g})$ and stop. The strategies $(\mathbf{f}^*, \mathbf{g}^*)$ are uniform discount and long run average optimal in the stochastic game Γ for Player 1 and Player 2, respectively.

Note that all the algebraic operations and the order signs are to be intended in the field $F(\mathbb{R})$.

Remark 3.1.2. Unlike the solution in [48], Algorithm 3.1.1 does not require the strategy search for Player 1 to be lexicographic. In fact, Player 1 faces in step 2 a classic Blackwell optimization.

Remark 3.1.3. Clearly, the roles of Player 1 and 2 are interchangeable in Algorithm 3.1.1. For simplicity, throughout this section the Player 1 will be assigned to step 2.

Remark 3.1.4. In step 3 of Algorithm 3.1.1, once the state s_{t_1+k} is found, the adjacent improvement involves the pivoting of any of the non-basic variable $x_{s_{t_1+k},a}$ to which corresponds a nonpositive (in the field $F(\mathbb{R})$) reduced cost.

Let us prove the convergence in finite time of Algorithm 3.1.1.

Theorem 3.4. *Algorithm 3.1.1 stops in a finite time and the pair of strategies* $(\mathbf{f}^*, \mathbf{g}^*)$ *are both uniform discount and long run average optimal in the stochastic game* Γ.

Proof. The proof follows the lines of the analogous one in the real field (see [48]). It proceeds by induction on the overall number of actions and it exploits Lemma 3.5 and Lemma 3.8. The main difference from [48], that does not affect the correctness of the proof, is that Player 1 is not constrained in optimizing lexicographically the MDP $\Gamma_1(\mathbf{g})$. For Theorem 3.3, $(\mathbf{f}^*, \mathbf{g}^*)$ are long run average optimal as well. □

Sensitivity to round-off errors

The first non-zero coefficients of the polynomials (numerator and denominator) of the tableaux obtained throughout the algorithm unfolding determine the positiveness of the elements of the tableaux themselves. Hence, Algorithm 3.1.1 is highly sensitive to the round-off errors that affect the null coefficients.

If the rewards and the transition probabilities for each pair of strategies are rational, then it is possible to work in the exact arithmetic and such inconveniences are completely avoided. In fact, if all the input data are rational, they will stay rational after the algorithm execution.

3.1.5 Computation of the optimality range factor

Let us report the analogous result to Lemma 3.5 when the discount factor is fixed.

Lemma 3.9 **([48]).** *Let* Γ *be a stochastic game with perfect information. Let* $\overline{\beta} \in [0; 1)$. *The pure stationary strategies* $(\mathbf{f}^*, \mathbf{g}^*)$ *are* $\overline{\beta}$*-discount optimal if and only if no uniform adjacent improvements are possible for both players in the* $\overline{\beta}$*-discounted stochastic game* Γ.

Let us define with $\zeta(f_\rho)$, where $f_\rho \in F(\mathbb{R})$, the set of positive roots of f_ρ such that $df_\rho/d\rho|_{\rho=u} < 0,\ \forall\, u \in \zeta(f_\rho)$. The following Lemma suggests how to compute the optimality range of discount factors.

Lemma 3.10. *Let* C *be the set of the reduced costs associated to the two optimal tableaux obtained at the step 2 and 3 of the last iteration of Algorithm 3.1.1. Let* $\overline{\rho}^* = \min_c \zeta(c)$, *where* $c \in C$. *If* $\overline{\rho}^*$ *does not exist, then the uniform optimal strategies* $(\mathbf{f}^*, \mathbf{g}^*)$ *are optimal for all* $\overline{\beta} \in [0; 1)$. *Otherwise,* $\overline{\beta}^* = (1 + \overline{\rho}^*)^{-1}$ *is the smallest value such that* $(\mathbf{f}^*, \mathbf{g}^*)$ *are* $\overline{\beta}$*-discount optimal in the game* Γ*, for all* $\overline{\beta} \in [\overline{\beta}^*; 1)$.

Proof. If $\overline{\rho}^*$ does not exist, then the reduced costs are non-negative for any $\overline{\rho} > 0$. Hence, $(\mathbf{f}^*, \mathbf{g}^*)$ are optimal $\forall\, \overline{\beta} \in [0; 1)$. Otherwise, $\forall\, \overline{\rho} \in (0; \overline{\rho}^*]$, the reduced costs are positive, hence no adjacent improvements are possible for both players. So, for Lemma 3.9 they are discounted optimal. If $\overline{\rho} > \overline{\rho}^*$ and $\overline{\rho}^* \in \mathbb{R}$, then at least one reduced cost is negative, hence at least an adjacent improvement is possible and $(\mathbf{f}^*, \mathbf{g}^*)$ are not $\overline{\beta}$-discount optimal, where $\overline{\beta} = (1 + \overline{\rho})^{-1}$. □

3.1.6 Best response algorithm

Let Γ be a zero-sum two-player stochastic game with perfect information. Consider the following best-response algorithm.

Algorithm 3.1.2.

Step 1 Select a stationary pure strategy $\mathbf{g}_0$ for Player 2. Set $k := 0$.

Step 2 Find the Blackwell optimal strategy $\mathbf{f}_k$ for Player 1 in the MDP $\Gamma_1(\mathbf{g}_k)$.

Step 3 If $\mathbf{g}_k$ is Blackwell optimal in $\Gamma_2(\mathbf{f}_k)$, then set $(\mathbf{f}^*, \mathbf{g}^*) := (\mathbf{f}_k, \mathbf{g}_k)$ and stop. Otherwise, find the Blackwell optimal strategy $\mathbf{g}_{k+1}$ for Player 2 in the MDP $\Gamma_2(\mathbf{f}_k)$, set $k := k + 1$ and go to step 2.

Obviously, if Algorithm 3.1.2 stops, $(\mathbf{f}^*, \mathbf{g}^*)$ is a pair of uniform discount and long run average optimal strategies, since they are both Blackwell optimal in the respective MDPs, $\Gamma_1(\mathbf{g}^*)$ and $\Gamma_2(\mathbf{f}^*)$.
The proof that Algorithm 3.1.2 never cycles is still an open problem. We found that $\mathbf{\Phi}_\rho(\mathbf{f}_{k+1}, \mathbf{g}_{k+1}) \leq_l \mathbf{\Phi}_\rho(\mathbf{f}_k, \mathbf{g}_k)$, is not true in general. However, if the conjecture in [48] were valid, then we could conclude that Algorithm 3.1.2 terminates in a finite time.

3.1.7 Complexity: simulation results

In Algorithm 3.1.1, Player 1 faces at each iteration an MDP optimization problem in the field of rational functions with real coefficients, which is solvable in polynomial time. Player 2, instead, is involved in a lexicographic search throughout the algorithm unfolding, whose complexity is at worst exponential in the number of states N. Player 2 lexicographically expands its search of its optimal strategy, and at the k-th iteration the two players find the solution of a subgame $\Gamma^{(k)}$ which monotonically tends to the entire stochastic game Γ.
The efficiency of Algorithm 3.1.1 is mostly due to the fact that most of the actions totally dominate other actions. In other words, it often occurs that an optimal action found in the subgame $\Gamma^{(k)}$, is optimum also in Γ, and consequently remains the same during all the remaining iterations. This exponentially reduces the policy space in which the algorithm needs to search.

Remark 3.1.5. Since in Algorithm 3.1.1 players' roles are interchangeble and since most of the actions dominate totally other actions, we suggest to assign the step 2 of the algorithm to the player whose total number of available actions is bigger.

Differently from [48], the search for Player 1 does not need to be lexicographic, and Player 1 is left totally free to optimize the MDP that it faces at each iteration of the algorithm in the most efficient way.

Let us compare in terms of number of pivotings the following three algorithms:

$\mathbf{M_1}$: Algorithm 3.1.1, in which in step 2 Player 1 pivots with respect to the variable with the minimum reduced cost until it finds its own Blackwell optimal strategy.

$\mathbf{M_2}$: Algorithm 3.1.1, in which in step 2 Player 1 pursues a lexicographic search, pivoting iteratively with respect to the *first* non-basic variable with a negative (in the field $F(\mathbb{R})$) reduced cost. This method is analogous to the one shown in [48], but in the field $F(\mathbb{R})$.

$\mathbf{M_3}$: Algorithm 3.1.2.

The results are shown in Tables 1 and 2. The simulations were carried out on 10^4 randomly generated stochastic games with 4 states, 2 for Player 1 and 2 for Player 2. In each state 5 actions are available for the controlling player.

	n. pivotings
M_1	40.59
M_2	41.87
M_3	24.93

Table 3.1: Average number of pivotings for the 3 methods.

$<$ (%)	M_1	M_2	M_3
M_1	-	52.85	18.57
M_2	42.18	-	15.26
M_3	80.05	82.75	-

Table 3.2: M_i (row) $<$ M_j (column) when, fixing the game, the number of pivotings in M_i is strictly smaller than the number of pivoting in M_j.

It is evident that Algorithm M_3 is much faster than the other two. In our numerical experiment, M_3 never cycled. The difference between M_1 and M_2 is due to the more efficient simplex method used by Player 1 in M_1.

3.1.8 Transient games

Let $p_t(s'|s)$ be the probability that the process is in state s' at time t given that s is the initial state. Let us give the definition of transient games.

Definition 9. A stochastic game is transient if and only if $\sum_{t=0}^{\infty} \sum_{s' \in S} p_t(s'|s, \mathbf{f}, \mathbf{g})$ is finite for all $s \in S$ and for any pair of stationary strategies $(\mathbf{f}, \mathbf{g})$.

Here we present the result of this section.

Theorem 3.5. *The uniform optimal strategies $(\mathbf{f}^*, \mathbf{g}^*)$ for a transient stochastic game with perfect information are optimal in the undiscounted criterion, i.e. $\overline{\beta} = 1$, as well.*

Proof. The uniform optimal strategies are still optimal when $\overline{\rho} \downarrow 0$, since the reduced costs of the tableaux built at the end of Algorithm 3.1.1 are non-negative when $\overline{\rho} \downarrow 0$. We know from [46] that, for transient stochastic games, the reward associated to each pair of stationary strategies $(\mathbf{f}, \mathbf{g})$ is finite. By invoking Abel's Theorem on power series [53], we claim that the reward associated to any stationary $(\mathbf{f}, \mathbf{g})$ tends to the undiscounted reward when $\overline{\rho} \downarrow 0$. Hence, the saddle-point relation (3.3) is still valid when $\overline{\rho} = 0$ and $(\mathbf{f}^*, \mathbf{g}^*)$ are optimal in the undiscounted criterion as well. □

3.2 Cooperative Markov Decision Processes

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 MDPs, otherwise 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 section 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.

This section is organized as follows. Subsection 3.2.1 is a short survey on non-cooperative and cooperative multi-agent MDPs. Following the lines of Petrosjan's work, in Subsection 3.2.2 we propose a stationary stage-wise CPDP for cooperative discounted MDPs (MDP-CPDP). In Subsection 3.2.3 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 Subsection 3.2.4 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 Subsection 3.2.5 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. Subsection 3.2.6 deals with a special case of our model, entailing that the transition probabilities among the states do not depend on the players' strategies. In Subsection 3.2.7 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 $\text{val}(\mathbf{A})$ 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 $\{\sum_i c_i : \; c_i \in C_i, \; \forall\, i\}$.

3.2.1 Discounted Cooperative Markov Decision Processes

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

$$\mathbf{\Phi}_\Lambda^{(\beta)}(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}) = \sum_{t=0}^{\infty} \beta^t \mathbf{P}^t(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda})\, \mathbf{r}_\Lambda(\mathbf{f}_\Lambda, \mathbf{f}_{\mathcal{C}\setminus\Lambda}), \tag{3.7}$$

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 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

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

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:

$$\begin{aligned} v^{(\beta)}(\mathcal{C},\Gamma_s) &= \Phi_{\Lambda}^{(\beta)}(s,\mathbf{f}_{\mathcal{C}}^{(\beta)*}) \\ \mathbf{f}_{\mathcal{C}}^{(\beta)*} &= \operatorname*{argmax}_{\mathbf{f}_{\mathcal{C}}\in\mathbf{F}_{\mathcal{C}}} \boldsymbol{\Phi}_{\mathcal{C}}^{(\beta)}(\mathbf{f}_{\mathcal{C}}), \qquad \forall\,\beta\in[0;1) \end{aligned} \tag{3.8}$$

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.

$$\begin{aligned} v^{(\beta)}(\Lambda,\Gamma_s) &= \max_{\mathbf{f}_{\Lambda}\in\mathbf{F}_{\Lambda}} \min_{\mathbf{f}_{\mathcal{C}\setminus\Lambda}\in\mathbf{F}_{\mathcal{C}\setminus\Lambda}} \Phi_{\Lambda}^{(\beta)}(s,\mathbf{f}_{\Lambda},\mathbf{f}_{\mathcal{C}\setminus\Lambda}) \\ &= \min_{\mathbf{f}_{\mathcal{C}\setminus\Lambda}\in\mathbf{F}_{\mathcal{C}\setminus\Lambda}} \max_{\mathbf{f}_{\Lambda}\in\mathbf{F}_{\Lambda}} \Phi_{\Lambda}^{(\beta)}(s,\mathbf{f}_{\Lambda},\mathbf{f}_{\mathcal{C}\setminus\Lambda}), \qquad \forall\,\Lambda\subseteq\mathcal{C}/\{\emptyset\}. \end{aligned} \tag{3.9}$$

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

Definition 10 (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 3.1. 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

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

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:

$$\begin{cases} \sum_{i\in I} b_i\mathbf{c}'_j(i) = \widetilde{\mathbf{c}}_j & \forall\,j:\ \{j\}\in\Lambda \\ \sum_{j:\{j\}\in\Lambda}\mathbf{c}'_j(i) \le v(\Lambda,\Delta_i) & \forall\,i\in I \end{cases} \tag{3.10}$$

Let us prove that (3.10) admits a solution. Let $\epsilon_i \geq 0$, for all $i \in I$, be such that

$$\sum_{i \in I} \epsilon_i = \sum_{i \in I} b_i v(\Lambda, \Delta_i) - \sum_{j:\{j\} \in \Lambda} \widetilde{\mathbf{c}}_j \geq 0 \tag{3.11}$$

We write the following linear system

$$\begin{cases} \sum_{i \in I} b_i \mathbf{c}'_j(i) = \widetilde{\mathbf{c}}_j & \forall\, j : \{j\} \in \Lambda \\ b_i \sum_{j:\{j\} \in \Lambda} \mathbf{c}'_j(i) = b_i v(\Lambda, \Delta_i) - \epsilon_i & \forall\, i \in I \end{cases} \tag{3.12}$$

Evidently, any solution to (3.12) is also a solution to (3.10). Thanks to (3.11), the sum of the first $|\Lambda|$ equations of (3.12) equals the sum of the remaining $|I|$ equations. By discarding the last equation of (3.12) we get a linear system with $|\Lambda| + |I| - 1$ linearly independent equations in $|\Lambda||I| > |\Lambda| + |I| - 1$ unknowns. Hence, a solution to (3.12) 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 11 (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

$$\begin{cases} \sum_{i \in \mathcal{C}} \mathbf{x}_i = v^{(\beta)}(\mathcal{C}, \Gamma_s) \\ \sum_{i \in \Lambda} \mathbf{x}_i \geq v^{(\beta)}(\Lambda, \Gamma_s), \ \forall\, \Lambda \subset \mathcal{C}. \end{cases} \tag{3.13}$$

The strict Core $\mathbf{sCo}^{(\beta)}(\Gamma_s)$ is defined in (3.13), 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 3.2. 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 3.1, 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 3.3. 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 3.1 and from the linearity property of the Shapley value. □

3.2.2 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

$$\sum_{t=0}^{\infty} \sum_{s' \in S} p_t(s'|s, \mathbf{f}_\mathcal{C}) < \infty, \quad \forall s \in S,\ \mathbf{f}_\mathcal{C} \in \mathbf{F}_\mathcal{C}. \tag{3.14}$$

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 3.1. All the results presented in the current section, as well as the ones in Sections 3.2.3, 3.2.4, 3.2.7, can be easily extended to undiscounted transient MDPs, i.e. games for which equation (3.14) 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 (3.14).

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 12 (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 13 (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 $\mathbf{g}^{(\beta)} : S \rightarrow \mathbb{R}^P$ 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 14 (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 $\mathbb{R} \rightarrow \mathbb{R}^N$ defined by

$$\boldsymbol{\gamma}^{(\beta)}(s) = \sum_{s' \in S} \left[\delta_{s,s'} - \beta\, p(s'|s, \mathbf{f}_{\mathcal{C}}^{(\beta)*})\right] \overline{\mathbf{T}}^{(\beta)}(\Gamma_{s'}), \qquad \forall\, s \in S. \tag{3.15}$$

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

3.2.3 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 (3.15) 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 3.2.1 (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 3.6. *The MDP-CPDP $\gamma^{(\beta)}$, defined in (3.15), fulfills the stage-wise efficiency Property 3.2.1, for all $\beta \in [0;1)$.*

Proof. The global optimum strategy $\mathbf{f}_{\mathcal{C}}^{(\beta)*}$ is pure, since the optimization problem (3.8) that it solves can be formulated as a MDP (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 (3.15) 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 3.2.2 (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 3.7. *The MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}(s) \in \mathbb{R}^P$, defined in (3.15) is the unique stationary CPDP that satisfies the terminal fairness Property 3.2.2, 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 (3.15) in the equation above, we find that $\boldsymbol{\gamma}_i^{(\beta)}$ defined in (3.15) 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 (3.15) also fulfills a *terminal efficiency* property, i.e.

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

3.2.4 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 3.2.2 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 3.2.3 (Time consistency). Set $\beta\in[0;1)$. The CPDP $\mathbf{g}^{(\beta)}$ in (3.15) 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 $n-1$,

$$E\left[\sum_{t=n}^{\infty}\beta^t\mathbf{g}^{(\beta)}(S_t,\mathbf{h}_{t-1})\Big|\mathbf{h}_{n-1}\right]\in\beta^n\mathbf{T}^{(\beta)}\left(\sum_{s'\in S}p(s'|S_{n-1}=\overline{s},\mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right), \tag{3.16}$$

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 3.8. *The stationary MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies the time consistency Property 3.2.3 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 (3.16) as

$$E\left[\sum_{t=0}^{\infty}\beta^t\boldsymbol{\gamma}^{(\beta)}(S_{t+n})\Big|S_{n-1}=\overline{s}\right]\in\mathbf{T}^{(\beta)}\left(\sum_{s'\in S}p(s'|\overline{s},\mathbf{f}_{\mathcal{C}}^{(\beta)*})\Gamma_{s'}\right). \tag{3.17}$$

Let us rewrite now equation (3.15), for all $s\in S$, as

$$\overline{\mathbf{T}}^{(\beta)}(\Gamma_s)=\boldsymbol{\gamma}^{(\beta)}(s)+\beta\sum_{s'\in S}p(s'|s,\mathbf{f}_{\mathcal{C}}^{(\beta)*})\overline{\mathbf{T}}^{(\beta)}(\Gamma_{s'}), \tag{3.18}$$

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 (3.18), 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 3.2 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 3.3 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, (3.17) is verified, and the thesis is proved. □

3.2.5 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 3.2.1.

Condition 3.2.1 (Maxmin coalition values). The coalition value $v^{(\beta)}(\Lambda, \Gamma_s)$ is computed as the max-min expression in (3.9).

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 (3.9), 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 3.2.4 (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 3.11. *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 MDP having $\mathbf{\Phi}_{\mathcal{C}}^{(\beta)}$ 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

$$y^{(\beta_{n_k})}(s,\mathbf{f}_{\mathcal{C}}) = \Phi_{\mathcal{C}}^{(\beta_{n_k})}(s,\underline{\mathbf{f}}_{\mathcal{C}}^*) - \Phi_{\mathcal{C}}^{(\beta_{n_k})}(s,\mathbf{f}_{\mathcal{C}}) \geq\ 0, \qquad \forall\, k \in \mathbb{N}. \tag{3.19}$$

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 (3.19), 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 $\Phi_{\mathcal{C}}^{(\beta)}(s,\underline{\mathbf{f}}_{\mathcal{C}}^*)$ 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:

$$\Theta_s :\ \left\{\mathbf{x} \in \mathbb{R}^P :\ \sum_{i\in\mathcal{C}} \mathbf{x}_i = \sum_{i\in\mathcal{C}} \mathbf{r}_i(s,\underline{\mathbf{f}}_{\mathcal{C}}^*)\right\}, \tag{3.20}$$

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

Corollary 3.1. *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 3.6 and from Lemma 3.11. □

Here we present a useful result.

Lemma 3.12. *Let $\mathbf{T} \equiv \mathbf{Sh}$. Under Condition 3.2.1,* $\lim_{\beta\downarrow 0}\boldsymbol{\gamma}^{(\beta)}(s) = \mathbf{Sh}^{(0)}(\Gamma_s) \equiv \mathbf{Sh}(\Omega_s)$.

Proof. Recall the expression (3.15) of $\boldsymbol{\gamma}^{(\beta)}$, that we rewrite as

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

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)

$$|\operatorname{val}(\mathbf{B}) - \operatorname{val}(\mathbf{C})| \leq \max_{i,j}|\mathbf{B}_{i,j} - \mathbf{C}_{i,j}| \tag{3.21}$$

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

$$v^{(\beta)}(\Lambda, \Gamma_s) = \text{val}\Bigg(\Big[\sum_{i \in \Lambda} \mathbf{r}_i(s, a_\Lambda, a_{\mathcal{C}\setminus\Lambda}) + \dots \\ + \beta \sum_{s' \in S} p(s'|s, a_\Lambda, a_{\mathcal{C}\setminus\Lambda})\, v^{(\beta)}(\Lambda, \Gamma_{s'}) \Big]_{a_\Lambda=1, a_{\mathcal{C}\setminus\Lambda}=1}^{m_\Lambda(s), m_{\mathcal{C}\setminus\Lambda}(s)} \Bigg), \tag{3.22}$$

where $a_\Lambda \in A_\Lambda(s)$ and $a_{\mathcal{C}\setminus\Lambda} \in A_{\mathcal{C}\setminus\Lambda}(s)$. Thus, from (3.21,3.22) we can say that, for all $\Lambda \subseteq \mathcal{C}$,

$$|v^{(\beta)}(\Lambda, \Gamma_s) - v^{(0)}(\Lambda, \Gamma_s)| \leq \max_{a_\Lambda, a_{\mathcal{C}\setminus\Lambda}} \left| \beta \sum_{s' \in S} p(s'|s, a_\Lambda, a_{\mathcal{C}\setminus\Lambda})\, v^{(\beta)}(\Lambda, \Gamma_{s'}) \right| \\ \leq \frac{\beta}{1-\beta} M$$

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 3.2.2 (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 3.2.2 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 (3.20), 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 3.2. *Suppose that the stage-wise strict convexity Condition 3.2.2 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 3.9. *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 3.2.1 and 3.2.2, the greedy players satisfaction Property 3.2.4 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 3.2 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 3.2.2. Moreover, from Corollary 3.1 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 3.12 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 3.2.2, for all $\beta \in [0; \delta)$, all the greedy players are content with the stage-wise allocation as well.

3.2.6 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 3.2.3. The actions taken by players in state s do not influence the transition probabilities from state s, i.e. $p(s'|s, a_1, \dots, a_P) = p(s'|s)$, for all $a_i \in A_i(s)$ and for each $s, s' \in S$.

Like in Section 3.2.5, 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 3.2.1 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 3.2.3, 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 3.2.4 and whose discounted long run sum belongs to the Core of each long run game Γ_s, $s \in S$.

Theorem 3.10. *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 3.2.3, if* $\boldsymbol{\gamma}^{(\beta)}$ *fulfills the greedy players satisfaction Property 3.2.4 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

$$\begin{bmatrix} \mathcal{V}(\Gamma_{s_1}, \Lambda) \\ \vdots \\ \mathcal{V}(\Gamma_{s_N}, \Lambda) \end{bmatrix} = (\mathbf{I} - \beta\mathbf{P})^{-1} \begin{bmatrix} \mathcal{V}(\Omega_{s_1}, \Lambda) \\ \vdots \\ \mathcal{V}(\Omega_{s_N}, \Lambda) \end{bmatrix}, \qquad \forall \Lambda \subseteq \mathcal{C}. \tag{3.23}$$

Since the matrix $(\mathbf{I} - \beta\mathbf{P})^{-1}$ is non negative, the thesis follows straighforward from Proposition 3.2. □

In Section 3.2.5 we showed that, when the transition probabilities among the states depend on the players' actions, a MDP-CPDP fulfills the greedy players satisfaction Property 3.2.4 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 3.2.4 for *all* $\beta \in [0; 1)$.

The reader should also notice that the converse of Theorem 3.10 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 3.10 for the Shapley value. The proof follows straightforward from (3.23) and from Proposition 3.3.

Corollary 3.3. *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 3.2.3,* $\boldsymbol{\gamma}^{(\beta)}(s) = \mathbf{Sh}(\Omega_s)$*, 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.

3.2.7 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 3.2.5 (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 3.2.5 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.

n-tuple step cooperation maintenance

We now generalize Property 3.2.5, 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 3.2.6 (*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 (3.15), 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 3.11. *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 3.2.6 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 (3.15), that can be rewritten as:

$$\boldsymbol{\gamma}_i^{(\beta)} = \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \overline{\mathbf{T}}_i^{(\beta)}, \quad \forall\, i \in \mathcal{C} \tag{3.24}$$

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 3.2.6 as

$$\begin{cases} \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \sum_{i\in\Lambda} \overline{\mathbf{T}}_i^{(\beta)} \geq \left[\mathbf{I} - \left[\beta \mathbf{P}^{*(\beta)}\right]^n\right] \mathbf{v}^{(\beta)}(\Lambda), \quad \forall\, \Lambda \subset \mathcal{C} \\ \sum_{i\in\mathcal{C}} \overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}) \end{cases} \tag{3.25}$$

where the second relation in (3.25) 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 (3.25) as

$$\begin{cases} \sum_{i\in\Lambda} \overline{\widetilde{\mathbf{T}}}_i^{(\beta,n)} \geq \widetilde{\mathbf{v}}^{(\beta,n)}(\Lambda),\ \forall\, \Lambda \subset \mathcal{C} \\ \sum_{i\in\mathcal{C}} \overline{\widetilde{\mathbf{T}}}_i^{(\beta,n)} = \widetilde{\mathbf{v}}^{(\beta,n)}(\mathcal{C}) \end{cases} \tag{3.26}$$

where

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

Since the relations in the systems of inequalities in (3.26) 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 3.11 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 3.12. *Set* $\beta \in [0; 1)$. *If the MDP-CPDP* $\boldsymbol{\gamma}^{(\beta)}$ *satisfies the single step cooperation maintenance Property 3.2.5, then it satisfies the* n*-tuple step cooperation maintenance Property 3.2.6, for all* $n > 1$.

Proof. Let $\boldsymbol{\gamma}^{(\beta)}$ be defined in (3.24), where $\overline{\mathbf{T}}^{(\beta)}$ satisfies the single step cooperation maintenance Property 3.2.5, i.e., from (3.25),

$$\begin{cases} \beta\mathbf{P}^{*(\beta)}\left[\sum_{i\in\Lambda}\overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda)\right] \geq \sum_{i\in\Lambda}\overline{\mathbf{T}}_i^{(\beta)} - \mathbf{v}^{(\beta)}(\Lambda), \quad \forall\,\Lambda\subset\mathcal{C} \\ \sum_{i\in\mathcal{C}}\overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}) \end{cases} \tag{3.27}$$

By iteratively left multiplying by the nonnegative matrix $\beta\mathbf{P}^{*(\beta)}$ both sides of the first relation in (3.27), 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. □

Core selection criterion

In the following we prove that the single step cooperation maintenance Property 3.2.5 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 3.4. *Set $\beta\in[0;1)$. If a MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ satisfies the single step cooperation maintenance Property 3.2.5, then*

$$E\left[\sum_{t\geq 0}\beta^t\boldsymbol{\gamma}^{(\beta)}(S_t)|S_0 = s\right] \in \mathbf{Co}^{(\beta)}(\Gamma_s), \quad \forall\, s\in S. \tag{3.28}$$

Proof. Let us define $\boldsymbol{\gamma}^{(\beta)}$ as in (3.24). We reformulate (3.28) as

$$\begin{cases} \sum_{i\in\Lambda}\overline{\mathbf{T}}_i^{(\beta)} \leq \mathbf{v}^{(\beta)}(\Lambda), \qquad \forall\,\Lambda\subset\mathcal{C}, \\ \sum_{i\in\mathcal{C}}\overline{\mathbf{T}}_i^{(\beta)} = \mathbf{v}^{(\beta)}(\mathcal{C}). \end{cases} \tag{3.29}$$

Since $\boldsymbol{\gamma}^{(\beta)}$ satisfies Property 3.2.5, then (3.25) is verified, with $n=1$. By left multiplying each set of inequalties in (3.25) by the nonnegative matrix $(\mathbf{I}-\beta\mathbf{P}^{*(\beta)})^{-1}$, we obtain the system of inequalities in (3.29). □

In this section we showed how appealing the single step cooperation maintenance property is. For Theorem 3.12, 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 3.4 suggests that the point of the Core $\overline{\mathbf{T}}^{(\beta)}$ used to compute the MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ in equation (3.15) should be picked such that $\overline{\mathbf{T}}^{(\beta)}$ also satisfies the single step cooperation maintenance property. In this sense, Property 3.2.5 is also a *Core selection* criterion.

Counterexample for the converse of Corollary 3.4

It is natural to ask whether the converse of Corollary 3.4 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 (3.29), then not necessarily the single step cooperation maintenance Property 3.2.5 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 Table 3.3.

	(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 3.3: 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 3.4, it is sufficient to find a specific long run allocation $\overline{\mathbf{T}}^{(0.8)}$ such that

$$[\overline{\mathbf{T}}_1^{(0.8)}(s_k)\ \overline{\mathbf{T}}_2^{(0.8)}(s_k)] \in \mathbf{Co}^{(0.8)}(\Gamma_{s_k}), \qquad k = 1, 2, 3, 4, \tag{3.30}$$

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

$$\begin{cases} \overline{\mathbf{T}}_1^{(0.8)} \geq \mathbf{v}^{(0.8)}(\{1\}) \\ \overline{\mathbf{T}}_2^{(0.8)} \geq \mathbf{v}^{(0.8)}(\{2\}) \\ \overline{\mathbf{T}}_1^{(0.8)} + \overline{\mathbf{T}}_2^{(0.8)} = \mathbf{v}^{(0.8)}(\{1,2\}) \end{cases} \tag{3.31}$$

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

$$\begin{aligned} \widetilde{\overline{\mathbf{T}}}_i^{(0.8)} &= \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \overline{\mathbf{T}}_i^{(0.8)} \\ \widetilde{\mathbf{v}}^{(0.8)}(\{i\}) &= \left[\mathbf{I} - \beta \mathbf{P}^{*(\beta)}\right] \mathbf{v}^{(0.8)}(\{i\}) \quad i = 1, 2. \end{aligned} \tag{3.32}$$

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 (3.31). 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

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

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 3.4 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 (3.30), in about the 99.45% of the trials the converse of Corollary 3.4 was verified.

Strictly convex single stage games

In the spirit of Section 3.2.5, we show that the strict convexity Condition 3.2.2 on the single stage games ensures the MDP-CPDP $\boldsymbol{\gamma}^{(\beta)}$ to satisfy Property 3.2.5 for all discount factors small enough.

Theorem 3.13. *Suppose that the strict convexity Condition 3.2.2 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 3.2.5 is valid for all* β *close enough to 0.*

Proof. Thanks to the linearity property of coalition values (see Proposition 3.1) we can reformulate Property 3.2.5 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 (3.15), 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 3.12), 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. □

3.3 Confidence Intervals of Shapley-Shubik Power Index in Markovian Games

Cooperative game theory is a powerful tool to analyze, predict and, especially, influence the interactions among several players capable to stipulate deals and form subcoalitions in order to pursue a common interest. Under the assumption that the grand coalition, comprising all the players, is formed, it is a delicate issue to share the payoff earned by the grand coalition among its participants.

Introduced by Lloyd S. Shapley in its seminal paper [54], the Shapley value is one of the best known payoff allocation rules in a cooperative game with TU. It is the only allocation procedure fulfilling three reasonable conditions of symmetry, additivity and dummy player compensation (see [54] for details). Moreover, it always exists under a superadditive assumption on the coalition values. The significance of the Shapley value is witnessed by the breadth of its applications, spanning from pure economics [55] to Internet economics [56–58], politics [59], and telecommunications [60].

The concept of Shapley value was successfully applied to weighted voting games as well. In this case, it is commonly referred to as Shapley-Shubik power index [61]. Such games imply that the coalition values are binary. Each player possesses a different amount of resources and a coalition is effective, i.e. its value is 1, whenever the sum of the resources shared by its participants is higher than a certain quota; otherwise, its value is 0. The Shapley-Shubik index proves to be particularly suitable to assess *a priori* the power of the members of a legislation committee, and has many applications to politics (see [62] for an overview).

However, politics or economics is more like a process of continuing negotiation and bargaining. This motivates the introduction of dynamic cooperative game theory. In this work we consider that the game is not played one-shot but rather over an infinite horizon: there exists a finite set of static cooperative games that come one after the other, following a discrete-time homogeneous Markov process. We call this interaction model repeated over time as Markovian game. Our Markovian game model arises naturally in all situations in which several individuals interact and cooperate over time, and an exogenous Markov process influences the value of each coalition, and consequently also the power of each player within coalitions. A very similar model, but with non transferable utilities, was considered in [63]. Our model can also be viewed as a particular case of the cooperative MDP described in Section 3.2, or in [64], in which the transition probabilities among the states do not depend on the players' actions.

We take into account two criteria to sum over time the payoffs earned in each single stage game, specifically the average and the discount criterion. In this section we extend the work in [65] to Markovian games. In [65], the authors considered a weighted voting static game and proved that any deterministic algorithm which approximates one component of the Banzhaf index with accuracy better than $c/\sqrt{P}$, where $c > 0$ and P is the number of players, needs $\Omega(2^P/\sqrt{P})$ queries. Hence, when P grows large, it is crucial to find a suitable way to approximate the power index with a manageable number of queries. Hence, in [65] a confidence interval for Banzhaf index and Shapley-Shubik power index in weighted voting games has been developed, based on Hoeffding's inequality. In this section, we assume that the estimator agent knows the transition probabilities among the states. We first show that it is still beneficial to utilize a randomized approach to approximate the *S*hapley-*S*hubik index in *M*arkovian games (SSM) for a number of players P sufficiently high. Thus, we propose three methods to compute a confidence interval for the SSM, that also apply to the Shapley value of *any* Markovian game. Then, we will essentially demonstrate that, asymptotically in the number of steps of the Markov chain and by exploiting the Hoeffding's inequality, the estimator agent does not need to have access to the coalition values in all the states at the same time. Indeed, it suffices for the estimator agent to learn the coalition values in each state along the course of the game to "well" approximate SSM.

Let us overview the content of this section. We provide some useful definitions, background results, and motivations of our dynamic model in Section 3.3.1. In Section 3.3.2 we study the trade-off between complexity and accuracy of deterministic algorithms approximating SSM. An exponential number of queries is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy. Motivated by this, we propose three different randomized approaches to compute a confidence interval for SSM. Their complexity does not even depend on the number of players. Such approaches also hold for the classic *Sh*apley value of any cooperative *M*arkovian game (ShM). In Section 3.3.3 we provide the expression of our first confidence interval, SCI, which relies on the static assumption that the estimator agent has access to the coalition values in all the states at the same time, even before the Markov process initiates. Although SCI relies on an impractical assumption, it is still a valid benchmark for the performance of the approaches yielding the confidence intervals described in Sections 3.3.4 and 3.3.4, dubbed DCI1 and DCI2 respectively. DCI1 and DCI2 also hold under the more realistic dynamic assumption that the estimator agent learns the value of coalitions along the course of the game. Then, we propose a straightforward way to optimize the tightness of DCI1. In Section 3.3.5 we compare the three proposed approaches in terms of tightness of the confidence interval. Finally, in Section 3.3.6 we provide a trade-off complexity/accuracy of our randomized algorithm, holding for any cooperative Markovian game.

We point out that our results for SSM are also valid for ShM in simple Markovian games, i.e. TU cooperative Markovian games with binary coalition values in each state. Moreover, the extension of our approaches to Banzhaf index [66] is straightforward.

In this section we adopt the following notation. If $\mathbf{a}$ is a vector, then $\mathbf{a}_i$ is its i-th component. If A is a random variable, then A_t is its t-th realization. Given a set S, $|S|$ is its cardinality. The expression $b^{(s)}$ indicates that the quantity b, standing possibly for Shapley value, Shapley-Shubik index, coalition value, feasibility region etc., is related to the static game played in state s. The expression $\Pr(B)$ stands for the probability that the event B is verified. The indicator function is written as $\mathbb{I}(.)$.

3.3.1 Markovian Model and Background results

In this section we consider cooperative Markovian games with TU. Let $\mathcal{P} = \{1, \dots, P\}$ be the grand coalition of players. We have a finite set of states $S = \{s_1, \dots, s_{|S|}\}$. In state s, each coalition $\Lambda \subseteq \mathcal{P}$ can ensure for itself the value $v^{(s)}(\Lambda)$, that can be shared in any manner among the players under the TU assumption. Hence, in each state $s \in S$ the game $\Psi^{(s)} \equiv (\mathcal{P}, v^{(s)})$ is played among P players. Let $\mathcal{V}^{(s)}(\Lambda)$ be the half-space of all feasible allocations for coalition Λ in the TU game $\Psi^{(s)}$, i.e. the set of real $|\Lambda|$-tuple $\mathbf{a} \in \mathbb{R}^{|\Lambda|}$ such that $\sum_{i=1}^{|\Lambda|} \mathbf{a}_i \leq v^{(s)}(\Lambda)$. We suppose that the coalition values are superadditive, i.e.

$$v^{(s)}(\Lambda_1 \cup \Lambda_2) \geq \; v^{(s)}(\Lambda_1) + v^{(s)}(\Lambda_2), \qquad \forall\, \Lambda_1 \cap \Lambda_2 = \emptyset.$$

The succession of the states is a discrete-time homogeneous Markov chain, whose transition probability matrix is $\mathbf{P}$. Let $\mathbf{x}^{(s)} \in \mathbb{R}^P$ be a payoff allocation among the players in the single stage game $\Psi^{(s)}$. Under the β-discounted criterion, where $\beta \in [0;1)$, the discounted allocation in the Markovian dynamic game Γ_{s_k}, starting from state s_k, can be expressed as

$$\sum_{t=0}^{\infty} \beta^t \sum_{i=1}^{|S|} p_t(s_i|s_k)\, \mathbf{x}^{(s_i)} = \sum_{i=1}^{|S|} \boldsymbol{\nu}_i^{(\beta)}(s_k)\, \mathbf{x}^{(s_i)}$$

where $p_t(s_i|s_k)$ is the probability for the process to be in state s_i after t steps when the initial state is s_k, and $\boldsymbol{\nu}^{(\beta)}(s_k)$ is the k-th row of the nonnegative matrix $(\mathbf{I} - \beta\mathbf{P})^{-1}$. Under the average criterion, if the transition probability matrix $\mathbf{P}$ is *irreducible*, then the allocation in the long run game Γ_{s_k} can be written as

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{i=1}^{|S|} p_t(s_i|s_k)\, \mathbf{x}^{(s_i)} = \sum_{i=1}^{|S|} \boldsymbol{\pi}_i\, \mathbf{x}^{(s_i)}$$

where $\boldsymbol{\pi}$ is the stationary distribution of the matrix $\mathbf{P}$.

We define $\mathcal{V}(\Lambda, \Gamma_s)$ as the set of feasible allocations in the long run game Γ_s for coalition Λ, i.e. the Minkowski sum:

$$\mathcal{V}(\Lambda, \Gamma_s) \equiv \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, \mathcal{V}^{(s_i)}(\Lambda).$$

where $\boldsymbol{\sigma}_i(s) \equiv \boldsymbol{\nu}_i^{(\beta)}(s)$ if the β-discounted criterion is adopted, and $\boldsymbol{\sigma}_i(s) \equiv \boldsymbol{\pi}_i$ under the average criterion.

Proposition 3.4. $\mathcal{V}(\Lambda, \Gamma_s)$ is equivalent to the set $\mathcal{A}$ of real $\mathbb{R}^{|\Lambda|}$-tuples $\mathbf{a}$ such that $\sum_{i=1}^{|\Lambda|} \mathbf{a}_i \leq v(\Lambda, \Gamma_s)$, where $v(\Lambda, \Gamma_s) = \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, v^{(s_i)}(\Lambda)$, for all $s \in S$, $\Lambda \subseteq \mathcal{P}$.

Thanks to Proposition 3.4, it is legitimate to define $v(\Lambda, \Gamma_s)$ as the value of coalition $\Lambda \subseteq \mathcal{P}$ in the long run game Γ_s.

Let us now define the Shapley value in static games [54].

Definition 3.1. The *Shapley value* $\mathbf{Sh}^{(s)}$ in the static game played in state $s \in S$ is a real P-tuple whose j-th component is the payoff allocation to player j:

$$\mathbf{Sh}_j^{(s)} = \sum_{\Lambda \subseteq \mathcal{P}/\{j\}} \frac{|\Lambda|!(P-|\Lambda|-1)!}{P!} \big[v^{(s)}(\Lambda \cup \{j\}) - v^{(s)}(\Lambda)\big].$$

We are ready to define the *Sh*apley value in the *M*arkovian game Γ_s, $\mathrm{ShM}(\Gamma_s)$, that can be expressed, thanks to Proposition 3.4 and to the standard linearity property of the Shapley value, as

$$\mathrm{ShM}_j(\Gamma_s) = \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, \mathbf{Sh}_j^{(s_i)}, \quad \forall\, s \in S,\; 1 \leq j \leq P. \tag{3.33}$$

In the next sections we will exploit Hoeffding's inequality [67] to derive basic confidence intervals for the Shapley value of Markovian games.

Theorem 3.14 (Hoeffding's inequality)**.** *Let* $A_1, \ldots, A_n$ *be* n *independent random variables, where* $A_i \in [a_i, b_i]$ *almost surely. Then, for all* $\epsilon > 0$*,*

$$\Pr\left(\sum_{i=1}^{n} A_i - E\left[\sum_{i=1}^{n} A_i\right] \geq n\,\epsilon\right) \leq\; 2\exp\left(-\frac{2n^2\,\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

In this work, several results are shown in the case of weighted voting Markovian games, that we define in the following.

Definition 3.2. A *weighted voting Markovian game* is a Markovian game in which each single stage game $\Psi^{(s)}$ is associated to the triple $(P, T^{(s)}, \mathbf{w}^{(s)})$, where $1, \ldots, P$ are the players, $\mathbf{w}^{(s)} \in \mathbf{R}^P$ is the set of weights, and $T^{(s)}$ is a threshold. The binary coalition values $v^{(s)}$ in state s are such that $v^{(s)}(\Lambda) = 1$ whenever $\sum_{i\in\Lambda} \mathbf{w}_i^{(s)} \geq T^{(s)}$ and $v^{(s)}(\Lambda) = 0$ whenever $\sum_{i\in\Lambda} \mathbf{w}_i^{(s)} < T^{(s)}$.

We say that player i is *critical* for coalition $\Lambda \subseteq \mathcal{P}\backslash\{i\}$ in state s if $v^{(s)}(\Lambda \cup \{i\}) - v^{(s)}(\Lambda) = 1$.

The concept of Shapley value applied to weighted voting static games is referred to as *S*hapley-*S*hubik index (SS) [61], and its formulation is the same as in Definition 3.1. Technically, weighted voting games are not TU cooperative game, since the value of a coalition does not represent a payoff to be shared among the players, but rather it indicates the effectiveness of a coalition. Anyway, even for these games, we will still assume that the coalition values in the long run game Γ possess the linearity property of Proposition 3.4. In other words, $v(\Lambda, \Gamma_s)$ can be interpreted as the expected effectiveness of coalition Λ in the long run game Γ_s. Accordingly, the SSM can be written as the expected - discounted or average - sum of the Shapley-Shubik indices over time, i.e.

$$\mathrm{SSM}_j(\Gamma_s) = \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, \mathrm{SS}_j^{(s_i)}, \quad \forall\, s \in S,\; 1 \leq j \leq P. \tag{3.34}$$

where $\mathrm{SSM}_j(\Gamma_s)$ is the Shapley-Shubik power index in the Markovian game Γ_s for player j and $\mathrm{SS}_j^{(s_i)}$ is the Shapley-Shubik index for player j in state s_i.

In the following we derive some properties of randomized algorithms to approximate SSM especially for weighted voting Markovian games. In the static case, fixed a state s, to each member i the weight $\mathbf{w}_i^{(s)}$ is assigned and a coalition Λ is effective, i.e. $v^{(s)}(\Lambda) = 1$, whenever the sum of its weights $\sum_{i \in \Lambda} \mathbf{w}_i^{(s)}$ exceeds a quota $T^{(s)}$. We address a situation in which repeated ballots are cast over times and the weights assigned to each member, as well as the quota T, change according to a Markov chain. This reflects the fluctuation of the clout that each member has over the vote over time, that depends on factors independent of the voting outcome. The expected sum of Shapley-Shubik indices computed in each stage of the process is the most natural way to assess the power of members over the whole course of voting procedure.

Of course, all the results that we will derive for weighted voting games are also valid in the case of simple Markovian games, i.e. TU cooperative Markovian games with binary coalition values in each state.

3.3.2 Complexity of deterministic algorithms

Since the *exact* computation of the Shapley value - or, equivalently, of the Shapley-Shubik index - involves the calculation of the incremental asset brought by a player to each coalition, then its complexity is proportional to the number of such coalitions, i.e. 2^{P-1}, under oracle access to the characteristic function. In this section we evaluate the complexity that any *deterministic* algorithm needs to *approximate* the Shapley-Shubik index in a weighted voting Markovian game.
Before starting the analysis, let us introduce some ancillary concepts. We mean by *game instance* a specific collection of coalition values. In this section, we implicitly assume that all the algorithms considered - deterministic or randomized - aim at approximating the Shapley value for player j, without loss of generality. Let us clarify our notion of query.

Definition 3.3. A *query* of an algorithm - deterministic or randomized - consists in the evaluation of the marginal contribution of player j to a coalition $\Lambda \subseteq \mathcal{P} \backslash \{i\}$, i.e. $v(\Lambda \cup \{i\}) - v(\Lambda)$.

Now we define the notion of accuracy of a deterministic algorithm.

Definition 3.4. Let us assume that the Shapley-Shubik index for player j in the game Γ_s is $\mathrm{SSM}_j(\Gamma_s) = a$. Let ALG be a *deterministic* algorithm needing a fixed number of queries. We say that ALG has an *accuracy* of at least $d > 0$ whenever, for all the game instances, ALG always answers $\mathrm{SSM}_j(\Gamma_s) \in [a - d; a + d]$.

We will first show that an exponential number of queries is necessary in order to achieve a polynomial accuracy for any deterministic algorithm aiming to approximate the Shapley-Shubik index in the static case. This is an extension of Theorem 3 in [65] to the Shapley-Shubik index.

Theorem 3.15. *Any deterministic algorithm computing one component of the Shapley-Shubik index in weighted voting static game in state* s *requires* $\Omega(2^P/\sqrt{P})$ *queries to achieve an accuracy of at least* $1/(2P)$*, for all* $s \in S$.

The proof of this theorem and the following corollary is omitted here. The interested reader can refer to [68]

From the previous results we derive the complexity of a deterministic algorithm computing the Shapley-Shubik index in a weighted voting Markovian game, as a function of the number of players P.

Corollary 3.5. *There exists* $c > 0$ *such that any deterministic algorithm computing one component of the Shapley-Shubik index in the weighted voting Markovian game* Γ_s *requires* $\Omega(2^P/\sqrt{P})$ *queries to achieve an accuracy of at least* c/P*, for all* $s \in S$.

The results of the current section clearly discourage from computing exactly or even approximating SSM with a deterministic algorithm when the number of players P is high. Motivated by this, in the next sections we will direct our attention towards randomized approaches to construct confidence intervals for SSM, whose complexity does not even depend on P.

3.3.3 Randomized static approach

In this section we will propose our first approach to compute a confidence interval for the Shapley value in Markovian games. The expression of the confidence interval that we will propose holds for the Shapley value of *any* Markovian game (ShM). Nevertheless, in the following sections we will provide some results holding specifically for the Shapley-Shubik index in the particular case of weighted voting Markovian games (SSM).
Let us first define our performance evaluator for a randomized algorithm.

Definition 3.5. Let us fix a probability of confidence $1-\delta$. The *accuracy of a randomized algorithm* approximating the SSM is the expected length of the confidence interval in which SSM lies with a probability of at least $1-\delta$.

In parallel, the reader learns the notion of accuracy of a deterministic algorithm from Definition 3.4.

We presuppose throughout the section that the transition probability matrix $\mathbf{P}$ is known by the estimator agent. In this section we also assume that the value of all coalitions in each single stage games are available *off-line* to the estimator agent.

Assumption 3.1. The estimator agent has access to all the coalition values in each state:

$$\{v^{(s)}(\Lambda),\ \forall\, \Lambda \subseteq \mathcal{P},\ s \in S\}$$

at the same time, before the Markovian game starts.

It is clear that, under Assumption 3.1, the estimator agent can perform an off-line randomized algorithm to approximate ShM.

Remark 3.2. Assumption 3.1 seems to be impractical for the intrinsic dynamics of the model we consider. Nevertheless, the randomized approach based on Assumption 3.1 that we next propose (SCI) will prove to be an insightful performance benchmark for two methods (DCI1 and DCI2) described in Section 3.3.4, based on a more realistic dynamic assumption.

First, let us find a formulation of the Shapley value in the Markovian game which is suitable for our purpose. Let X be the set of all the permutations of $\{1, \dots, P\}$. Let $\mathcal{C}_\chi(j)$ be the coalition of all the players whose index precedes j in the permutation $\chi \in X$, i.e.

$$\mathcal{C}_\chi(j) \equiv \{i : \chi(i) < \chi(j)\}. \tag{3.35}$$

We can write the Shapley value of the Markovian game Γ_s, both for the discount and for the average criterion, as

$$\begin{aligned}
\mathrm{ShM}_j(\Gamma_s) &= \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\,\mathbf{Sh}_j^{(s_i)} \\
&= \frac{1}{P!} \sum_{\chi \in X} \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s) \big[v^{(s_i)}(\mathcal{C}_\chi(j) \cup \{j\}) - v^{(s_i)}(\mathcal{C}_\chi(j)) \big] \\
&= E_\chi \left[\sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s) \big[v^{(s_i)}(\mathcal{C}_\chi(j) \cup \{j\}) - v^{(s_i)}(\mathcal{C}_\chi(j)) \big] \right],
\end{aligned}$$

where E_χ is the expectation over all the permutations $\chi \in X$, each having the same probability $1/P!$.

We now propose our first algorithm to compute a confidence interval for $\mathrm{ShM}_j(\Gamma_s)$, for each player j and initial state s. For each query, labeled by the index $k = 1, \dots, m$, let us select independently over a uniform distribution on X a permutation χ_k of $\{1, \dots, P\}$. Let us define $Z(j)$ as the random (over $\chi \in X$) variable

$$\begin{aligned}
Z(j) &\equiv \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s) \big[v^{(s_i)}(\mathcal{C}_\chi(j) \cup \{j\}) - v^{(s_i)}(\mathcal{C}_\chi(j)) \big] \\
&= v(\mathcal{C}_\chi(j) \cup \{j\}, \Gamma_s) - v(\mathcal{C}_\chi(j), \Gamma_s)
\end{aligned} \tag{3.36}$$

and let $Z_k(j)$ be the k-th realization of $Z(j)$. We remark that $Z(j)$ implies the computation of $|S|$ queries, one in each state. Thanks to Hoeffding's inequality, we can write that, for all $\epsilon > 0$,

$$\Pr\left(\left| \frac{1}{m} \sum_{k=1}^{m} Z_k(j) - \mathrm{ShM}_j(\Gamma_s) \right| \geq \epsilon \right) \leq 2 \exp\left(- \frac{2m\,\epsilon^2}{[\overline{y} - \underline{y}]^2} \right)$$

where

$$\overline{y} = \max_{\mathcal{C}\subseteq\mathcal{P}} \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\big[v^{(s_i)}(\mathcal{C}\cup\{j\}) - v^{(s_i)}(\mathcal{C})\big],$$

$$\underline{y} = \min_{\mathcal{C}\subseteq\mathcal{P}} \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\big[v^{(s_i)}(\mathcal{C}\cup\{j\}) - v^{(s_i)}(\mathcal{C})\big].$$

We remark that, in the case of weighted voting games, $\big[\overline{y}-\underline{y}\big]^2 \leq \big[\sum_{i=1}^{|S|}\boldsymbol{\sigma}_i(s)\big]^2$. Now we are ready to propose our first confidence interval, based on Assumption 3.1.

Static Confidence Interval 3.3.1 (**SCI**). Let $1 \leq j \leq P$, $s \in S$. Fix an integer n and set $\delta \in (0;1)$. Then, with probability of confidence $1-\delta$, $\mathrm{ShM}_j(\Gamma_s)$ belongs to the confidence interval

$$\left[\frac{1}{m}\sum_{k=1}^{m} Z_k(j) - \epsilon(m,\delta)\;;\;\frac{1}{m}\sum_{k=1}^{n} Z_k(j) + \epsilon(m,\delta)\right],$$

where

$$\epsilon(m,\delta) = \sqrt{\frac{[\overline{y}-\underline{y}]^2\log(2/\delta)}{2m}}. \tag{3.37}$$

In the case of weighted voting games, (3.37) becomes

$$\epsilon(m,\delta) = \sqrt{\frac{\Big[\sum_{i=1}^{|S|}\boldsymbol{\sigma}_i(s)\Big]^2\log(2/\delta)}{2m}}. \tag{3.38}$$

Under the average criterion, (3.38) can be written as $\epsilon(m,\delta) = \sqrt{\log(2/\delta)/[2m]}$.

Not surprisingly, the confidence interval SCI is analogous to the one found in [65] for static games. Indeed, the intrinsic dynamics of the game is surpassed by Assumption 3.1, for which the estimator has global knowledge of all the coalition values, even before the Markov process initiates. Therefore, from the estimator agent's point of view, there exists no conceptual difference between the approach in [65] and SCI, except for the complexity, which in the dynamic game increases by a factor $|S|$.

3.3.4 Randomized dynamic approaches

In this section we will propose two methods to compute a confidence interval for SSM, for which Assumption 3.1 on global knowledge of coalition values is no longer necessary. Indeed, the reader will notice that their conception naturally arises from the assumption that the estimator agent learns the coalition values in each single stage game while the Markov chain process unfolds, as formalized below.

Assumption 3.2. The state in which the estimator agent finds itself at each time step follows the same Markov chain process of the Markovian game itself. The estimator agent has local knowledge of the game that is being played, i.e. at step $t \geq 0$ the estimator agent has access only to the coalition values associated to the static game in the current state S_t.

Remark 3.3. The approaches described in this section can also be employed under Assumption 3.1. Indeed, any algorithm requiring the query on coalition values separately in each state can also be run under a static assumption.

In the following we still assume that the transition probability matrix $\mathbf{P}$ is known by the estimator agent. As in Section 3.3.3, the randomized approaches that we are going to introduce hold for the Shapley value of *any* Markovian game.

First dynamic approach

Now we propose our first randomized approach to compute a confidence interval for ShM, holding both under the static Assumption 3.1 and under the dynamic Assumption 3.2. Let $\chi \in X$ be, as in Section 3.3.3, a random permutation uniformly distributed on the set $\{1, \ldots, P\}$. Let us define $Y^{(s_i)}(j)$ as the random (over $\chi \in X$) variable associated to state s_i:

$$Y^{(s_i)}(j) \equiv \ v^{(s_i)}(\mathcal{C}_\chi(j) \cup \{j\}) - v^{(s_i)}(\mathcal{C}_\chi(j)). \tag{3.39}$$

Let us assume that $Y^{(s_i)}(j)$ has been queried n_i times in state s_i, and let $\sum_{i=1}^{|S|} n_i = n$. We can still exploit Hoeffding's inequality to say that, for all $\epsilon' > 0$,

$$\Pr\left(\left|\sum_{i=1}^{|S|}\frac{\boldsymbol{\sigma}_i(s)}{n_i}\sum_{t=1}^{n_i} Y_t^{(s_i)}(j) - \mathrm{ShM}_j(\Gamma_s)\right| \geq n\,\epsilon'\right) \leq \ \ldots$$

$$2\exp\left(-\frac{2[n\,\epsilon']^2}{\sum_{i=1}^{|S|}\boldsymbol{\sigma}_i^2(s)[\overline{x}(i)-\underline{x}(i)]^2/n_i}\right)$$

where, for all $i = 1, \ldots, |S|$,

$$\overline{x}(i) = \max_{\mathcal{C}\subseteq\mathcal{P}} \ v^{(s_i)}(\mathcal{C} \cup \{j\}) - v^{(s_i)}(\mathcal{C})$$

$$\underline{x}(i) = \min_{\mathcal{C}\subseteq\mathcal{P}} \ v^{(s_i)}(\mathcal{C} \cup \{j\}) - v^{(s_i)}(\mathcal{C})$$

We notice that, in the case of weighted voting games, $\overline{x}(i) = 1$ and $\underline{x}(i) = 0$ for all $i = 1, \ldots, |S|$. Set $\widetilde{\epsilon} = n\,\epsilon'$. Now we are ready to propose our second confidence interval for $\mathrm{ShM}_j(\Gamma_s)$, the first holding under Assumption 3.2.

Dynamic Confidence Interval 3.3.1 (**DCI1**). Let $1 \leq j \leq P$, $s \in S$. Fix the number of queries n and set $\delta \in (0; 1)$. Then, with probability of confidence $1-\delta$, $\mathrm{ShM}_j(\Gamma_s)$

belongs to the confidence interval

$$\left[\sum_{i=1}^{|S|}\frac{\boldsymbol{\sigma}_i(s)}{n_i}\sum_{t=1}^{n_i}Y_t^{(s_i)}(j)-\widetilde{\epsilon}(n,\delta)\;;\;\sum_{i=1}^{|S|}\frac{\boldsymbol{\sigma}_i(s)}{n_i}\sum_{t=1}^{n_i}Y_t^{(s_i)}(j)+\widetilde{\epsilon}(n,\delta)\right],$$

where

$$\widetilde{\epsilon}(n,\delta)=\sqrt{\frac{\log(2/\delta)}{2}\sum_{i=1}^{|S|}\frac{\boldsymbol{\sigma}_i^2(s)}{n_i}[\overline{x}(i)-\underline{x}(i)]^2}. \tag{3.40}$$

In the case of weighted voting games, (3.40) becomes

$$\widetilde{\epsilon}(n,\delta)=\sqrt{\frac{\log(2/\delta)}{2}\sum_{i=1}^{|S|}\frac{\boldsymbol{\sigma}_i^2(s)}{n_i}}\,. \tag{3.41}$$

Optimal sampling strategy

In this section we focus exclusively on *weighted voting Markovian games*. It is interesting to investigate the optimum number of times n_i^* in which the variable $Y^{(s_i)}(j)$ should be sampled in each state s_i, in order to minimize the length of the confidence interval DCI1, keeping the confidence probability fixed. We notice that, by fixing $1-\delta$, we can find the optimal values for $n_1,\ldots,n_{|S|}$ by setting up the following integer programming problem:

$$\begin{cases}\min\limits_{n_1,\ldots,n_{|S|}}\sum_{i=1}^{|S|}\boldsymbol{\sigma}_i^2(s)[\overline{x}^2(i)-\underline{x}^2(i)]/n_i\\ \sum_{i=1}^{|S|}n_i=n, \qquad n_i\in\mathbb{N}\end{cases} \tag{3.42}$$

Remark 3.4. If the static Assumption 3.1 holds, then the computation of the optimum values $n_1^*,\ldots,n_{|S|}^*$ in (3.42) is the only information we need to maximize the accuracy of DCI1, since the sampling is done off-line. Otherwise, if Assumption 3.2 holds, the estimator does not know in advance the succession of states hit by the process, hence it is crucial to plan a sampling strategy of the variable $Y^{(s_i)}(j)$ along the Markov chain. Of course, a possible strategy would be, when n is fixed, to sample n_i^* times the variable $Y^{(s_i)}(j)$ only the first time the state s_i is hit, until all the states are hit. Nevertheless, this approach is clearly not efficient, since in several time steps the estimator is forced to remain idle.

Motivated by Remark 3.4, now we devise an efficient and straightforward sampling strategy, consisting in sampling $Y^{(s_i)}(j)$, *each* time the state s_i is hit, an equal number of times over all $i=1,\ldots,|S|$. Let us first show a useful classical result for Markov chains. Let η be the number of steps performed by the Markov chain $\{S_t,\ t\in[0;\eta-1]\}$. Let η_i be the number of visits to state s_i, i.e.

$$\eta_i=\sum_{t=0}^{\eta-1}\mathbb{I}(S_t=s_i).$$

Theorem 3.16 ([69]). *Let* $\{S_t,\ t \geq 1\}$ *be an ergodic Markov chain. Let* $\hat{\boldsymbol{\pi}}_i^{(\eta)} \equiv \eta_i/\eta$*. Then, for any distribution on the initial state and for all* $i = 1, \ldots, |S|$*,*

$$\hat{\boldsymbol{\pi}}_i^{(\eta)} \xrightarrow{\eta \uparrow \infty} \boldsymbol{\pi}_i \qquad \text{with probability } 1,$$

where $\boldsymbol{\pi}$ *is the stationary distribution of the Markov chain.*

It is evident from (3.40) that $\widetilde{\epsilon}(n, \delta) \in \Theta(n^{-1/2})$. We will now show under which conditions the straightforward and efficient sampling strategy described above allows to achieve asymptotically for $n \uparrow \infty$ the best rate of convergence of $\widetilde{\epsilon}(n, \delta)$, for δ fixed. The reader can find the proof of the next Theorem in [70].

Theorem 3.17. *Suppose that Assumption 3.2 holds. Let the Markov chain of the weighted voting Markovian game be ergodic. Fix the confidence probability* $1 - \delta$*. Under the average criterion, if each time the state* s_i *is hit then the estimator agent samples the random variable* $Y^{(s_i)}(j)$ *a constant number of times not depending on* i *(e.g. 1), then with probability 1:*

$$\sqrt{n}\, \widetilde{\epsilon}(n, \delta) \xrightarrow{n \uparrow \infty} \inf_{n \in \mathbb{N}} \min_{\substack{n_1, \ldots, n_{|S|}: \\ \sum_i n_i = n}} \sqrt{n}\, \widetilde{\epsilon}(n, \delta) = \sqrt{\frac{\log(2/\delta)}{2}}.$$

Second dynamic approach

Since Hoeffding's inequality has a very general applicability and does not refer to any particular probability distribution of the random variables at issue, it is natural to look for confidence intervals especially suited to particular instances of games. In this section we will show a third confidence interval for the Shapley value of the Markovian game Γ which is tighter *i*) the higher the confidence probability $1 - \delta$ is and *ii*) the tighter the confidence intervals $[l_i; r_i]$ are. As an example, in the following we will show a tight confidence interval for weighted voting Markovian games.

We suppose that we have at our disposal beforehand a general confidence interval $[l_i; r_i]$ for the Shapley value $\mathbf{Sh}_j^{(s_i)}$ in the static games in state s_i, for all states $s_i \in S$. In general, the extrema l_i and r_i may depend on n_i, $\sum_{t=1}^{n_i} Y_t^{(s_i)}(j)$, and δ_i.

As in the case of DCI1, the randomized approach proposed in this section also holds both under the static Assumption 3.1 and under the dynamic Assumption 3.2. It is based on the following Lemma, whose proof can be found in [70].

Lemma 3.13. *Let* $A_1, \ldots, A_k$ *be* k *random variables such that* $\Pr(A_i \in [l_i; r_i]) \geq 1 - \delta_i$*. Let* $c_i \geq 0$*, for* $i = 1, \ldots, k$*. Then,*

$$\Pr\left(\sum_{i=1}^{k} c_i\, A_i \in \left[\sum_{i=1}^{k} c_i l_i\ ;\ \sum_{i=1}^{k} c_i r_i\right]\right) \geq \prod_{i=1}^{k} [1 - \delta_i]$$

The reader should keep in mind that, the smaller the single confidence levels $\delta_1, \dots, \delta_k$ are, the tighter the lower bound on the confidence probability $\prod_{i=1}^{k}(1-\delta_i)$ is.
Now we are ready to present our second dynamic approach. Let the random variable $Y^{(s_i)}(j)$ be defined as in (3.39). The following approach, as well as DCI1, implies that $Y^{(s_i)}(j)$ is sampled n_i times in state s_i, for all $s_i \in S$.

Dynamic Confidence Interval 3.3.2 (**DCI2)**. Set $\delta_i \in (0;1)$, for all $i = 1, \dots, |S|$. Let

$$\left[l^{(s_i)} \left(n_i, \sum_{t=1}^{n} Y_t^{(s_i)}(j), \delta_i \right) \; ; \; r^{(s_i)} \left(n_i, \sum_{t=1}^{n} Y_t^{(s_i)}(j), \delta_i \right) \right] \tag{3.43}$$

be the confidence interval for $\mathbf{Sh}^{(s_i)}$, with probability of confidence $1-\delta_i$, for all $i = 1, \dots, |S|$. Let $1 \le j \le P$, $s \in S$. Then, with probability of confidence $\prod_{i=1}^{|S|}(1-\delta_i)$, $\mathrm{ShM}_j(\Gamma_s)$ belongs to the confidence interval

$$\left[\sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, l^{(s_i)} \left(n_i, \sum_{t=1}^{n_i} Y_t^{(s_i)}(j), \delta_i \right) \; ; \; \sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\, r^{(s_i)} \left(n_i, \sum_{t=1}^{n_i} Y_t^{(s_i)}(j), \delta_i \right) \right].$$

We notice that the confidence interval DCI2 reveals the most natural connection between the issue of computing confidence intervals of Shapley value in static games, already addressed in [65], and in Markovian games under the dynamic Assumption 3.2.

We already saw in Section 3.3.4 that the accuracy of DCI1 can be maximized by adjusting the number of queries $n_1, \dots, n_{|S|}$ in each state. Here, in addition, we could optimize DCI2 also over the set of confidence levels $\delta_1, \dots, \delta_{|S|}$, under the nonlinear constraint:

$$\prod_{i=1}^{|S|} [1-\delta_i] = 1 - \delta \,.$$

Weighted voting Markovian games

The aim of this section is twofold. Firstly, we suggest methods to compute a confidence interval for the Shapley-Shubik index in weighted voting static games, as a complement of the study in [65]. Secondly, such methods can be utilized to compute efficiently the confidence interval DCI2 for SSM, as it is clear from the definition of DCI2 itself.
In [65], the authors derived a confidence interval for the Shapley value of a single stage game, based on Hoeffding's inequality. Nevertheless, for weighted voting static games, a tighter confidence interval can be obtained, by applying the following approach. Let $\chi \in X$ be a random permutation of $\{1, \dots, P\}$. Let us assume that $\{\chi_k \in X\}$, $k \ge 1$, are uniform and independent. Let us define the Bernoulli variable $Y^{(s)}(j)$ as in (3.39). As pointed out in [65], we can interpret the Shapley-Shubik index $\mathrm{SS}_j^{(s)}$ as

$$\mathrm{SS}_j^{(s)} = \Pr\left(Y^{(s)}(j) = 1 \right).$$

Let $Y_1^{(s)}(j), \ldots, Y_n^{(s)}(j)$ be independent realization of $Y^{(s)}(j)$. It is evident that

$$\sum_{k=1}^{n} Y_k^{(s)}(j) \sim \mathcal{B}(n, \mathrm{SS}_j^{(s)}),$$

where $\mathcal{B}(a, b)$ is the binomial distribution with parameters a, b. Hence, computing a confidence interval for $\mathrm{SS}_j^{(s)}$ boils down to the computation of confidence intervals of the probability of success of the Bernoulli variable $Y^{(s)}(j)$ given the proportion of successes $\sum_{k=1}^{n} Y_k^{(s)}(j)/n$, which is a well know problem in literature. Of course, this might be accomplished by using the general Hoeffding's inequality as in [65], but over the last decades some more efficient methods have been proposed, like the Chernoff bound [71], the Wilson's score interval [72], the Wald interval [73], the adjusted Wald interval [74], and the "exact" Clopper-Pearson interval [75].

3.3.5 Comparison among the proposed approaches

In this section we focus on *weighted voting Markovian games*, and we compare the accuracy of the proposed randomized approaches. We know that, under the static Assumption 3.1, we are allowed to use any of the three methods presented in this section, SCI, DCI1, and DCI2, to compute a confidence interval for the Shapley-Shubik index in weighted voting Markovian games. In fact, DCI1 and DCI2 involve independent queries over the different states, and this can also be done under Assumptions Assumption 3.1. Therefore, it makes sense to compare the tightness of the two confidence intervals SCI and DCI1.

Lemma 3.14. *Consider weighted voting Markovian games. Let $2\epsilon(n, \delta)$ be the accuracy of SCI (see eq. 3.38). Let $2\widetilde{\epsilon}(n, \delta)$ be the accuracy of DCI1 (see eq. 3.41). Then, for any integer n and for any confidence probability $1 - \delta$,*

$$\epsilon(n, \delta) \leq \widetilde{\epsilon}(n, \delta).$$

An interested reader can find the proof of Lemma 3.14 in [68].

Remark 3.5. The reader should not be misled by the result in Lemma 3.14. In fact, n being equal in the two cases, the number of queries needed for confidence interval SCI is $|S|$ times bigger than for DCI1, since each sampling of the variable $Z(j)$, defined in (3.36), requires $|S|$ queries, one per each state. The comparison between the two confidence interval would be fair only if the estimator agent knew beforehand the coalition values of the long run game $\{v(\Lambda, \Gamma_s)\}_{s,\Lambda}$.

According to Remark 3.5, we should compare the length of the confidence interval for the static case, $2\,\epsilon(n, \delta)$, with the one for the dynamic case, $2\,\widetilde{\epsilon}(|S|n, \delta)$, calculated with $|S|$ times many queries. Intriguingly, the relation between the tightness of SCI and DCI is now, for a suitable query strategy, reversed.

Theorem 3.18. *In the case of weighted voting Markovian games, for any integer n,*

$$\min_{\substack{n'_1,\dots,n'_{|S|}:\\ \sum_i n'_i=|S|n}} \widetilde{\epsilon}(|S|n,\delta) \leq \epsilon(n,\delta).$$

Proof. We can write

$$\min_{\substack{n'_1,\dots,n'_{|S|}:\\ \sum_i n'_i=|S|n}} \sum_{i=1}^{|S|} \frac{\boldsymbol{\sigma}_i^2(s)}{n'_i} \leq \sum_{i=1}^{|S|} \frac{\boldsymbol{\sigma}_i^2(s)}{\sum_{k=1}^{|S|} n'_k/|S|} = \sum_{i=1}^{|S|} \frac{\boldsymbol{\sigma}_i^2(s)}{n} \leq \frac{\left[\sum_{i=1}^{|S|} \boldsymbol{\sigma}_i(s)\right]^2}{n} \qquad (3.44)$$

where the last inequality holds since $\boldsymbol{\sigma}_i(s) \geq 0$. Hence, by inspection over the expressions (3.38) and (3.41), the thesis is proved. □

Theorem 3.18 clarifies the relation between the confidence intervals SCI and DCI1, under the condition of weighted voting Markovian games. We highlight its significance in the next two remarks.

Remark 3.6. Theorem 3.18 claims that the approach DCI1 is more accurate than SCI for a suitable choice of $n'_1,\dots,n'_{|S|}$, when the number of queries is equal for the two methods. In essence, this occurs because the dynamic approach allows us to tune the number of queries in the coalition values according to the weight $\boldsymbol{\sigma}_i(s)$ of each state s_i in the long run game. Moreover, the queries on coalition values are independent among the states, hence providing more diversity to the statistics.

Remark 3.7. As we already remarked, the dynamic Assumption 3.2 is more pragmatic and less restrictive than the static Assumption 3.1. Let us now give some insights on the accuracy that can be achieved by the approaches SCI and DCI1 under Assumptions Assumption 3.1 and Assumption 3.2 themselves. The approach DCI1 can be also utilized under static Assumption 3.1, and in finite time DCI1 is more accurate under Assumption 3.1 than under Assumption 3.2. Indeed, for a fixed n and under the static Assumption 3.1, the value of $n'_1,\dots,n'_{|S|}$ in (3.44) can always be set to the optimum value, since the algorithm DCI1 is run off-line. Instead, under the dynamic Assumption 3.2, the sequence of states over time $S_0,S_1,S_2,\dots$ is unknown *a priori* by the estimator agent, hence $n'_1,\dots,n'_{|S|}$ cannot be optimized for a finite n. Hence, in finite time, the static Assumption 3.1 has still an edge over the dynamic Assumption 3.2.
Nevertheless, we know from Theorem 3.17 that, for the average criterion in ergodic Markov chains, there exists a query strategy enabling to achieve an optimum rate of convergence for DCI1's accuracy. Therefore we can conclude with the following consideration. Under the average criterion, DCI1, when employed under the dynamic Assumption 3.2, can be *asymptotically* as accurate as DCI1 itself and more accurate than SCI, when they are both employed under the stronger static Assumption 3.1.

In addition to what has just been discussed, simulations showed that, when the number of queries n and the confidence level δ are equal for the two methods, then the *effective* confidence probability for SCI is generally higher than for DCI1, i.e. the lower bound $1-\delta$ is less tight. We explain this by reminding that the centers of the confidence intervals SCI and DC1, respectively

$$\frac{1}{m}\sum_{k=1}^{m} Z_k(j) \quad , \quad \sum_{i=1}^{|S|} \frac{\boldsymbol{\sigma}_i(s)}{n_i} \sum_{t=1}^{n_i} Y_t^{(s_i)}$$

are already two estimators for $\mathrm{SSM}(\Gamma_s)$, and the former possesses a smaller variance than the second one.

$1-\delta$	$a_{2\succ 2}$ (%)
.97	100
.95	99.9
.9	87.5
.8	57.7

Table 3.4: Percentage $a_{2\succ 1}$ of cases in which the confidence interval DCI2 is narrower than confidence interval DCI1, at different confidence probabilities. The Clopper-Pearson interval is considered for DCI2.

About the performance of confidence interval DCI2, the simulations confirmed our intuitions. We utilized the Clopper-Pearson interval to compute a confidence interval for the Shapley-Shubik index in weighted voting static games, and we saw that the confidence interval is more and more tight when the confidence probability approaches 1. Let $a_{2\succ 1}$ be the percentage of weighted voting Markovian game instances, generated randomly, in which the confidence interval DCI2 is narrower than confidence interval DCI1. In Table 3.4 we show, for each value of confidence probability $1-\delta$, the values of $a_{2\succ 1}$ obtained from simulations. We see that, for $1-\delta < 0.8$, the two confidence interval have a comparable length. For $1-\delta \geq 0.8$, the confidence interval DCI2 is evidently tighter than DCI1 under these settings.

3.3.6 Complexity of confidence intervals

In Section 3.3.2 we motivated the importance of devising an algorithm that approximates SSM with a polynomial accuracy in the number of players P without the need of an exponential number of queries. In this section we show that the proposed randomized approaches SCI and DCI1 fulfill this requirement. Before, we still need to clarify the notion of accuracy of a randomized approach, in parallel with the one of a deterministic approach shown in Definition 3.4.

Definition 3.6. Fix a confidence level δ and a number of queries n. The *accuracy of a randomized algorithm* approximating SSM is the length its confidence interval.

In the previous sections we derived three confidence intervals for ShM/SSM. Now we show that the number of queries required by SCI and DCI1 does not even depend on the number of players P.

Proposition 3.5. Fix the confidence level δ and the length of confidence interval $2\,\epsilon$. Then n queries are required for the confidence interval SCI, where

$$n = \frac{\left[\overline{y} - \underline{y}\right]^2 \log(2/\delta)}{2\,\epsilon^2}.$$

Proof. The proof follows straightforward from the expression of confidence interval SCI. □

Proposition 3.6. Fix the confidence level δ and the length of confidence interval $2\,\widetilde{\epsilon}$. Then, there exists values of $n_1, \ldots, n_{|S|}$, with $\sum_i n_i = n$, such that n queries are required for the confidence interval DCI1, where

$$n \leq \frac{|S|\,\left[\overline{y} - \underline{y}\right]^2\,\log(2/\delta)}{2\,\widetilde{\epsilon}^2}.$$

Proof. The proof follows straightforward from Theorem 3.18. □

From Propositions Proposition 3.5 and Proposition 3.6 we derive the following fundamental result on the complexity of SCI and DCI1.

Theorem 3.19. *Let $p(P)$ be a polynomial in the variable P. The number of queries required to achieve an accuracy of $1/p(P)$ is $O(p^2(P))$, for both the approaches SCI and DCI1.*

Since we did not provide an explicit expression for the confidence interval DCI2, then we can not provide a result analogous to Theorem 3.19 for DCI2 either. Anyway, we notice that the expression (3.43) of its confidence interval does not depend on the number of players P. Moreover, if the Hoeffding's inequality is used to compute the confidence interval for the Shapley value in the static games, then a result similar to Theorem 3.19 can be derived for DCI2.

Remark 3.8. Corollary 3.5 and Theorem 3.19 explain in what sense the proposed randomized approaches SCI and DCI1 are better than any deterministic approach, according to Definition 3.4 and Definition 3.6 of "accuracy". In order to achieve an accuracy in the order of P^{-1}, for a number of players P sufficiently high, the number of queries needed by SCI and DCI1 is always smaller than the number of queries employed by any deterministic algorithm.

3.3.7 Conclusive Considerations

We proved in Section 3.3.2 that an exponential number of queries is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy. Hence, we directed our attention to randomized algorithms and we proposed three different methods to compute a confidence interval for SSM. The first one, described in Section 3.3.3 and called SCI, assumes that the coalition values in each state are available off-line to the estimator agent. SCI can be seen as a benchmark for the performance of the other two methods, DCI1 in Sections 3.3.4 and DCI2 in Section 3.3.4. The last two methods can be utilized also if we pragmatically assume that the estimator learns the coalition values in each static game while the Markov chain process unfolds. DCI2 reveals the most natural connection between confidence intervals of Shapley value in static games, presented in [65], and in Markovian games. As a by-product of the study of DCI2, we provided confidence intervals for the Shapley-Shubik index in static games, which are tighter than the one proposed in [65]. We proposed a straightforward way to optimize the tightness of DCI1. We compared in Section 3.3.5 the proposed three approaches in terms of tightness of the confidence interval. We proved that DCI1 is tighter than SCI, with an equal number of queries and for a suitable choice of the number of queries on coalition values in each state. This occurs essentially because DCI1 allows us to tune the number of samples according to the weight of the state. Hence we showed that, *asymptotically*, the dynamic Assumption 3.2 is not restrictive with respect to the much stronger static Assumption 3.1, under the average criterion and for what concerns SCI and DCI1. The simulations confirmed that DCI2 is more accurate than the SCI and DCI1 when both the confidence probability is close to 1 and a tight confidence interval for the Shapley-Shubik index of static games is available, like the Clopper-Pearson interval. Finally, in Section 3.3.6 we showed that a polynomial number of queries is sufficient to achieve a polynomial accuracy for the proposed algorithms. Hence, in order to compute SSM, the proposed randomized approaches are more accurate than any deterministic approach for a number of players sufficiently high. The three proposed randomized approaches also produce confidence intervals for the Shapley value in *any* cooperative Markovian game.

3.4 Stochastic Games for Cooperative Network Routing and Epidemic Spread

Several providers share a network to provide connection towards a unique common destination to their customers. We provide a framework of a coalition game to facilitate the design of the available network links and their costs such that there exists an optimum routing strategy and a cost sharing satisfying all the subsets of providers. More specifically, we provide algorithms to compute the coalition values, i.e. the minimum costs that each coalition can ensure for itself. The proposed algorithm is based on some results for two-player zero-sum stochastic games with perfect information in Section 3.1.

It is worth noticing that the analyzed problem differs substantially from the noncooperative routing games thoroughly studied in literature (for additional details see e.g. [76] and references therein). At the best of the authors' knowledge, this work is the first one applying coalition games to determine an optimum routing solution and cost allocation in a shared network.

3.4.1 Routing model

We consider a network consisting of a set of nodes $V = \{1, \ldots, N\}$. M service providers share the network to offer their customers connection toward a single destination node N. The customers's traffic is injected in the network at $n \leq N-1$ nodes, called sources, located in nodes $\mathcal{T} = \{i_1, \ldots, i_n\} \subseteq V/\{N\}$. There is only one destination, in node N. We assume that all the sources transmit at the same rate the packets of a provider k, for all possible k. Let $c_k(i,j) > 0$ represent the cost per unit time that provider k has to sustain to convey its own packets, sent by any of the sources in $\mathcal{T}$, through the link $i \rightarrow j$.

The k-th service provider controls the routing, i.e. the activation of outgoing links, in the set of nodes V_k. We suppose that a node is controlled at most by one provider, i.e., $V_i \cap V_j = \emptyset, \ \forall i \neq j$ and $\bigcup_i V_i \subseteq V$. Each node i is assigned a subset $\alpha_i \subseteq V$, such that the *directed* link $i \rightarrow j$ can be activated if and only if $j \in \alpha_i$. In the generic node $i \in V_k$ controlled by provider k, provider k himself can assign a probability distribution $\mathbf{f}_k$ to the each node $j \in \alpha_i$ such that the probability that the network link (i,j) is utilized for routing is $\mathbf{f}_k(i,j)$ at *any* routing decision moment. The destination node is a "sink", and it does not route the incoming packets to any of the other nodes. We remark that all the nodes $\{1, \ldots, N-1\}$, included the sources, serve as routing nodes.

Let $\mathbf{\Phi}_\beta^{(k)}$, with $\beta \in [0;1]$, be a N-by-1 vector whose i-th component is the expected β-discounted sum of costs:

$$E_{\mathbf{f}_1,\ldots,\mathbf{f}_M}\Big[\sum_{t \geq 0} \beta^t c_k(i_t, i_{t+1})\Big], \text{ with } i_0 = i,$$

where i_t is the t-th node crossed by the packets. It is worth noticing that, for $\beta = 1$, $\mathbf{\Phi}_1^{(k)}$ is the undiscounted sum and its l-th component, with $l \in \mathcal{T}$, is the cost per unit time that provider k incurs for the stream of packets going from the l-th source to the destination.

3.4.2 Routing coalition game

Let $\mathcal{M} = \{1, \ldots, M\}$ be the grand coalition of service providers. We assume that the providers belonging to a generic coalition $\mathcal{C} \subseteq \mathcal{M}$ can stipulate binding agreements among them to enforce the optimum strategy for the coalition and distribute the costs among themselves.
Let $\mathbf{F}_\mathcal{C}$ be the set of strategies available to coalition $\mathcal{C} \subseteq \mathcal{M}$. It is easy to show that $\mathbf{F}_\mathcal{C}$ is the Cartesian product of the strategies available to all the members of $\mathcal{C}$, i.e. $\mathbf{F}_\mathcal{C} = \times_{k \in \mathcal{C}} \mathbf{F}_k$, and the set of strategies $\mathbf{F}_\mathcal{C}$ is dubbed *not correlated*. Moreover, thanks

to available results on stochastic games (see Section 3.2), we can focus only on *pure* (deterministic) strategies. Let $\mathbf{F}_{\mathcal{C}}$ be the set of pure strategies for $\mathcal{C}$, i.e.

$$\mathbf{F}_{\mathcal{C}} = \Big\{\mathbf{f}_k : \; \{k\} \in \mathcal{C}; \; \forall\, i \in V_k, \; \exists\, j : \mathbf{f}_k(i,j) = 1\Big\}.$$

Let us define, for any $\beta \in [0;1]$, the expected β-discounted sum

$$\mathbf{\Phi}_{\beta}^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \sum_{\{k\}\in\mathcal{C}} \mathbf{\Phi}_{\beta}^{(k)}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$$

Let $\mathbf{e}_{\mathcal{I}}$ be a N-by-1 vector containing 1's in correspondance of the sources and 0's otherwise. In this section we are interested in the case $\beta = 1$, since the quantity $\mathbf{e}_{\mathcal{I}}^T \mathbf{\Phi}_{\beta=1}^{(\mathcal{C})}$ is the total cost per unit time that $\mathcal{C}$ incurs to sustain its $n|\mathcal{C}|$ information streams. The minimum cost $v(\mathcal{C})$ that coalition $\mathcal{C}$ can ensure for itself is

$$v(\mathcal{C}) = \min_{\mathbf{f}_{\mathcal{C}}\in\mathbf{F}_{\mathcal{C}}} \; \max_{\mathbf{f}_{\mathcal{M}/\mathcal{C}}\in\mathbf{F}_{\mathcal{M}/\mathcal{C}}} \mathbf{e}_{\mathcal{I}}^T \mathbf{\Phi}_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}). \tag{3.45}$$

Under the TU condition, we suppose that $v(\mathcal{C})$ can be partitioned among the providers of $\mathcal{C}$ in any manner, thanks to a binding agreement among its members. We can say that $v(\mathcal{C})$ is the minmax value of a zero-sum game between the coalition $\mathcal{C}$ and the rest of the providers $\mathcal{M}/\mathcal{C}$, who are willing to "punish" the coalition $\mathcal{C}$.

The formulation of this conflict among coalitions as a two-player stochastic game with perfect information is described hereinafter.

Player 2 is the coalition $\mathcal{C}\subset\mathcal{M}$, while player 1 is the rest of the providers $\mathcal{M}/\mathcal{C}$. There exist a bijective association between the network nodes V and the states S. Let S_1 and S_2 be the set of states associated to the set of nodes $\bigcup_{\{k\}\in\mathcal{M}/\mathcal{C}} V_k$ and to $\bigcup_{\{k\}\in\mathcal{C}} V_k$, respectively. The network link $i\rightarrow j$ is activated if and only if player k selects the action $a_j^{(k)}(s_i)$, where $j \in \alpha_i$, $k : s_i \in S_k$. The instantaneous reward $r(s_i, a_j^{(k)}(s_i)) = \sum_{\{p\}\in\mathcal{C}} c_p(i,j)$, where k is the player that controls the node i. The transition probability is $p(s_w|s_i, a_j^{(k)}(s_i)) = \mathbb{1}(w = j)$, where $\mathbb{1}$ is the indicator function. Note that $\sum_{s'\in S} p(s'|s, \mathbf{f}, \mathbf{g}) = 1$, $\forall\, s \in S/\{s_N\}$ and for each couple of stationary strategies $(\mathbf{f}, \mathbf{g})$. The destination node is a "sink", i.e. $p(s_i|s_N) = 0, \; \forall\, i \in [1; N]$, and no actions are available in it for both players.

The overall optimum global routing strategy $\mathbf{F}^{\mathrm{o}}$ satifies

$$v(\mathcal{M}) = \mathbf{e}_{\mathcal{I}}^T \mathbf{\Phi}_1^{(\mathcal{M})}(\mathbf{F}^{\mathrm{o}}) = \min_{\mathbf{f}_{\mathcal{M}}\in\mathfrak{F}_{\mathcal{M}}} \mathbf{e}_{\mathcal{I}}^T \mathbf{\Phi}_1^{(\mathcal{M})}(\mathbf{f}_{\mathcal{M}})$$

where $\mathbf{F}_{\mathcal{M}}$ is the set of strategies available to the grand coalition $\mathcal{M}$. It is easy to see that the superadditivity property of the characteristic function v:

$$v(\mathcal{C}_1) + v(\mathcal{C}_2) \geq v(\mathcal{C}_1 \cup \mathcal{C}_2), \; \forall\, \mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{M}, \; \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$$

holds directly from the minmax definition (3.45) of v.

Algorithm for computing coalition values

The values $v(\mathcal{C})$ may be *infinite*. In Section 3.2 it is shown that $v(\mathcal{C})$ is the value of the game at Nash equilibrium. If $v(\mathcal{C}) = +\infty$, then the optimal strategies for the players, i.e. the strategies at Nash equilibrium, impede at least one source-destination path by causing a loop in the network. In practice, $v(\mathcal{C}) = +\infty$ is not the cost that coalition $\mathcal{C}$ has to bear; anyway, it shows well that any service provider cannot accept to lose its own packets.
The theory of stochastic games provides an approach to *avoid infinities in the computation of coalition values*. The details are illustrated in the following lemma whose proof can be found in [77].

Lemma 3.15. *Suppose that all the instantaneous rewards are nonnegative. Let us utilize the extended line of real numbers, i.e. treat $\pm\infty$ as a number ($\pm\infty = \pm\infty$, $-\infty < a \in \mathbb{R} < +\infty$). Then, the uniform optimal strategies are optimal in the undiscounted criterion as well, i.e.*

$$\mathbf{\Phi}_1(\mathbf{f}, \mathbf{g}^*) \leq \mathbf{\Phi}_1(\mathbf{f}^*, \mathbf{g}^*) \leq \mathbf{\Phi}_1(\mathbf{f}^*, \mathbf{g}) \quad \forall \, \mathbf{f}, \mathbf{g} \tag{3.46}$$

The idea is to compute the optimal strategies $(\mathbf{f}^*_{\mathcal{M}/\mathcal{C}}, \mathbf{f}^*_{\mathcal{C}})$, for coalitions $\mathcal{M}/\mathcal{C}$ and $\mathcal{C}$ respectively, for *all* the discount factors sufficiently close to 1. Then, we adopt the strategy that is still optimal in the limit for $\beta \to 1$.

In the following, we illustrate the proposed approach. Fix a pure strategy $\mathbf{f}_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$. We say that the pure strategy $\mathbf{f}'_{\mathcal{C}}$ is an improvement for coalition $\mathcal{C}$ with respect to $\mathbf{f}'_{\mathcal{C}}$ for the discount factor β iff

$$\mathbf{\Phi}^{(\mathcal{C})}_{\beta}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}'_{\mathcal{C}}) \leq \mathbf{\Phi}^{(\mathcal{C})}_{\beta}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$$

where the relation $\leq$ is component-wise and $<$ is valid for at least one component. Let $\Gamma_{\mathcal{M}/\mathcal{C}}(\bar{\mathbf{f}}_{\mathcal{C}})$ be the optimization problem that $\mathcal{M}/\mathcal{C}$ faces when $\mathcal{C}$ fixes its own strategy $\bar{\mathbf{f}}_{\mathcal{C}}$. Then, the optimum strategy for $\mathcal{M}/\mathcal{C}$ in $\Gamma_{\mathcal{M}/\mathcal{C}}(\bar{\mathbf{f}}_{\mathcal{C}})$ maximizes $\mathbf{\Phi}^{(\mathcal{C})}_{\beta}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \bar{\mathbf{f}}_{\mathcal{C}})$ component-wisely.

Algorithm 3.4.1.

1. Pick a pure routing strategy $f_{\mathcal{C}}$ for coalition $\mathcal{C}$.
2. Find the best strategy $f_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$ in the optimization problem $\Gamma_{\mathcal{M}/\mathcal{C}}(f_{\mathcal{C}})$, for all the discount factors close enough to 1.
3. Find the *first* node controlled by coalition $\mathcal{C}$ in which a change of strategy $f'_{\mathcal{C}}$ is a benefit for coalition $\mathcal{C}$ for all the discount factors close enough to 1. If it does not exists, then set $(f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}}) := (f_{\mathcal{M}/\mathcal{C}}, f_{\mathcal{C}})$ and go to step 4. Otherwise, set $f_{\mathcal{C}} := f'_{\mathcal{C}}$ and go to step 2.
4. If $\lim_{\beta\to 1} \mathbf{e}^T_{\mathcal{I}} \mathbf{\Phi}^{(\mathcal{C})}_{\beta}(f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}}) = l < +\infty$ then set $v(\mathcal{C}) = l$. Otherwise, set $v(\mathcal{C}) = +\infty$.

We remark that the optimal strategy in step 2 and the strategy refinement in step 3 are found with the help of simplex tableaux in the non-archimedean ordered field $F(\mathbb{R})$ of rational functions with real polynomial coefficients (for all details, see Section 3.1).

Transient case

Suppose now that the following assumption holds.

Assumption 3.3. For any couple of pure strategies $(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ for $\mathcal{M}/\mathcal{C}$ and $\mathcal{C}$ respectively, and for all $i \in V$, there exists a path[1] $\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ of finite length[2] $L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ and without loops linking node i to the destination node N.

The following result shows that the assumption above ensures $\mathbf{\Phi}_1^{(\mathcal{C})}$ to be finite, for any couple of strategies.

Proposition 3.7. Suppose that Assumption 3.3 holds. Then, for all the pure strategies $\mathbf{f}_{\mathcal{M}/\mathcal{C}} \in \mathbf{F}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}} \in \mathbf{F}_{\mathcal{C}}$:

(i) the path $\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ is unique;

(ii) $\mathbf{\Phi}_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) < +\infty$.

Proof. Let $\tau_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \{i_0 = i, i_1, \dots, i_{L_i} = N\}$ be the nodes crossed by the path τ_i when $\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}$ are fixed. If there existed more than one path linking two nodes then there would exist at least one node in which more than one arc go out of it. This is impossible since the strategies are pure. Then, *(i)* is proved. Therefore, we can say that

$$\begin{cases} p_t(j|i_0 = i, \mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = \mathbb{1}(j = i_t), & \forall\, t \in [1; L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})] \\ p_t(j|i_0 = i, \mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) = 0, & \forall\, t > L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}}) \end{cases}$$

where $p_t(j|i_0)$ is the probability that the t-th node crossed by the packets starting in node i_0 is j. Thus, $\forall\, i \in V$, the i-th component of $\Phi_1^{(\mathcal{C})}(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})$ is bounded by

$$L_i(\mathbf{f}_{\mathcal{M}/\mathcal{C}}, \mathbf{f}_{\mathcal{C}})\,|\,\mathcal{C}\,|\, \max_{k,i,j} c_k(i,j) < +\infty$$

□

Adapted algorithm for finite coalition values

If Assumption 3.3 holds, then the algorithm 3.4.1 can be adapted as follows.

Algorithm 3.4.2.

1. Pick a pure routing strategy $f_{\mathcal{C}}$ for coalition $\mathcal{C}$.

[1] a path is a sequence of connected nodes

[2] the length of the path is the number of edges that it is composed of.

2. Find the best strategy $f_{\mathcal{M}/\mathcal{C}}$ for coalition $\mathcal{M}/\mathcal{C}$ in the optimization problem $\Gamma_{\mathcal{M}/\mathcal{C}}(f_{\mathcal{C}})$, for $\beta = 1$.
3. Find the *first* node controlled by coalition $\mathcal{C}$ in which a change of strategy $f'_{\mathcal{C}}$ is a benefit for coalition $\mathcal{C}$, for $\beta = 1$. If it does not exists, then set $(f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}}) := (f_{\mathcal{M}/\mathcal{C}}, f_{\mathcal{C}})$ and go to step 4. Otherwise, set $f_{\mathcal{C}} := f'_{\mathcal{C}}$ and go to step 2.
4. Set $v(\mathcal{C}) = \mathbf{e}^T_{\mathcal{T}} \mathbf{\Phi}^{(\mathcal{C})}_1 (f^*_{\mathcal{M}/\mathcal{C}}, f^*_{\mathcal{C}})$.

We remark that the algorithm 3.4.2 is analogous to the one described by Raghavan and Syed in [48] when $\beta = 1$ and restricted to the transient case, with the difference that in step 2 the search is not necessarily lexicographic for coalition $\mathcal{M}/\mathcal{C}$. Indeed, at each iteration $\mathcal{M}/\mathcal{C}$ is allowed to find its own temporarily optimal strategy with *any* MDP solving method.

3.4.3 Network design

The main contribution of this section consists in describing how to compute the coalition values, and the network design is not our purpose. Nevertheless, we suggest which steps could be followed in this direction.

An eventual network designer should aim at devising both the routing decisions α_i available to each provider in each node $i \in V$ and the cost of the links $c_k(i,j)$, in order to ensure that each coalition of providers has an interest in not deviating from the global optimum policy $\mathbf{F}^{o}$. Formally, a network designer should ensure the non-emptiness of the *core* of the TU (transferable utility) coalition game (M, v), i.e. that set of cost $Co(v) = \{g_1, \ldots, g_M\} \in \mathbb{R}^M$ that providers can share among themselves through binding agreements, such that

$$\begin{cases} \sum_{k=1}^{M} g_k = v(\mathcal{M}) \\ \sum_{\{k\} \in \mathcal{C}} g_k \leq v(\mathcal{C}), \quad \forall\, \mathcal{C} \subset \mathcal{M}. \end{cases}$$

We see from the former equation that the core is globally *efficient* for the network and from the latter that it is also *stable* with respect to the formation of greedy coalitions.

3.4.4 Hacker-Provider routing game

The routing game with just two players described in section 3.4.1 can also be re-interpreted in the framework of the conflicts between one service provider and one hacker.
There is a set $V_1 \subseteq V$ of vulnerable nodes, where the routing control may be got hold by a hacker. V_0 is the set of nodes in which the routing is handled by a service provider. The set $V_2 = V_0/V_1$ is the set of unattackable nodes among the ones controlled by the service provider. Each link $i \to j$ is assigned $c(i,j) > 0$, that in this case can be also interpreted as a *delay*, i.e. the time that a packet of provider k spends to go from node i to node j. In such a case, let us assume that the nodes are capable to re-direct all the incoming

packets as soon as they receive them, without any additional delay due to the buffering. The service provider here wants to find the routing rule that *jointly* minimizes the packet delay $\mathbf{\Phi}_1$ for all the sources; conversely, the hacker wants to slow down the network.

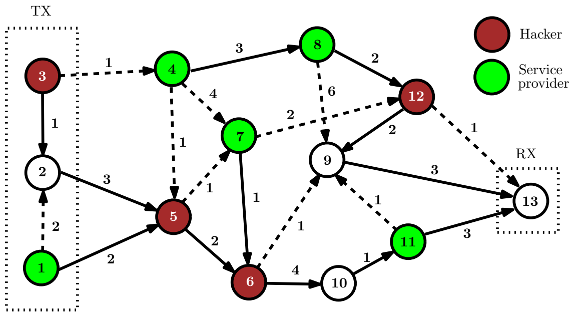

Figure 3.1: Nash equilibrium in routing game. The continuous arrows are the activated links. The costs are specified next to each arrow. Green and red nodes are controlled by the service provider and by the hacker, respectively. In white nodes there are no routing choice.

As in section 3.4.1, there may be some couple of strategies for the two players for which there exist loops in the network, that cause the packet delay from some sources to be infinite. Note that the hacker can also disrupt some nodes, by forcing a loop on them. Hence, here we deal with the general case of undiscounted stochastic games described in Section 3.2. The undiscounted optimal strategies can be computed by the algorithm 3.4.1, in which player 1 is now the hacker which controls nodes V_1, and player 2 is the service provider, which controls nodes V_2.

Note that in this case, in contrast to the coalition game, we are more interested in the computation of the optimal strategies, and not in the value of the game at the Nash equilibrium. Indeed, the optimal strategy for the service provider is the pure routing policy it should adopt in order to minimize the source-wise packet delay in the worst case. For Lemma 3.8, the worst situation for the provider is when the hacker is able to control all the vulnerable nodes V_1 and has at its disposal as many routing policies as possible. Note that the optimal strategies for both players are pure, i.e. the routing policy is deterministic in each node.

An example of optimal strategies for both players in a delay routing game is shown in Figure 3.1.

3.4.5 Natural disaster

Let us reformulate the model described in section 3.4.4, where player 1 is now a natural agent that can put out of order some nodes $V_1 \subset V$ of the network, independently of the routing action taken by the service provider in such nodes. This addresses the practical situation in which nodes V_1 are located in areas subject to catastrophic natural phenomena. It is straightforward to see that the computation of the optimal strategies for the service provider boils down to the calculation of a MDP uniform optimal solution (see [50]), in which the set of nodes of interest is reduced to V_2, that is the collection of nodes controlled by the service provider.

3.4.6 Epidemic network

In this section we model an epidemic network with N nodes; $N-1$ possibly infected individuals are located in nodes $\{1, \dots, N-1\}$ respectively. Each individual can infect, with some probability, only one among a subset of other individuals in its neighborhood. There is a probability $\boldsymbol{\mu}_i$ that the infection process starts from the i-th individual. The infection spread terminates when the virus reaches the healer, located in node N. Hence, there is a probability $\boldsymbol{\mu}_N$ that the epidemic spread is averted. There are two player: player 2, the "good" one, wants to design and force the connections among the individuals such that the lowest expected number of individuals are infected, while player 1 has the opposite goal. The assumption of perfect information still holds, i.e. the set of nodes in which player 1 and player 2 have more than one action available are disjoint.

The formulation of the problem is analogous to the two-player game described in section 3.4.1, in which the cost of the link (i, j) is 1 for all nodes i, j. The nodes are substituted by the individuals, the destination with the healer, the sources become the first infected entity, the packet routing is replaced by the virus transmission. In this context, we wish $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1$ to represent the average number of infected individuals. Therefore, for each couple of routing strategies, *no loops* in the network are allowed, i.e. we suppose that the Assumption 3.3 holds. Hence, thanks to Proposition 3.7, for every couple of pure stationary strategies $(\mathbf{f}, \mathbf{g})$, $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}, \mathbf{g})$ is actually the expected number of infected individuals.

It can be shown (see [77]) that Algorithm 3.4.2 can be used to find the optimal strategy for the "good" player, who is interested in minimizing the objective function $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}, \mathbf{g})$. If $(\mathbf{f}^*, \mathbf{g}^*)$ are the undiscounted optimal strategies, then the value $\boldsymbol{\mu}^T \boldsymbol{\Phi}_1(\mathbf{f}^*, \mathbf{g}^*)$ is the most pessimistic estimate for player 2 for the expected number of infected individuals.

3.4.7 Conclusive Considerations

Several providers share the same network and control the routing in disjoint sets of nodes. There are several information sources and one destination. By using the framework of stochastic games, we provided algorithms to compute the minimum costs that each coalition of providers can ensure for itself. This helps the optimum design of a network, which should guarantee the existence of an efficient and stable costs partition among the providers. We also modeled situations in which there are two players with conflicting interests, like a hacker against a service provider, or in which a service provider wants to reduce the damages to the network caused by a natural disaster. An epidemic spread network model was shown as well. From a theoretical perspective, we extended some results on uniform optimal strategies in stochastic game to the case of undiscounted criterion.

3.5 Cooperative Games on Markovian MACs with Jamming Users

In the last few years, the computing capability of terminals has rapidly increased, as much as their ability to sense the behavior of the other terminals and to decide what is the best strategy for themselves, in terms of power consumption, handover, choice of symbol constellation and so on. Thus, we are witnessing a paradigm shift, from fully centralized networks with dumb terminals to distributed networks, in which users can cooperate and pursue their own interest. Hence, from a system designer point of view, it is more and more crucial to devise a rate allocation being both optimum for the whole network and stable, i.e. no subset of users is dissatisfied with the service and decide to withdraw from the communication [78]. This challenging issue is addressed by cooperative game theory with non-transferable utility (NTU) [79], which provides powerful tools to derive efficient and stable allocations in a setting in which the players - in this context the users - can cooperate to reach a common goal.

In [80], La and Anantharam dealt with a cooperative game on multiple access channels (MAC) in which several users attempt to send information to a single receiver. Users being dissatisfied with the assigned rate can threaten to jam the network. In this case, only users with enough available power can transmit with positive rate. They considered a static channel and defined the Core of the game as the set of rates such that no subsets of players can attain a better allocation when the remaining users jam.

In this section we consider a setting similar to the one in [80], except for the fact that the channel is quasi-static, i.e. they vary slowly enough to be assumed constant for the whole duration of a codeword. Thus, the channel coefficients vary at each codeword, and follow a homogeneous Markov chain (HMC) on a discrete set of channel states. Hence, at each step of the HMC, the same game as in [80] is played. We first study the relation between the static and the Markovian game. The Core of the Markovian game is still the most attractive set of rate allocations, both from a centralized point of view and for the single users. We consider both the discounted and the average criterion to sum the rate over time. We find in Section 3.5.4 that the Core of the Markovian game is nonempty and we study its connections with the Core of the single stage game. We also consider the possibility that coalitions can change over time, along the Markov process, and we find that our allocation is still stable, for any subgame. In Section 3.5.6 we show that, under the discounted criterion, the procedure of joint rate allocations in the Markovian game for each starting state of the HMC is a delicate procedure. Indeed, the associated single stage allocations may not be feasible. Thus we propose a way to ensure their feasibility. Section 3.5.7 analyzes the α-fair allocation procedures, with particular attention to the max-min fair ($\alpha \to \infty$) and to the proportional fair ($\alpha \to 1$) [81]. We find a condition on the single stage games ensuring that, if the single stage allocations all satisfy such criteria, they also do it in the long run game. Moreover, we define in our setting the Nash bargaining solution. Although the Core is a reasonably stable solution, there is still the chance that an agreement among the users is not found, and everybody threatens to jam. For example, users may be envious of the rate assigned to some other user. Thus, we investigate this situation in Section 3.5.8 and we prove that, if the number of players P increases, the probability that some user can still communicate tends to 0 with exponential rate. As a by product of this analysis, we find that the Nash bargaining solution that we defined tends to all the three fair criteria cited above when P tends to infinity. Finally, in Section 3.5.9 we prove that in the Markovian channel the expected sum rate in the long run game tends to 0 when P tends to infinity.

A notation remark. Any order relation between vectors is to meant to be component-wise.

3.5.1 System Model

We consider a wireless system in which P terminals attempt to send information to a single receiver or base station. Let $\mathcal{P} = \{1, \ldots, P\}$ be the set of all users. Each user i has at its disposal a coding scheme C_i of length n. Let $\{M_i(k)\}_{k\in\mathbb{N}}$ be the set of messages that user i can transmit. Then, $C_i \equiv \{C_i(k) :\ M_i(k) \to \mathbb{C}^n\}_{k\in\mathbb{N}}$, i.e., C_i maps the set of messages of user i into complex symbols of length n, which does not depend on the user. The knowledge of such codes is available to all the users and, of course, to the receiver. We assume that the power of C_i is subject to the constraint Δ_i, i.e., for any codeword $\mathbf{x}^{(i)} \in C_i$,

$$\frac{1}{n}\sum_{k=1}^{n} |\mathbf{x}_k^{(i)}|^2 \leq \Delta_i.$$

We assume a quasi-static channel, i.e. the channel coefficient can be considered constant for the whole duration of a codeword. Thus, the t-th signal block received by the unique receiver, for $t \in \mathbb{N}_0$, can be written as

$$\mathbf{y}[t] = \sum_{i=1}^{P} h^{(i)}[t]\,\mathbf{x}^{(i)}[t] + \mathbf{w}[t]$$

where $h^{(i)}[t]$ is the complex channel coefficient for user i at time step t, and $\mathbf{w}[t]$ is a collection of n zero mean white Gaussian noise samples with variance N_0. We assume that the set of channel coefficients $\{h^{(1)}, h^{(2)}, \ldots, h^{(P)}\}$ is finite and it follows a discrete time HMC, which can change state at every new codeword. In other words, if S_t is the channel state at time step t, where

$$S_t \equiv \left[h^{(1)}[t], \ldots, h^{(P)}[t]\right],$$

then the random process $\{S_t,\ t \geq 0\}$ is a HMC. We define S as the set of all the N possible states of the HMC.

We stress that our system model is different from the one usually investigated from an information theoretical point of view (e.g. see [82]), since the channel does not vary during a whole codeword.
We assume that any subset $J \subset \mathcal{P}$ of users can decide not to participate anymore to the transmission and hence to threaten to jam the network. We point out that it is not granted that they actually jam, but anyway the active users should brace themselves for the worst possible scenario, and assess which rate they can ensure in such a case.
Note that our system model is equivalent to the one in [80], except for the fact that the complex channel coefficients h_i are not static, but vary at every codeword, according to a Markov chain. Our goal in this section is indeed to study a fair rate allocation in the long run process, under the assumption of quasi-static Markov channel.

3.5.2 Theoretical Background

In this section we provide some useful definitions and background results on polymatroids and cooperative game theory with NTU.

3.5.3 Polymatroids

Definition 3.7. Let A be a ground set with cardinality n and let $g : 2^A \to \mathbb{R}$ be a rank function, i.e. a real valued set function such that

$$g(B_1 \cup \{i\}) - g(B_1) \geq g(B_2 \cup \{i\}) - g(B_2), \; \forall \, B_1 \subset B_2 \subseteq A \backslash \{i\}.$$

The *polymatroid* $\mathcal{R}$ associated to g is defined as

$$\mathcal{R} \equiv \left\{ \mathbf{x} \in \mathbb{R}_+^n : \sum_{i \in T} \mathbf{x}_i \leq g(T), \quad \forall \, T \subseteq A \right\}.$$

Definition 3.8. Let A, B be two sets. Then, the set C:

$$C = A + B = \{a + b \mid a \in A, \; b \in B\}$$

is called the *Minkowski sum* between A and B.

We present here an important result from [83], p.241, Theorem 12.1.5.

Theorem 3.20. *Let $\mathcal{R}_1, \ldots, \mathcal{R}_k$ be k polymatroids on the same ground set A, and let g_i be the rank function associated to $\mathcal{R}_i$. Then, the Minkowski sum $\sum_{i=1}^k \mathcal{R}_i$ is still a polymatroid with rank function $\sum_{i=1}^k g_i$.*

The proof of the following Corollary is straightforward.

Corollary 3.6. *Let $a_i \geq 0$, for all $i = 1, \dots, k$. Let $\mathcal{R}_1, \dots, \mathcal{R}_k$ be k polymatroids on the same ground set and let g_i be the rank function of $\mathcal{R}_i$. Then, the Minkowski sum $\sum_{i=1}^{k} a_i \mathcal{R}_i$ is a polymatroid with rank function $\sum_{i=1}^{k} a_i g_i$.*

In the following we will see that the main facet of a polymatroid plays a fundamental role in our game theoretic model.

Definition 3.9. Let $\mathcal{R}$ be a polymatroid and let g be its rank function on the set A. The main facet $M(\mathcal{R})$ is defined as

$$M(\mathcal{R}) \equiv \left\{ \mathbf{x} \in \mathcal{R} : \sum_{i \in A} \mathbf{x}_i = g(A) \right\}.$$

The facet $M(\mathcal{R})$ has at most $n!$ extreme points, and each of them has an explicit characterization as a function of the rank function g. Let $\mathcal{H}(n)$ be the set of permutations of $\{1, \dots, n\}$. A point $\mathbf{w} \in M(\mathcal{R})$ is a vertex if and only if there exists a permutation σ of $\{1, \dots, n\}$ such that

$$\mathbf{w}_i \equiv \mathbf{w}_i(\sigma) = g(\{\sigma_1, \dots, \sigma_{i-1}, \sigma_i\}) - g(\{\sigma_1, \dots, \sigma_{i-1}\})$$

Let us show a classical result on polymatroids (e.g. see [84]).

Theorem 3.21. *For any $\mathbf{c} \in \mathbb{R}^{|A|}$, a solution of the linear program $\max \sum_{i=1}^{|A|} \mathbf{c}_i \mathbf{x}_i$, where $\mathbf{x} \in \mathcal{R}$, is attained at a vertex $\mathbf{w}(\sigma)$ of $M(\mathcal{R})$ such that $\mathbf{c}_{\sigma_1} \geq \dots \geq \mathbf{c}_{\sigma_{|A|}}$.*

Corollary 3.7. *For each $\mathbf{x} \in \mathcal{R}$, there exists $\overline{\mathbf{x}} \in M(\mathcal{R})$ such that $\overline{\mathbf{x}} \geq \mathbf{x}$.*

The following result directly follows from Theorem 3.20 and Corollary 3.6.

Corollary 3.8. *Let $a_i \geq 0$, for all $i = 1, \dots, k$. Let $\mathcal{R}_1, \dots, \mathcal{R}_k$ be k polymatroids on the same ground set A. Let $\mathcal{R} = \sum_{i=1}^{k} a_i \mathcal{R}_i$. Let $\mathbf{w}(\sigma)(i)$ be the vertex of $M(\mathcal{R}_i)$ associated to the permutation σ of $1, \dots, |A|$. Let $\mathbf{w}(\sigma)$ be a vertex of $M(\mathcal{R})$. Then,*

$$\mathbf{w}(\sigma) = \sum_{i=1}^{k} a_i \, \mathbf{w}(\sigma)(i), \quad \forall \, \sigma \in \mathcal{H}(|A|).$$

Cooperative game theory

In cooperative game theory, we generally deal with a set $\mathcal{P}$ of P players in which each coalition $T \subseteq \mathcal{P}$ can stipulate an off-line agreement and cooperate. Under the NTU assumption, each coalition $T \subseteq \mathcal{P}$ can ensure for itself, regardless of the action of the other players, a set of feasible payoff allocations $\mathcal{R}(T) \subseteq \mathbb{R}^{|T|}$, such that, if, $\mathbf{r} \in \mathcal{R}$, $\mathbf{r}_i$ is the payoff allocated to the i-th player belonging to T. If all the players in $\mathcal{P}$ cooperate, we say that the grand coalition $\mathcal{P}$ is formed.

Traditionally, the main objective of cooperative game theory is to find a payoff allocation which is both efficient for the whole community and stable, i.e. no coalition has interest in withdrawing from the grand coalition. A coalition T is enticed to withdraw from the grand coalition when it can ensure for itself a payoff which is strictly bigger, for each of its participants, than the one assigned by the grand coalition. This concept is known as the Core of the NTU game.

Definition 3.10. Let $\mathcal{R}(T)$ be the feasible set of coalition $T \subseteq \mathcal{P}$ in a NTU game. The *Core* of the game $\mathbf{Co}$ is the set of real P-tuples which are feasible and optimum for $\mathcal{P}$ and which are not dominated by any vector in $\mathcal{R}(T)$, for all $T \subseteq \mathcal{P}$, i.e.

$$\mathbf{Co} \equiv \Big\{ \overline{\mathbf{x}} \in \mathcal{R}(\mathcal{P}) : \nexists \, (T \subset \mathcal{P}, \widetilde{\mathbf{x}} \in \mathcal{R}(T)) \text{ s.t. } \widetilde{\mathbf{x}} > \overline{\mathbf{x}}^{(T)}, \\ \text{and } \sum_{i \in \mathcal{P}} \overline{\mathbf{x}}_i \geq \sum_{i \in \mathcal{P}} \mathbf{x}_i, \ \forall \mathbf{x} \in \mathcal{R}(\mathcal{P}) \Big\}$$

where $\mathbf{x}^{(T)}$ stands for the set of components of $\mathbf{x}$ related to players of coalition T, and the order sign $>$ is component-wise.

Nevertheless, the Core might not be synonymous with stability. For example, the envy [85] felt by a player towards the allocation assigned to another player may push the first player not to cooperate anymore. Nash addressed in [86] the problem of payoff allocations among non-cooperative players, who cannot find any agreement. He found a unique allocation, called the Nash bargaining solution, fulfilling four sensible axioms (see [86] for any further detail).

Definition 3.11. Let $\mathbf{d}$ be the disagreement point, i.e. $\mathbf{d}_i$ is the payoff that player i can ensure for itself when no agreement is attained among the players. Assume that the set $B \equiv \{\mathbf{x} \in \mathcal{R}(\mathcal{P}) : \ \mathbf{x}_i > \mathbf{d}_i, \ \forall i \in \mathcal{P}\}$ is nonempty. The *Nash bargaining solution* $\mathbf{x}^{NB} \in \mathbb{R}^P$ is defined as

$$\mathbf{x}^{NB} = \operatorname*{argmax}_{\mathbf{x} \in B} \prod_{i=1}^{P} (\mathbf{x}_i - \mathbf{d}_i)$$

3.5.4 Cooperative games

In this section we analyze the feasibility rate region of the system, assuming that any subset of users can jam the remaining active users. In our game theoretic model, the users are the players and the feasibility region for each set of active users A_J is their non transferable utility region. In Section 3.5.5 we apply some results obtained in [80] to our settings, while in Section 3.5.5 we provide new results on the feasible rate regions on Markovian MAC.

3.5.5 Static Channel

Set the channel state $s \equiv [h^{(1)}(s), \ldots, h^{(P)}(s)]$. Fix the set of jamming users $J \subset \mathcal{P}$. Let $\Lambda(J, s)$ be the maximum power achievable by the signal transmitted by the jamming users. Under the assumption that the users in J can stipulate an off-line deal and hence cooperate effectively against the set of active users $A_J \equiv \mathcal{P} \backslash J$, their signals can sum coherently and the expression of $\Lambda(J, s)$ becomes

$$\Lambda(J, s) = \left(\sum_{i \in J} |h^{(i)}(s)| \sqrt{\Delta_i} \right)^2 .$$

Let $\mathcal{R}(A_J, s)$ be the capacity region for the active users A_J in the channel state s. As usual, it is defined as the set of rates jointly achievable by all the users in A_J with a probability of detection error that tends to 0 when the block length n tends to infinity.

Assumption 3.4. In our NTU cooperative game model, we assume that $\mathcal{R}(A_J, s)$ is the set of feasible allocations for A_J in state s.

Let $\widehat{A}_J(s)$ be the set of active users such that:

$$\widehat{A}_J(s) \equiv \left\{ i \in A_J : \; |h^{(i)}(s)|^2 \Delta_i > \Lambda(J, s) \right\}$$

In [80], the authors showed that $\mathcal{R}(A_J, s)$ is the set of non-negative real $|A_J|$-tuples $\mathbf{r}_i(s)$ such that

$$\mathbf{r}_i(s) = \; 0, \quad \forall \, i \in A_J \backslash \widehat{A}_J$$
$$\sum_{i \in T} \mathbf{r}_i(s) \leq C\Bigg(\sum_{i \in T \cap \widehat{A}_J(s)} |h^{(i)}(s)|^2 \Delta_i, \Lambda(J, s) + N_0 \Bigg), \; \forall \, T \subseteq A_J$$

where $C(a, b) = \log_2(1 + a/b)$. In other words, only the users in $\widehat{A}_J(s)$, whose signal is strong enough to overwhelm the jamming signal, can communicate. It is easy to check that the capacity region $\mathcal{R}(A_J, s)$ can be rewritten as the set of $|A_J|$-tuples $\{\mathbf{r}_i \in \mathbb{R}_0^+\}_{i \in A_J}$ such that

$$\sum_{i \in T} \mathbf{r}_i(s) \leq \; C\Bigg(\sum_{i \in T} |h^{(i)}(s)|^2 \widetilde{\Delta}_i, \Lambda(J, s) + N_0 \Bigg), \; \forall \, T \subseteq A_J$$

where $\widetilde{\Delta}_i \equiv \Delta_i$ for $i \in \widetilde{A}_J(s)$ and $\widetilde{\Delta}_i = 0$ for all $i \in A_J \backslash \widetilde{A}_J(s)$. For simplicity of notations, let us define, for all $T \subseteq A_J$ and $s \in S$,

$$g_{(A_J)}(T, s) \equiv \; C\Bigg(\sum_{i \in T} |h^{(i)}(s)|^2 \widetilde{\Delta}_i, \Lambda(J, s) + N_0 \Bigg).$$

It is straightforward to check that, for any state $s \in S$ and any set A_J, $g_{(A_J)}(T, s)$ is a rank function. Therefore, in each channel state $s \in S$ and for any set of jamming users J, the feasible region $\mathcal{R}(A_J, s)$ is a polymatroid.

As remarked before, we consider $\mathcal{R}(A_J, s)$ as the set of feasible allocations for coalition A_J in the single stage game played in state $s \in S$. The Core of the game can be interpreted, in this case, as the set of efficient allocations for the grand coalition of users such that there is no subcoalition that can ensure a better allocation for all its participants when the rest of the users jam. Therefore, if a rate allocation does not lie in the Core is not efficient from a centralized point of view, and/or it is not fair for some users. In [80], La and Anatharam found the Core of the game by relying on cooperative game theory with transferable utilities. Their approach is not completely rigorous, since the rate cannot be shared in any manner among the users but only within the capacity region. Nevertheless, NTU cooperative game theory yields the same result as [80], as it is clear from the following Theorem.

Theorem 3.22. *Let Ψ_s be the single stage game played in state $s \in S$. The Core $\mathbf{Co}(\Psi_s)$ coincides with the main facet $M(\mathcal{R}(\mathcal{P}, s))$.*

Proof. For Theorem 3.21, all the points in $M(\mathcal{R}(\mathcal{P}, s))$ solve the linear program $\max_{\mathbf{r}\in\mathcal{R}(\mathcal{P},s)} \sum_{i\in\mathcal{P}} \mathbf{r}_i$. Hence, all the points in $M(\mathcal{R}(\mathcal{P}, s))$ are efficient for $\mathcal{P}$. Moreover, in [80] it is shown that, for all $\mathbf{r} \in M(\mathcal{R}(\mathcal{P}, s))$,

$$\sum_{i\in A_J} \mathbf{r}_i \geq g_{(A_J)}(A_J, s), \quad \forall A_J \subset \mathcal{P}.$$

Hence, we can say that, for all $\mathbf{r} \in M(\mathcal{R}(\mathcal{P}, s))$, there exists no allocation belonging to $M(\mathcal{R}(A_J, s))$ that dominates $\mathbf{r}$ for coalition A_J. Since any rate allocations belonging to $\mathcal{R}(A_J, s)$ is dominated by a rate allocation in $M(\mathcal{R}(A_J, s))$, then $M(\mathcal{R}(\mathcal{P}, s)) \subseteq \mathbf{Co}(\Psi_s)$. If $\mathbf{r} \notin M(\mathcal{R}(\mathcal{P}, s))$, either it is not feasible or it is not efficient for $\mathcal{P}$. Then, $M(\mathcal{R}(\mathcal{P}, s)) = \mathbf{Co}(\Psi_s)$. □

Markovian channel

In this section we mainly deal with Markovian cooperative game theory. We consider a finite set of channel states $S = \{s_1, \ldots, s_N\}$ and in each $s \in S$ the game described in Section 3.5.4 is played among the same set of players $\mathcal{P}$. The succession of states at discrete time steps $\{S_t,\ t \geq 0\}$ is a HMC, characterized by the transition probability matrix $\mathbf{P}$.

Assumption 3.5. The transition probability matrix $\mathbf{P}$ is irreducible, i,e, from any state, all the states S are reachable with positive probability.

Under Assumption 3.5, the stationary distribution π of $\mathbf{P}$ exists and the matrix $(\mathbf{I} - \beta\mathbf{P})^{-1}$ is positive. Along the Markov process, for each state $s \in S$, a rate $\mathbf{r}(s) \in \mathcal{R}(A_J, s)$ needs to be assigned to the active users A_J. We will now make our fundamental assumption on the rate allocation procedure.

Assumption 3.6. The rate allocation procedure along the HMC is stationary, i.e. it is a function exclusively of the state of the channel at each step of the HMC, and not of the rate allocations and of the channel states in the previous steps.

Let $\mathbf{r}_i(s) \in \mathcal{R}(T,s)$ be the payoff in state $s \in S$ assigned to player $i \in T$. Let Γ_s be the Markovian game when the initial state of the underlying HMC is $s \in S$. In literature, two criteria to sum the payoffs over time are usually considered. The former is the β-discounted criterion, where $\beta \in [0;1)$. In this case, the long run discounted allocation $\mathbf{r}_i^{(\beta)}(\Gamma_s)$ in the game Γ_s for player $i \in T$, provided that the initial state of the HMC is $s \in S$, is expressed as

$$\begin{aligned} \mathbf{r}_i^{(\beta)}(\Gamma_{s_j}) &= \sum_{t=0}^{\infty} \beta^t \sum_{k=1}^{N} p_t(s_k|s_j)\, \mathbf{r}_i(s_k) \\ &= \sum_{k=1}^{N} \nu^{(\beta)}(s_k, s_j)\, \mathbf{r}_i(s_k), \quad \forall\, s_j \in S \end{aligned} \tag{3.47}$$

where $\nu^{(\beta)}(., s_j)$ is the j-th row of the non negative matrix $(\mathbf{I} - \beta\mathbf{P})^{-1}$, and $p_t(s'|s)$ is the probability of being in state s' after t steps, when the initial state is s. Expression (3.47) can be inverted and described by a bijective function $\Phi^{(\beta)} :\ \mathbb{R}^{NP} \to \mathbb{R}^{NP}$ such that

$$\begin{aligned} \big[\mathbf{r}(\Gamma_{s_1})\ \ldots\ \mathbf{r}(\Gamma_{s_N})\big]^{\mathrm{T}} &= \Phi^{(\beta)}\big(\{\mathbf{r}(s)\}_{s\in S}\big) \\ &= (\mathbf{I} - \beta\mathbf{P})^{-1}\big[\mathbf{r}(s_1)\ \ldots\ \mathbf{r}(s_N)\big]^{T} \end{aligned}$$

The latter criterion considered is referred to as average criterion, and in this case $\mathbf{r}_i(\Gamma_s)$ is defined as the Cesaro limit:

$$\mathbf{r}_i(\Gamma_s) = \liminf_{n\to\infty} \frac{1}{n+1} \sum_{t=0}^{n} \sum_{s'\in S} p_t(s'|s)\mathbf{r}_i(s').$$

Since $\mathbf{P}$ is irreducible, then $\mathbf{r}_i(\Gamma_s)$ can be expressed as [87]:

$$\mathbf{r}_i(\Gamma_s) = \sum_{k=1}^{N} \pi(s_k)\, \mathbf{r}_i(s_k), \quad \forall\, s \in S. \tag{3.48}$$

where π is the stationary distribution of the HMC. We can rewrite (3.48) as a function $\Phi :\ \mathbb{R}^{NP} \to \mathbb{R}^{P}$ such that:

$$\begin{aligned} \mathbf{r}(\Gamma_s) &= \Phi\big(\{\mathbf{r}(s)\}_{s\in S}\big) \\ &= \sum_{k=1}^{N} \pi(s_k)\, \mathbf{r}(s_k) \end{aligned}$$

Let us define now the feasible rate region in the long run process under discounted and average criterion. Fix the set of jamming users J, who are supposed to jam for the whole duration of the transmission. Let $A_J = \mathcal{P} \backslash J$ be the set of active users.

Definition 3.12. Set $\beta \in [0;1)$. The *discounted feasible rate region* $\mathcal{R}^{(\beta)}(A_J, \Gamma_s)$ is defined as the weighted Minkowski sum:

$$\mathcal{R}^{(\beta)}(A_J, \Gamma_s) = \sum_{k=1}^{N} \nu^{(\beta)}(s_k, s)\mathcal{R}(A_J, s), \quad \forall\, A_J \subseteq \mathcal{P}.$$

Definition 3.13. The *average feasible rate region* $\mathcal{R}(A_J, \Gamma_s)$ is defined as the weighted Minkowski sum:

$$\mathcal{R}(A_J, \Gamma_s) = \sum_{k=1}^{N} \pi(s_k)\mathcal{R}(A_J, s_k), \quad \forall\, A_J \subseteq \mathcal{P}.$$

Note that we can define $\mathcal{R}(A_J, \Gamma) \equiv \mathcal{R}(A_J, \Gamma_s)$, since the average feasible rate region does not depend on the initial state s.
Now we are ready to derive the exact expressions for both $\mathcal{R}^{(\beta)}(A_J, \Gamma_s)$ and $\mathcal{R}(A_J, \Gamma_s)$. The following two Lemmas follows directly from Corollary 3.6.

Lemma 3.16. *Set $\beta \in [0;1)$ and $s \in S$. Let $A_J \subseteq \mathcal{P}$ be the set of active users. The discounted feasible rate region $\mathcal{R}^{(\beta)}(A_J, \Gamma_s)$ is a polymatroid with rank function:*

$$g^{(\beta)}_{(A_J)}(T, \Gamma_s) = \sum_{i=1}^{N} \nu^{(\beta)}(s_i, s)\, g_{(A_J)}(T, s_i), \quad \forall\, T \subseteq A_J.$$

Lemma 3.17. *Let $A_J \subseteq \mathcal{P}$ be the set of active users. For any initial state $s \in S$, the average feasible rate region $\mathcal{R}(A_J, \Gamma)$ is a polymatroid with rank function:*

$$g_{(A_J)}(T, \Gamma) = \sum_{i=1}^{N} \pi(s_i)\, g_{(A_J)}(T, s_i), \quad \forall\, T \subseteq A_J.$$

Thanks to Lemma 3.16 and Lemma 3.17, we can extend the results on the Core of a single stage game in Theorem 3.22 to our Markovian setting.

Lemma 3.18. *Under the β-discounted criterion, the Core of the game Γ_s, $\mathbf{Co}^{(\beta)}(\Gamma_s)$ coincides with the main facet of $\mathcal{R}^{(\beta)}(\mathcal{P}, \Gamma_s)$, i.e.*

$$\mathbf{Co}^{(\beta)}(\Gamma_s) = M\Big(\mathcal{R}^{(\beta)}(\mathcal{P}, \Gamma_s)\Big), \quad \forall\, s \in S.$$

Analogously, under the average criterion, the Core $\mathbf{Co}(\Gamma_s)$ coincides with

$$\mathbf{Co}(\Gamma_s) = M\Big(\mathcal{R}(\mathcal{P}, \Gamma)\Big), \quad \forall\, s \in S.$$

Hence, all the rate allocations in $M(\mathcal{R}(\mathcal{P}, \Gamma))$ are auspicable both from a system designer point of view, since they are efficient in the long run process, and for each subset of users A. In fact, A would not be able to ensure a better rate if the remaining users $\mathcal{P}\backslash A$ would be jamming for the whole duration of the transmission.

Time consistency

Up to now, we have assumed that the coalitions of active and jamming users are predefined at the beginning of the game and they do not change in the course of the game. Of course, a more realistic setting allows the coalition to change throughout the transmission, at each step of the channel HMC. Under this assumption, we want to investigate when an allocation is stable throughout the whole game.

Let us consider the discounted criterion. If a long run allocation $\mathbf{r} \in \mathbf{Co}^{(\beta)}(\Gamma_s)$, then if any coalition of users faces the dilemma:

do I withdraw now or I'll cooperate forever?

before the game Γ_s starts, then it always decides to cooperate, by the definition of the Core. We still ignore what is the answer to this dilemma at any intermediate step of the game. Hence, we want strengthen our concept of stability, and we wish that, for any trajectory of the HMC up to the n-th step, the expected β-discounted sum of rate allocations from state S_n on still belongs to the Core of Γ_{S_n}, for any n. In dynamic cooperative game theory, this concept is usually referred to as *time consistency* (e.g. see [88]). In our case it is naturally satisfied, as we can see from the following Theorem.

Theorem 3.23. *Let S_t be the state visited by the channel HMC at time step t. Let the set of long run allocations $\{\mathbf{r}(\Gamma_s) \in \mathbf{Co}^{(\beta)}(\Gamma_s)\}_{s\in S}$ and let $\{\mathbf{r}(s)\}_{s\in S}$ be its associated set of single stage rate allocation. Suppose that $\mathbf{r}(s) \in \mathcal{R}(\mathcal{P}, s)$, for all $s \in S$. Then, for every integer n,*

$$\mathrm{E}\left(\sum_{t=n}^{\infty} \beta^t \mathbf{r}(S_t) \Big| \mathbf{h}(n)\right) \in \beta^n \mathbf{Co}^{(\beta)}(\Gamma_{S_n})$$

where $\mathbf{h}(n)$ is the history of channel states and rate allocations up to time step n.

Proof. Since the rate allocation is assumed to be stationary, then

$$\begin{aligned} \mathrm{E}\left(\sum_{t=n}^{\infty} \beta^t \mathbf{r}(S_t) \Big| \mathbf{h}(n)\right) &= \mathrm{E}\left(\sum_{t=n}^{\infty} \beta^t \mathbf{r}(S_t) \Big| S_n\right) \\ &= \beta^n \mathrm{E}\left(\sum_{t=0}^{\infty} \beta^t \mathbf{r}(S_{t+n}) \Big| S_n\right) \\ &= \beta^n \, \mathbf{r}(\Gamma_s) \\ &\in \beta^n \, \mathbf{Co}^{(\beta)}(\Gamma_{S_n}) \end{aligned} \tag{3.49}$$

where (3.49) comes from hypothesis. Then, the thesis is proved. □

3.5.6 Core Allocations

Lemma 3.18 is not sufficient yet to provide a method to allocate the rate at each stage of the game. We should design both the rate allocation $\mathbf{r}(\Gamma_s)$ that each user gets in the long run process Γ_s and a rate allocation $\mathbf{r}(s)$ at each stage s of the Markov process, such that their discounted/average expected sum matches $\mathbf{r}(\Gamma_s)$. Of course, the best solution would be to find, under the β-discounted criterion,

$$\begin{aligned} &\Big(\{\mathbf{r}(\Gamma_s) \in \mathbf{Co}^{(\beta)}(\Gamma_s)\}_{s\in S},\ \{\mathbf{r}(s) \in \mathbf{Co}(\Psi_s)\}_{s\in S}\Big) \\ &\text{s.t. } \mathbf{r}(\Gamma_s) = \sum_{i=1}^{N} \nu^{(\beta)}(s_i, s)\, \mathbf{r}(s_i),\ \forall\, s \in S \end{aligned} \tag{3.50}$$

and, under the average criterion,

$$\begin{aligned} &\Big(\{\mathbf{r}(\Gamma_s) \in \mathbf{Co}(\Gamma_s)\}_{s\in S},\ \{\mathbf{r}(s) \in \mathbf{Co}(\Psi_s)\}_{s\in S}\Big) \\ &\text{s.t. } \mathbf{r}(\Gamma_s) = \sum_{i=1}^{N} \pi(s_i)\, \mathbf{r}(s_i),\ \forall\, s \in S \end{aligned} \tag{3.51}$$

Remark 3.9. Let us suppose the existence of greedy users, having a myopic perspective of the game and only care for the rate assigned to them in the current stage of the game. If, depending on the criterion considered, either condition (3.50) or (3.51) is verified, then also such users are content with the assigned rate allocation.

A promising result comes from the following Lemma.

Lemma 3.19. *Let $\mathcal{R}_1, \ldots, \mathcal{R}_k$ be k polymatroids on the same ground set A. Let g_i be the rank function of $\mathcal{R}_i$. Let $a_i > 0$, for all $1 \le i \le k$. Let $\mathcal{R} = \sum_{i=1}^{k} a_i \mathcal{R}_i$, with rank function $g \in \sum_{i=1}^{k} a_i g_i$ for Corollary 3.6. Then, the weighted Minkowski sum of the main facets $\sum_{i=1}^{n} a_i M(\mathcal{R}_i)$ coincides with $M(\mathcal{R})$.*

Proof. Let $\mathbf{r}(i) \in M(\mathcal{R}_i)$, for all $1 \le i \le k$. Then, it is easy to see from Corollary 3.6 that $\sum_{i=1}^{k} a_i \mathbf{r}(i) \in M(\mathcal{R})$. Hence, $\sum_{i=1}^{n} a_i M(\mathcal{R}_i) \subseteq M(\mathcal{R})$. Let now $\mathbf{r} \in \mathcal{R}$. Then there exists a convex combination such that

$$\mathbf{r} = \sum_{\sigma \in \mathcal{H}(|A|)} c(\sigma)\, \mathbf{w}(\sigma)$$

Hence, if we choose, for all $1 \le i \le k$,

$$\mathbf{r}(i) = \sum_{\sigma \in \mathcal{H}(|A|)} c(\sigma)\, \mathbf{w}(\sigma)(i)$$

then we can see that

$$\begin{aligned} \sum_{i=1}^{k} a_i \mathbf{r}(i) &= \sum_{\sigma \in \mathcal{H}(|A|)} c(\sigma) \sum_{i=1}^{k} a_i \mathbf{w}(\sigma)(i) \\ &= \sum_{\sigma \in \mathcal{H}(|A|)} c(\sigma) \mathbf{w}(\sigma) = \mathbf{r}. \end{aligned}$$

where the second equality comes from Corollary 3.8. So, $M(\mathcal{R}) \subseteq \sum_{i=1}^{n} a_i M(\mathcal{R}_i)$ and the thesis is proved. □

Average criterion

Thanks to Lemma 3.18 and Lemma 3.19 we can derive the first important result for rate allocation under the average criterion.

Theorem 3.24. *Under the average criterion, given a long run allocation $\mathbf{r}(\Gamma) \in \mathbf{Co}(\Gamma)$, it is always possible to find a set of single stage allocations:*

$$\left[\mathbf{r}(s_1) \ \dots \ \mathbf{r}(s_N)\right]^{\mathrm{T}} \in (\Phi)^{-1}(\mathbf{r}(\Gamma))$$

such that $\mathbf{r}(s) \in \mathbf{Co}(\Psi_s)$, for all $s \in S$.

Moreover, under the average criterion, the proof of Lemma 3.19 suggests a way to find the single stage allocations from the log run one.

Discounted criterion

Under the discounted criterion, the situation is more tricky. There exists a one-to-one linear transformation between the set $\{\mathbf{r}(\Gamma_s) \in M(\mathcal{R}^{(\beta)}(\mathcal{P}, \Gamma_s))\}_{s \in S}$ and $\mathbb{R}^{NP}$, namely

$$\left\{\mathbf{r}(s)\right\}_{s \in S} = \left(\Phi^{(\beta)}\right)^{-1} \left(\left\{\mathbf{r}(\Gamma_s) \in M(\mathcal{R}^{(\beta)}(\mathcal{P}, \Gamma_s))\right\}_{s \in S}\right)$$

such that $\mathbf{r}(s_1) \dots \mathbf{r}(s_N)$ is the only stage-wise allocation whose β-discounted sum, when the process starts in state s, equals $\mathbf{r}(\Gamma_s)$. Lemma 3.19 ensures that a single long run allocation is always reachable by stationary allocations. Under the β-discounted criterion, a set of N long run allocations need to be jointly reached by the same set of stationary allocation, which is not always verified. Indeed, it is quite easy to find a counterexample.

Counterexample 3.5.1. Set $\beta = 0.9$, $N_0 = 0.1$. Set $P = 2$, with power constraints $\Delta_1 = \Delta_2 = 1$. Consider two states. In s_1, $|h^{(1)}(s_1)|^2 = 0.1$, $|h^{(2)}(s_1)|^2 = 0.2$. In s_2, $|h^{(1)}(s_2)|^2 = 0.15$, $|h^{(2)}(s_2)|^2 = 0.3$. Let $\mathbf{P} = [0.8 \ 0.2; \ 0.3 \ 0.7]$. Choose the feasible allocations in the long run game

$$\mathbf{r}(\Gamma_{s_1}) \cong [0.67; \ 1.48] \in M(\mathcal{R}^{(0.9)}(\mathcal{P}, \Gamma_{s_1}))$$
$$\mathbf{r}(\Gamma_{s_2}) \cong [0.77; \ 1.52] \in M(\mathcal{R}^{(0.9)}(\mathcal{P}, \Gamma_{s_2}))$$

It is straightforward to check that the corresponding single stage allocations:

$$[\mathbf{r}(s_1) \ \mathbf{r}(s_2)]^{\mathrm{T}} = (\Phi^{(0.9)})^{-1} \left([\mathbf{r}(\Gamma_{s_1}) \ \mathbf{r}(\Gamma_{s_2})]^{\mathrm{T}}\right)$$
$$\mathbf{r}(s_1) \cong [0.0476; \ 0.1061] \notin \mathcal{R}(\mathcal{P}, s_1)$$
$$\mathbf{r}(s_2) \cong [0.1416; \ 1.1619] \notin \mathcal{R}(\mathcal{P}, s_2)$$

are both not feasible.

These considerations lead to the following fact.

Fact 3.5.1. For any set of single stage allocations $\{\mathbf{r}(s) \in \mathbf{Co}(\Psi_s)\}_{s\in S}$, the corresponding long run allocations $\{\mathbf{r}(\Gamma_s)\}_{s\in S}$ belong respectively to $\{\mathbf{Co}^{(\beta)}(\Gamma_s)\}_{s\in S}$. The converse is not true: some set of long run allocations $\{\mathbf{r}(\Gamma_s) \in \mathbf{Co}^{(\beta)}(\Gamma_s)\}_{s\in S}$ are not reachable, since $(\Phi^{(\beta)})^{-1}(\{\mathbf{r}(\Gamma_s))$ are not feasible stationary single stage allocations.

Therefore, our effort should be focused on devising a suitable set of reachable long run allocation $\{\mathbf{r}(\Gamma_s) \in \mathbf{Co}^{(\beta)}(\Gamma_s)\}_{s\in S}$, such that the associated single stage rate allocations $(\Phi^{(\beta)})^{-1}(\{\mathbf{r}(\Gamma_s))$ are all feasible and belonging to the Core of the respective games. The following Theorem suggests a simple way to do it.

Theorem 3.25. *Choose a set of convex coefficients $\{c(\sigma)\}$, for all $\sigma \in \mathcal{H}(P)$. Let $\mathbf{w}^{(\beta)}(\sigma)(\Gamma_s)$ be the vertex of $M(\mathcal{R}^{(\beta)}(\mathcal{P},\Gamma_s))$ associated to the permutation σ. Compute the set of long run allocations as*

$$\mathbf{r}(\Gamma_s) = \sum_{\sigma\in\mathcal{H}(P)} c(\sigma)\mathbf{w}^{(\beta)}(\sigma)(\Gamma_s), \quad \forall\, s \in S.$$

Then, the set of single stage allocations

$$\{\mathbf{r}(s)\}_{s\in S} = \left(\Phi^{(\beta)}\right)^{-1}(\{\mathbf{r}(\Gamma_s)\}_{s\in S})$$

are feasible and $\mathbf{r}(s) \in M(\mathcal{R}(\mathcal{P}, s))$, for all $s \in S$, i.e. condition (3.50) holds.

Proof. Let us write

$$\begin{bmatrix}\mathbf{r}(s_1)\\ \vdots \\ \mathbf{r}(s_N)\end{bmatrix} = (\mathbf{I}-\beta\mathbf{P}) \begin{bmatrix}\sum_{\sigma\in\mathcal{H}(P)} c(\sigma)\mathbf{w}^{(\beta)}(\sigma)(\Gamma_{s_1})\\ \vdots \\ \sum_{\sigma\in\mathcal{H}(P)} c(\sigma)\mathbf{w}^{(\beta)}(\sigma)(\Gamma_{s_N})\end{bmatrix}$$
$$= \sum_{\sigma\in\mathcal{H}(P)} c(\sigma)(\mathbf{I}-\beta\mathbf{P}) \begin{bmatrix}\mathbf{w}^{(\beta)}(\sigma)(\Gamma_{s_1})\\ \vdots \\ \mathbf{w}^{(\beta)}(\sigma)(\Gamma_{s_N})\end{bmatrix}$$

For Corollary 3.8, we can say that

$$\begin{bmatrix}\mathbf{r}(s_1)\\ \vdots \\ \mathbf{r}(s_N)\end{bmatrix} = \sum_{\sigma\in\mathcal{H}(P)} c(\sigma) \begin{bmatrix}\mathbf{w}(\sigma)(s_1)\\ \vdots \\ \mathbf{w}(\sigma)(s_N)\end{bmatrix}.$$

Hence, condition (3.50) holds and the thesis is proved. □

The method provided in Theorem 3.25 to choose a suitable long run rate allocation is not the only available of course, but leads to an intuitive remark. Each vertex $\mathbf{w}(\sigma)(s)$ can be achieved by letting the receiver decode sequentially, in the reverse order of σ, the signals coming from each user under the channel state $s \in S$, and by considering the signals not decoded yet as Gaussian noise (e.g. see [89, 90]). Therefore, any rate allocation on the main facet $M(\mathcal{R}(\mathcal{P}, s))$ can be achieved by time sharing such decoding configurations. Theorem 3.25 hence suggests that time sharing in the same way in each single stage is an efficient and fair strategy also in the long run process.

3.5.7 Fairness Criteria

Let us define three well known fair allocations.

Definition 3.14. An allocation $\mathbf{r}^{MM}$ is *max-min fair* whenever no user j with rate $\mathbf{r}_j$ can yield resources to a user i with $\mathbf{r}_i < \mathbf{r}_j$ without violating feasibility.

Definition 3.15. The α*-fair* allocation $\mathbf{r}^{\alpha F}$, with $\alpha \geq 0$, is defined as

$$\mathbf{r}^{\alpha F} = \underset{\mathbf{r} \in \mathcal{R}(\mathcal{P},s)}{\operatorname{argmax}} \prod_{i=1}^{P} \frac{\mathbf{r}_i^{1-\alpha}}{1-\alpha}.$$

Definition 3.16. The *proportional fair* allocation $\mathbf{r}^{PF}$ coincides with the α-fair allocation when $\alpha \to 1$, i.e.

$$\mathbf{r}^{PF} = \underset{\mathbf{r} \in \mathcal{R}(\mathcal{P},s)}{\operatorname{argmax}} \prod_{i=1}^{P} \mathbf{r}_i \tag{3.52}$$

From [81] we know the following result.

Theorem 3.26. *If the feasible rate region is a polymatroid, then the α-fair allocation coincides, for all $\alpha \geq 0$, with the max-min fair allocation and with the proportional fair allocation.*

Since the three allocation procedures are equivalent in our settings, we will encapsulate them in the concept of *general fair* allocation $\mathbf{r}^F$

$$\mathbf{r}^F \equiv \mathbf{r}^{MM} = \mathbf{r}^{PF} = \mathbf{r}^{\alpha F}, \quad \forall \, \alpha \geq 0.$$

We study the relation between the general fair allocation associated to each single stage game with the one associated to the long run game. Generally, if we assign a general fair allocation in each stage of the game, the associated long run allocation is not necessarily general fair in the long run game. Nevertheless, we will provide a sufficient condition for this property to be satisfied.

In [91], the following algorithm to compute the general fair allocation within a general polymatroid $\mathcal{R}$, with rank function g and on the set A, is provided.

Algorithm 3.5.1. Set $k \equiv 1$. Set $A' \equiv A$, $g' \equiv g$.

1) Compute

$$T^*_{(k)} = \underset{T \subseteq \mathcal{P}'}{\operatorname{argmin}} \frac{g'(T)}{|T|}$$
$$\mathbf{r}_i^F = \frac{g'(T^*_{(k)})}{|T^*_{(k)}|}, \quad \forall \, \{i\} \in T^*_{(k)}.$$

2) If $T^*_{(k)} = \mathcal{P}$, then stop. The rate allocation $\mathbf{r}^F$ is general fair for $\mathcal{R}$. Otherwise, set $k \equiv k+1$, $\mathcal{P}' \equiv \mathcal{P}' \backslash T^*_{(k)}$ and

$$g'(T) \equiv \; g'(T \cup T^*_{(k)}) - g'(T^*_{(k)}) \quad \forall \, T \subseteq \mathcal{P}'$$

Of course, Algorithm 3.5.1 can be utilized to compute the general fair allocation both in single stage games and in the long run process Γ_s for both the summation criteria.
The following Theorem is a straightforward extension of the results in [91], and shows that the general fair allocation is a useful procedure to select a point in the Core of the long run game.

Theorem 3.27. *The general fair allocation of a polymatroid $\mathcal{R}$ belongs to its main facet $M(\mathcal{R})$.*

We now present the main result of this section. It provides a sufficient condition to ensure that the general fair allocation employed in each stage of the process is still consistent in the long run process. Let $\mathbf{r}^{F(\beta)}(\Gamma_s)$ be the general fair allocation under the β-discounted game Γ_s and let $\mathbf{r}^F(\Gamma_s)$ be the general fair allocation under the average criterion.

Theorem 3.28. *Let $T^*(1)(s), \ldots, T^*(k_s)(s)$ be the sequence of sets computed in step 1 of Algorithm 3.5.1 applied to the single stage game s. Let $\mathbf{r}^F(s)$ be the general fair allocation in state $s \in S$. Suppose that the sequences*

$$\{T^*_{(1)}(s), T^*_{(2)}(s), \ldots\} = \{\widetilde{T}^*_{(1)}, \widetilde{T}^*_{(2)}, \ldots\} \quad \forall \, s \in S$$

i.e. are equal in all states. Then the long run allocations

$$\mathbf{r}^{F(\beta)}(\Gamma_s) = \; \sum_{i=1}^{N} \nu^{(\beta)}(s_i, s) \, \mathbf{r}^F(s_i)$$

in case of β-discounted criterion, and

$$\mathbf{r}^F(\Gamma_s) = \; \sum_{i=1}^{N} \pi(s_i) \, \mathbf{r}^F(s_i)$$

in case of average criterion, are general fair in the long run game Γ_s.

Proof. Let us consider the discounted criterion. The proof for the average one is totally similar. Let us apply Algorithm 3.5.1 to compute the max-min allocation $\mathbf{r}^F(\Gamma_s) \in M(\mathcal{R}^{(\beta)}(\mathcal{P}, \Gamma_s))$. At the first iteration, at step 1, for Lemma 3.16 we can write:

$$\begin{aligned} T^*_{(1)} &= \operatorname*{argmin}_{T \subseteq \mathcal{P}} \frac{g^{(\beta)}_{(\mathcal{P})}(T, \Gamma_s)}{|T|} \\ &= \operatorname*{argmin}_{T \subseteq \mathcal{P}} \frac{\sum_{i=1}^{N} \nu^{(\beta)}(s_i, s) \, g_{(\mathcal{P})}(T, s_i)}{|T|} \\ &= \widetilde{T}^*_{(1)} \end{aligned}$$

Hence, we can compute the max-min fair allocation for the set $T^*_{(1)}$:

$$\begin{aligned}\mathbf{r}^F_j(\Gamma_s) &= \frac{\sum_{i=1}^N \nu^{(\beta)}(s_i,s)\, g_{(\mathcal{P})}(\widetilde{T}^*_{(1)}, s_i)}{|\widetilde{T}^*|} \\ &= \sum_{i=1}^N \nu^{(\beta)}(s_i,s)\mathbf{r}^F_j(s_i), \quad \forall \{j\} \in \widetilde{T}^*_{(1)}.\end{aligned}$$

Then, at step 2, the update of the rank function, for all $T \subseteq \mathcal{P}\backslash\widetilde{T}^*_{(1)}$,

$$\begin{aligned}\left(g^{(\beta)}_{(\mathcal{P})}(T,\Gamma_s)\right)' &= \sum_{i=1}^N \nu^{(\beta)}(s_i,s)[g_{(\mathcal{P})}(T\cup\widetilde{T}^*_{(1)}, s_i) - g_{(\mathcal{P})}(T,s_i)] \\ &= \sum_{i=1}^N \nu^{(\beta)}(s_i,s)\left(g_{(\mathcal{P})}(T,s_i)\right)'\end{aligned}$$

preserves the linearity property of the rank function also in the following iteration. Hence, by induction, the thesis is proved. □

3.5.8 Nash bargaining solution

In the previous sections we have seen that the Core of the long run game is nonempty, and that it is always possible to find single stage allocations belonging to the Core of the single stage games such that their long run sum still belongs to the Core of the long run game. Hence, under this point of view, the grand coalition is stable throughout the game. Nevertheless, there is still the chance that some user may perceive its own rate allocation as unfairly too low with respect to some other users.
As an example, consider in state s a couple of users (i,j) such that their received power is

$$|h^{(i)}(s)|^2\Delta_i > |h^{(j)}(s)|^2\Delta_j.$$

Let $\mathbf{r}_i(s), \mathbf{r}_j(s)$ be the rate allocations for users i and j, respectively. Of course, user i can decrease its power to

$$\frac{|h^{(j)}(s)|^2\Delta_j}{|h^{(i)}(s)|^2}$$

and he will get a rate $\mathbf{r}'_i(s)$. La and Anantharam pointed out in [80] that, if $\mathbf{r}'_i(s) < \mathbf{r}_j(s)$, then user i may "envy" user j and decide not to join the grand coalition, and in the worst case scenario to jam. Therefore, it still makes sense to probe the situation in which no agreement is reached among the users.

Let us focus on the static game in a state $s \in S$. We want to define the Nash bargaining solution for MAC with jamming users. We stress that in literature (e.g. see [92]) the disagreement point is generally set to 0 for each user. In our case, in which the game

structure is clear, it makes sense to define it as the set of maximum rate that each user can ensure for itself without agreement.
Let us first show the following result.

Proposition 3.8. Let $\mathbf{d}_i(P,s)$ be the rate that user $i \in [1;P]$ can ensure in state s when no cooperation is reached. In each state s there exists at most one user k such that $\mathbf{d}_i(P,s) > 0$.

Proof. User k, with

$$|h^{(k)}(s)|^2\Delta_k > \left(\sum_{i=1,i\text{NE}k}^{P} |h^{(i)}(s)|\sqrt{\Delta_i}\right)^2 = \Lambda(\mathcal{P}\backslash\{k\},s) \tag{3.53}$$

can ensure for itself a positive rate, which we call $\mathbf{d}_k(P,s)$:

$$\mathbf{d}_k(P,s) = C\left(|h^{(k)}(s)|^2\Delta_k, \Lambda(\mathcal{P}\backslash\{k\},s) + N_0\right).$$

Then,

$$|h^{(k)}(s)|^2\Delta_k > \sum_{i=1,i\text{NE}k}^{P} |h^{(i)}(s)|^2\Delta_i$$

and, for any user $j\text{NE}k$,

$$\begin{aligned} |h^{(j)}(s)|^2\Delta_j &< |h^{(k)}(s)|^2\Delta_k - \sum_{i=1,i\text{NE}k,j}^{P} |h^{(i)}(s)|^2\Delta_i \\ &< \left(\sum_{i=1,i\text{NE}j}^{P} |h^{(i)}(s)|\sqrt{\Delta_i}\right)^2 \end{aligned}$$

Therefore,

$$\mathbf{d}_i(P,s) = 0 \quad \forall\, i\text{NE}k.$$

□

Now we can define the Nash bargaining solution in our setting.

Definition 3.17. Fix the state $s \in S$. Let

$$\mathbf{d}_k(P,s) = \max_{1\le i\le P} \mathbf{d}_i(P,s) \ge 0.$$

The *Nash bargaining solution* $\mathbf{r}^{NB}(s)$ in MAC with static coefficients and jamming users is

$$\mathbf{r}^{NB}(P,s) = \underset{\substack{\mathbf{r}\in\mathcal{R}(\mathcal{P},s)\\ \mathbf{r}_k\ge\mathbf{d}_k(P,s)}}{\text{argmax}} \left(\mathbf{r}_k - \mathbf{d}_k(P,s)\right)\prod_{\substack{i=1\\ i\text{NE}k}}^{P} \mathbf{r}_i. \tag{3.54}$$

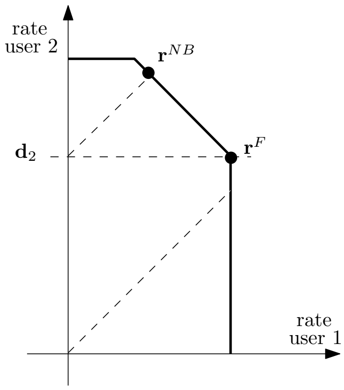

Figure 3.2: General fair and Nash bargaining allocations.

Analogously, the Nash bargaining solution in the Markovian game can be defined in the following way.

Definition 3.18. Set $\beta \in [0;1)$. Set the initial state $s \in S$. Compute the β-discounted disagreement point as

$$\mathbf{d}_j^{(\beta)}(P,\Gamma_s) = \sum_{i=1}^{P} \nu^{(\beta)}(s_i,s)\,\mathbf{d}_j(P,s_i), \quad \forall\, 1 \le j \le P.$$

Under β-discounted criterion, the *Nash bargaining solution* $\mathbf{r}^{NB(\beta)}(\Gamma_s)$ in Markovian MAC with static coefficients and jamming users is

$$\mathbf{r}^{NB(\beta)}(P,\Gamma_s) = \underset{\substack{\mathbf{r}\in\mathcal{R}^{(\beta)}(\mathcal{P},\Gamma_s)\\ \mathbf{r}_i \ge \mathbf{d}_i^{(\beta)}(P,\Gamma_s),\forall\, i}}{\operatorname{argmax}} \prod_{\substack{i=1\\ i\text{NE}k}}^{P} \big(\mathbf{r}_i - \mathbf{d}_i^{(\beta)}(P,\Gamma_s)\big).$$

Definition 3.19. Compute the average disagreement point as

$$\mathbf{d}_j(P,\Gamma) = \sum_{i=1}^{P} \pi(s_i)\,\mathbf{d}_j(P,s_i), \quad \forall\, 1 \le j \le P.$$

Under average criterion, the *Nash bargaining solution* $\mathbf{r}^{NB}(\Gamma_s)$ in Markovian MAC with static coefficients and jamming users is

$$\mathbf{r}^{NB}(P,\Gamma_s) = \underset{\substack{\mathbf{r}\in\mathcal{R}(\mathcal{P},s)\\ \mathbf{r}_i \ge \mathbf{d}_i(P,\Gamma),\forall\, i}}{\operatorname{argmax}} \prod_{\substack{i=1\\ i\text{NE}k}}^{P} \big(\mathbf{r}_i - \mathbf{d}_i(P,\Gamma)\big).$$

By comparing (3.54) with (3.52) it is evident that $\mathbf{r}^{NB} = \mathbf{r}^F$ whenever $\mathbf{d}_i(P) = 0$ for all i, i.e. when no user has sufficient available power and channel conditions good enough to be still able to communicate when no agreements is reached.

We now want to investigate the relation between the Nash bargaining solution and the general fair allocation in the single stage game when the number of user increases. To do so in probabilistic terms, we suppose that there exists a probability distribution among the states which preserves the independence of the channel coefficients among the users. Under this assumption, we will show that the probability to be in a state s such that there exists $k:\ \mathbf{d}_k(P,s) > 0$ tends to 0 when the number of users P tends to infinity. Hence, the probability that $\mathbf{r}^{NB}{}_{\text{NE}}\mathbf{r}^F$ tends to 0 as well.

First, we show a general result on independent random variables.

Lemma 3.20. *Let $\xi_1, \xi_2, \ldots$ be a countable set of independent positive random variables, with $\mathrm{E}(\xi_i) = m_1$ and $\mathrm{Var}(\xi_i) = v_i^2$. Let $\inf_i m_i > 0$ and $\sup_i v_i^2 < \infty$. Let the probability density function of ξ_i be bounded and with limited support, for all i. Then,*

$$\Pr\left(\xi_j \leq \sum_{i=1, i\text{NE}j}^{P} \xi_i,\ \forall\, 1 \leq j \leq P\right) \xrightarrow{P\uparrow\infty} 1. \tag{3.55}$$

with exponential rate of convergence in P.

Proof. Let $\underline{\mathrm{m}} \equiv \inf_i m_i > 0$ and $\overline{v}^2 \equiv \sup_i v_i^2$. Define the event

$$E_j^{(P)} \equiv \left(\xi_j \leq \sum_{i=1, i\text{NE}j}^{P} \xi_i\right).$$

Then, we can rewrite (3.55) as

$$\begin{aligned}
\Pr\left(\cap_{j=1}^{P} E_j^{(P)}\right) &= 1 - \Pr\left(\cup_{j=1}^{P} \overline{E}_j^{(P)}\right) \\
&= 1 - \sum_{j=1}^{P} \Pr\left(\overline{E}_j^{(P)}\right) && (3.56) \\
&\geq 1 - P \max_{j=1,\ldots,P} \Pr\left(\overline{E}_j^{(P)}\right) && (3.57)
\end{aligned}$$

where $\overline{E}_j^{(P)}$ is the complementary event of $E_j^{(P)}$ and (3.56) comes from the exclusivity of the events. We can prove that the (3.57) tends to 1 when P tends to infinity. Hence, it is sufficient to prove that

$$P \Pr\left(\overline{E}_j^{(P)}\right) \xrightarrow{P\uparrow\infty} 0, \quad \forall\, j = 1, \ldots, P. \tag{3.58}$$

Let $f_{\sum_i \xi_i}(.)$ be the probability density function, possibly discrete, of $\sum_i \xi_i$. Let $f_{\xi_i}(.)$ be bounded between $[0; a_i]$. For simplicity of notation, we show (3.58) for $j = 1$.

$$P \Pr\left(\overline{E}_1^{(P)}\right) = P \int_0^{a_1} f_{\xi_1}(x_1) \int_0^{x_1} f_{\sum_{j=2}^{P} \xi_i}(x_2)\, dx_2\, dx_1.$$

From the central limit Theorem, we can say that

$$\lim_{P\uparrow\infty} P \Pr\left(\overline{E}_1^{(P)}\right) =$$
$$\lim_{P\uparrow\infty} P \int_0^{a_1} f_{\xi_1}(x_1) \int_0^{x_1} \mathcal{N}\left(\sum_{j=2}^{P} m_i, \sum_{j=2}^{P} v_i^2\right)(x_2)\, dx_2\, dx_1.$$

where $\mathcal{N}(a,b)(.)$ is a gaussian univariate distribution with mean a and variance b. We continue by writing

$$\begin{aligned} P\int_0^{a_1} f_{\xi_1}(x_1) \int_0^{x_1} \mathcal{N}\left(\sum_{j=2}^{P} m_i, \sum_{j=2}^{P} v_i^2\right)(x_2)\, dx_2\, dx_1 \leq \\ \frac{P}{2}\int_0^{a_1} f_{\xi_1}(x_1) \exp\left(-\frac{(x_1 - \sum_{j=2}^{P} m_j)^2}{\sum_{j=2}^{P} v_j^2}\right) dx_1 \leq \\ \frac{P}{2}\int_0^{a_1} f_{\xi_1}(x_1) \exp\left(-\frac{(x_1 - [P-1]\,\underline{\mathrm{m}})^2}{[P-1]\,\overline{v}^2}\right) dx_1 \leq \\ \frac{P}{2}\exp\left(-\frac{(a_1 - [P-1]\,\underline{\mathrm{m}})^2}{[P-1]\,\overline{v}^2}\right) \max_x f_{\xi_1}(x) \stackrel{P\uparrow\infty}{\longrightarrow} 0 \end{aligned}$$

where the first inequality comes from the Chernoff bound on the tail probability of Gaussian variables, and the second and the third inequality are valid for all P large enough. Hence, (3.58) holds and the thesis is proved. □

Let us now make the following assumption on the distribution of the channel coefficients.

Assumption 3.7. The channel coefficients $\{h^{(i)}(s)\}_{i\in\mathbb{N}}$, under the probability distribution μ on S, are independent and bounded variables, with

$$\inf_i \mathrm{E}_\mu\left(|h^{(i)}(s)|^2\right) > 0, \quad \sup_i \mathrm{Var}_\mu\left(|h^{(i)}(s)|^2\right) < \infty.$$

Lemma 3.21. *Under Assumption 3.7, the probability that no user can still ensure for itself a positive rate without agreement tends to 1 when the number of users tends to infinity, i.e.*

$$\Pr\left(\mathbf{d}_i(P,s) = 0,\ \forall\, 1 \leq i \leq P\right) \stackrel{P\uparrow\infty}{\longrightarrow} 1.$$

Proof. The probability of the event

$$|h^{(k)}(s)|^2\Delta_k \leq \left(\sum_{i=1, i\mathrm{NE}k}^{P} |h^{(i)}(s)|\sqrt{\Delta_i}\right)^2,\ \forall\, k = 1,\ldots,P. \tag{3.59}$$

is not smaller than

$$\Pr\left(|h^{(k)}(s)|^2\Delta_k \leq \sum_{i=1, i\mathrm{NE}k}^{P} |h^{(i)}(s)|^2\Delta_i,\ \forall\, k = 1,\ldots,P\right). \tag{3.60}$$

For Lemma 3.20, (3.60) tends to 1 when P tends to infinity, so also the probability of (3.59) does. Hence, the thesis is proved. □

Hence, we obtain that the probability that the Nash bargaining solution coincides with the general fairness allocation tends to 1 when the number of users increases.

Theorem 3.29. *Under Assumption 3.7,*

$$\Pr\left(\mathbf{r}^{NB}(P,s) = \mathbf{r}^{F}(P,s)\right) \stackrel{P\uparrow\infty}{\longrightarrow} 1.$$

3.5.9 No cooperation in Markovian channels

In this section we want to prove the importance of reaching an agreement in the Markovian channel. Let us first derive an accessory result, on the sum rate in the single stage game, assuming a distribution probability on the states.

Lemma 3.22. *Suppose that Assumption 3.7 holds. Then,*

$$\sum_{i=1}^{P} \mathrm{E}_{\mu}(\mathbf{d}_i(P,s)) \stackrel{P\uparrow\infty}{\longrightarrow} 0$$

with exponential rate of convergence in P.

Proof. Let $|\overline{h}|^2$ be the supremum that a channel coefficient can achieve over all the states. Let $\underline{m} \equiv \inf_i \mathrm{E}_{\mu}(|h^{(i)}(s)|^2)$ and $\overline{v}^2 \equiv \sup_i \mathrm{Var}_{\mu}(|h^{(i)}(s)|^2)$. Let $\overline{\Delta}$ be the maximum possible power constraint for users. We write

$$\sum_{i=1}^{P} \mathrm{E}(\mathbf{d}_i(P)) \leq P \max_{1\leq i\leq P} \mathrm{E}_{\mu}(\mathbf{d}_i(P,s))$$

It is sufficient to prove that $P\mathrm{E}_{\mu}(\mathbf{d}_i(P,s))$ tends to 0 when $P\uparrow\infty$, for all $1 \leq i \leq P$. Then, for Lemma 3.20, for all i,

$$\begin{aligned} P\mathrm{E}_{\mu}(\mathbf{d}_i(P,s)) &\leq P \Pr(\mathbf{d}_i(P,s)_{\text{NE}} 0)\, C\left(|\overline{h}|^2\overline{\Delta}, N_0\right) \\ &\stackrel{P\uparrow\infty}{\longrightarrow} 0 \end{aligned}$$

where the rate of convergence is exponential, still for Lemma 3.20. Hence, the thesis is proved. □

We now suppose to know the distribution of the channel state on S at the beginning of the game is $\mu_0(.)$. In this case, under the discounted criterion, the feasibility region averaged over μ_0 can be expressed as

$$\mathcal{R}^{(\beta)}(\mathcal{P},\Gamma_{\mu_0}) = \sum_{i=1}^{N} \mu_0(s_i)\, \mathcal{R}^{(\beta)}(\mathcal{P},\Gamma_{s_i})$$

which is still a polymatroid, according to Corollary 3.6. Under the average criterion, the average of the capacity region $\mathcal{R}(\mathcal{P}, \Gamma_s)$ over μ_0 is still $\mathcal{R}(\mathcal{P}, \Gamma)$, since it does not depend on the initial state.

We want now to show under which conditions the sum throughput of the network in the Markovian game collapses for $P \uparrow \infty$, when no agreement is reached among the users. Let $\mathbf{d}_j(S_t, P)$ be the rate $\mathbf{d}_j(S_t, P)$ at time step t. We define $\mathbf{d}_j^{(\beta)}(\Gamma_{\mu_0}, P)$ as the rate that user j can ensure when no agreement is found:

$$\mathbf{d}_j^{(\beta)}(\Gamma_{\mu_0}, P) = \sum_{s \in S} \mu_0(s) \sum_{i=1}^{N} \nu^{(\beta)}(s_i, s)\, \mathbf{d}_j(P, s_i)$$

in the case of β-discounted criterion and

$$\mathbf{d}_j(\Gamma_{\mu_0}, P) = \sum_{i=1}^{N} \pi(s_i)\, \mathbf{d}_j(P, s_i) \equiv \mathbf{d}_j(\Gamma, P)$$

under the average criterion, since it does not depend on μ_0.

We now make an usual assumption on MAC, for which the channel HMC can be splitted into P independent processes.

Assumption 3.8. For any integers P, t,

$$p\Big(h^{(1)}(S_{t+1}), \ldots, h^{(P)}(S_{t+1}) \,\big|\, h^{(1)}(S_t), \ldots, h^{(P)}(S_t)\Big) = \prod_{i=1}^{P} p\left(h^{(i)}(S_{t+1}) \middle| h^{(i)}(S_t)\right)$$

Let $S^{(i)}$ be the set of possible states for the channel coefficient $h^{(i)}$, such that $S^{(i)} \equiv \{h^{(i)}\}_{s \in S}$ and S is the Cartesian product $\prod_{i=1}^{P} S^{(i)}$. Let $\pi^{(i)}$ be the stationary distribution of the irreducible HMC $\{S_t^{(i)},\ t \geq 0\}$.

Proposition 3.9. If Assumption 3.8 holds, then $\pi(s)$, where

$$\pi(s) = \prod_{i=1}^{P} \pi^{(i)}(h^{(i)}(s))$$

is the stationary distribution of the HMC on S.

Proof. We can write $\pi(s)$ as

$$
\begin{aligned}
\pi(s) &= \sum_{s' \in S} \pi(s')\, p(s|s') \\
&= \sum_{s' \in S} \prod_{i=1}^{P} \pi^{(i)}(h^{(i)}(s'))\, p(h^{(i)}(s)|h^{(i)}(s')) \\
&= \prod_{i=1}^{P} \sum_{s' \in S} \pi^{(i)}(h^{(i)}(s'))\, p(h^{(i)}(s)|h^{(i)}(s')) \\
&= \prod_{i=1}^{P} \pi^{(i)}(h^{(i)}(s)).
\end{aligned}
$$

□

The following result directly follows from Proposition 3.9.

Corollary 3.9. *Suppose that Assumption 3.8 holds. If $\mu_0 = \pi$, i.e. the stationary distribution is also the initial distribution of the HMC, then the variables $h^{(1)}(S_t), \ldots, h^{(P)}(S_t)$ are independent and distributed as $\pi^1, \ldots, \pi^{(P)}$ respectively, for any $t \geq 0$.*

We are now ready to show our main result of this section. It states that, under Markovian assumption, the guaranteed throughput of the network in case no agreement is found goes to 0 when the number of users P tends to infinity.

Theorem 3.30. *Suppose that Assumption 3.8 holds and that Assumption 3.7 is valid for the distribution π. Then, in the case of β-discounted criterion, the sum rate:*

$$
\sum_{i=1}^{P} \mathbf{d}_i^{(\beta)}(\Gamma_\pi, P) \overset{P \uparrow \infty}{\longrightarrow} 0
$$

and, under the average criterion, the sum rate:

$$
\sum_{i=1}^{P} \mathbf{d}_i(\Gamma_\pi, P) \overset{P \uparrow \infty}{\longrightarrow} 0.
$$

Proof. We can write, in the case of β-discounted criterion,

$$
\sum_{j=1}^{P} \mathbf{d}_j^{(\beta)}(\Gamma_\pi, P) = \sum_{t=0}^{\infty} \beta^t \sum_{j=1}^{P} \mathrm{E}_{S_t} \left(\mathbf{d}_j(S_t, P) \right) \tag{3.61}
$$

and, under the average criterion,

$$
\sum_{j=1}^{P} \mathbf{d}_j^{(\beta)}(\Gamma, P) = \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \sum_{j=1}^{P} \mathrm{E}_{S_t} \left(\mathbf{d}_j(S_t, P) \right). \tag{3.62}
$$

Thanks to Corollary 3.9, we can invoke Lemma 3.22 to say that

$$\sum_{j=1}^{P} \mathrm{E}_{S_t}\left(\mathbf{d}_j(S_t, P)\right) \stackrel{P\uparrow\infty}{\longrightarrow} 0, \quad \forall\, t \geq 0.$$

Hence, the thesis is proved. □

4 Game Theory for Spectrum and Infrastructure Sharing Networks

We directed our efforts into two directions, i.e. *spectrum* and *infrastructure* sharing.

With reference to the spectrum sharing, we considered the scenario where several operators decide to share with each other a part of their spectrum [93]. However, this choice requires the coordination for the access to the shared resources in order to avoid another resource waste due to the so called "Tragedy of the commons" [94]. Many algorithms can be proposed, belonging to two main categories: *orthogonal* and *non-orthogonal*. The former considers mutually exclusive access to the shared spectrum and hence does not tolerate any interference. The latter allows several BSs use the same transmission frequency at the same time, provided that the level of interference at the intended receivers is below a desired threshold. All of them have pros and cons and enable different gains for the operators. In our work, we focused on the orthogonal case and quantified an *upper bound* on the achievable gain. To do that, we proposed in [95] a centralized algorithm which exploits perfect channel information and performs the optimal allocation in a coordinated manner. Even though it is not directly amenable to practical implementation, due to the conflicting operator interests, this solution aims at calculating an upper bound on the sum capacity that can be reached in each cell. This study is presented in Section 4.1. A cooperative context is considered where the network operators are not selfish and try to maximise a common metric, i.e. the total system capacity.

With reference to infrastructure sharing, we studied the case of relay sharing in wireless networks ([96,97]). In particular, in Section 4.2 we consider the case of a network with a a cross-layer interaction between routing and medium access control and investigate the proper cooperation mechanism to be adopted by the users so that a gain is obtained by both those who have their transmission relayed to the final destination and also those who act as relays. In Section 4.3, we consider Bayesian Networks to compute the correlation between local network parameters and overall performance, whereas the selection of the nodes to be shared by a nework operator is made by means of a game theoretic approach.

4.1 An Upper Bound on Capacity Gains due to Orthogonal Spectrum Sharing

We consider two Long Term Evolution (LTE) mobile networks, operating in the same geographical region, i.e. having adjacent cells. In particular, for the sake of simplicity we consider the case of two cells. However, all considerations can be promptly extended to a more general case. We assume that the network operators own adjacent portions of the spectrum and agree upon sharing a part of them.

In LTE networks, the downlink of a cell is organized according to a TDMA/OFDMA scheme. The time is divided into 10 ms frames, each consisting of 10 sub-frames of 1 ms. The spectrum is divided into bundles of adjacent sub-carriers, called *sub-channels*, each one of which is being assigned to a single user for an entire sub-frame. Thus, the resources to be allocated by the BS managing the cell are represented by the sub-channels, and the allocation decision is taken per-sub-frame basis. The choice of which user to allocate on each sub-channel depends on the scheduling and resource allocation policy adopted, i.e. on the utility functions that are to be optimised by the allocation entity.

Let $\mathcal{K}$ be the set of available sub-channels for the downlink, split into two subsets $\mathcal{K}_1$ and $\mathcal{K}_2$, where $\mathcal{K}_j$ is assigned to operator j. Denote with $K = |\mathcal{K}|$, $K_j = |\mathcal{K}_j|$ and with $\alpha \in [0, 1]$ the sharing percentage, i.e. the fraction of spectrum each operator decides to share. For the sake of simplicity, we assume $|\mathcal{K}_1| = |\mathcal{K}_2| = K/2$ and indicate it as k. The extension to the general case is straightforward. The k sub-channels are split into a part $k^s = k\alpha$ that is shared, and a part k^p that remains private to the operator, with $k = k^s + k^p$. In this way, each operator has a final set of k^F available sub-channels that is made of its initial k plus the k^s shared by the other,

$$k^F = k(1 + \alpha). \tag{4.1}$$

We denote with k^c the number of sub-channels which compose the common pool (i.e. $k^c = 2k^s$). Since in this work we consider the case of *orthogonal spectrum sharing*, the access to the common resources is mutually exclusive. Therefore, coordination for the usage of such resources is needed.

We remark that, since OFDMA is used for the downlink access of multiple users, no interference occurs among different sub-channels and thus the transmit power on each one of them can be set to the maximum possible. The performance metric taken into consideration for operator m is the cell sum capacity, i.e. the sum of the Shannon capacities achievable on each sub-channel allocated, defined as

$$\begin{aligned} C_m &= \textstyle\sum_{i=1}^{N_{UE}} \sum_{j=1}^{k^F} B \log_2(1 + SINR_{ij} \cdot \delta_{ij}), \\ \delta_{ij} &= \begin{cases} 1, & \text{UE}_i \text{ allocated to sub-channel}_j \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4.2}$$

where B and N_{UE} are, respectively, the sub-channel bandwidth and the number of User Equipments (UEs) in the cell, while $SINR_{i,j}$ is the Signal-to-Interference-and-Noise Ratio (SINR) perceived by UE i on sub-channel j. Note that the actual cell throughput is lower than the capacity and depends on the modulation and coding scheme chosen for each sub-channel on the basis of the corresponding SINR.

4.1.1 Coordinated Scheduling Algorithm

At first, each BS runs independently its internal resource allocation procedure; then, the trading for the common pool usage starts. Therefore, two phases can be identified: (i) the proposal of a resource allocation and (ii) the contention resolution.

For every allocation opportunity, each BS belonging to an operator decides which are the sub-channels, among those it is entitled to use, that will be used in that allocation period and the UE the resource will be assigned to. We call the set of pairs <*sub-channel, UE*> as *channel allocation map*. We assume that the BS manages a different flow for each UE registered to it, and these flows are always backlogged, so that every time a flow is selected there is always a packet to transmit. Another fundamental assumption that we do is regarding the internal scheduling policy adopted by each BS. We suppose that short-term fairness among the flows is not taken into consideration, thus the only objective of the allocator is the maximisation of the cell capacity. Indeed, if fairness was taken into consideration, then the total throughput reached by a BS would be reduced, since the achievement of a fair situation could force the allocator to sub-optimal choices, as discussed in depth in [98].

The problem of selecting the channel allocation map that maximises the capacity is combinatorial in nature and thus with a high computational complexity. In a more formal way, each UE has a Channel Quality Indicator (CQI) for each sub-channel. We denote $CQI_{i,j}^{l}$ the CQI value of user j, served by BS i, for the sub-channel l. Each BS selects, for each sub-channel l, the user u_i^l with the largest CQI so as to maximise the sum capacity. In other words,

$$u_i^l = \arg\max_j CQI_{i,j}^l,\ l = 1, \ldots, k^F, \tag{4.3}$$

$$CQI_i^l = \max_j CQI_{i,j}^l,\ l = 1, \ldots, k^F. \tag{4.4}$$

In this way the BS i constructs the proposed allocation vector $\mathbf{u}_i = [u_i^1, u_i^2, \ldots, u_i^k, u_i^{k+1}, \ldots, u_i^{k^F}]$. If the resources that BS i can allocate are fewer than k^F, BS i selects the best k elements according to the value of CQI, $[u_i^{m_1}, u_i^{m_2}, \ldots, u_i^{m_k}]$, where $m_s \in \mathcal{K}$.

When each BS has determined its proposal of resource allocation, the trading phase starts in order to solve all the possible contentions for resource access and determine the final allocation maps. Many algorithms can be proposed to take care of the conflicts, each of them having pros and cons and reaching different performance. We propose such an algorithm, without considering the implementation aspects of it, for the purpose of establishing an upper bound on the cell capacity achieved in total by both operators.

This is a centralized algorithm that obtains an upper bound on the gain that can be achieved by distributed algorithms enabling orthogonal spectrum sharing. Even though it is not meant as a practical scheme, it can be used to benchmark the performance of other strategies. In particular, the proposed solution aims at maximising the total cell capacity. To do so, the operators behave as if they were a unique entity, a kind of *monopolist* having complete information on both cells, and allocate each sub-channel to the best UE, i.e. the one having the largest Channel Quality Indicator (CQI), without taking into consideration any fairness constraints amongst operators. The resulting capacity is hence the theoretical limit, i.e. the maximum achievable by full coordination. Of course, when deciding the allocation of the common sub-channels, the UE can belong to any of the involved operators. On the other hand, when allocating the sub-channels private to a certain operator, then only its UEs can be taken into consideration. The pseudo-code for the algorithm is given hereafter.

```
forall sub-channels j ∈ 𝒦 do
    if (j ∈ 𝒦^c) then                    /* it is a common sub-channel */
        if (CQI^j_{1,u^j_1} > CQI^j_{2,u^j_2}) then
            u^j = u^j_1;
        end
        u^j = u^j_2;
    else                                  /* it is private sub-channel */
        u^j = u^j_s;             /* s is the owner of sub-channel j */
    end
end
```

Algorithm 2: Coordinated Scheduling Algorithm

In this way, the final allocation vector $\mathbf{u} = (u^k)_{k \in \mathcal{K}}$ is constructed and contains the indication of which UE to allocate on each sub-channel.

4.1.2 Numerical Results

We run an extensive simulation campaign to assess the performance of the coordinated scheduling algorithm. To better appreciate the sharing gain achievable by orthogonal spectrum sharing, we considered two different scenarios: (i) symmetric load scenario, and (ii) asymmetric load scenario. We consider two partially overlapping disk-shaped cells and random user droppings. Table 4.1 contains the main system parameters common to both scenarios. In particular, those related to the PHY layer are taken directly from the LTE standard. The simulations were run by using the network simulator ns-3 [99] with the extension defined in [100] for the support of multi-cell multi-operator LTE networks with spectrum sharing.

Symmetric load scenario. Both cells are under saturation (i.e. all the UEs have an infinite number of packets to transmit) and each BS tries to use as many sub-channels as possible, i.e. k^F. The two cells are statistically equivalent and thus the average performance achieved by both coincides.

Asymmetric load scenario. In this scenario, the two BSs have different traffic load and the one with higher traffic can opportunistically exploit the shared resources not used by the other one. In order to model this scenario, we assume that each BS guarantees to

Parameter	**Value**
Center frequencies	2.115 GHz (BS0), 2.125 GHz (BS1)
Downlink channel bandwidth	10 MHz
Subcarrier bandwidth	15 kHz
Doppler frequency	60 Hz
Sub-channel bandwidth	180 kHz
Subcarriers per sub-channel	12
OFDM symbols per sub-carrier	14
TX power per sub-channel	27 dBm
Noise spectral density (N_0)	-174 dBm/Hz
Pathloss	$128.1 + (37.6 \cdot \log_{10}(d))$ dB (d is the BS-UE distance in km)
Shadow fading	log-normal ($\sigma = 8$ dB)
Multipath	Jakes model with 6 to 12 scatterers
Wall penetration loss	10 dB
Frame duration	10 ms
TTI	1 ms
Cell radius	1500 m
Distance between BSs	1000 m

Table 4.1: Main system parameters

each UE no more than two sub-channels at each allocation opportunity. The UEs are always allocated according to their CQI, so as to exploit the multiuser diversity as much as possible. In particular, each BS tries to give 2 time-frequency resources to each of its users, as long as there are free sub-channels in the part of the spectrum it can access. Therefore, the asymmetric load is generated by varying the number of users. Cell 1 is overloaded and has 40 UEs to serve; thus, it needs 80 sub-channels. For cell 2 we considered a varying number of UEs, ranging from 2 up to 40. When the sharing percentage is 0%, BS1 cannot use more than 50 sub-channels, independently from the load of cell 2. This may lead to large resource waste when the latter is underloaded. When the sharing percentage increases, BS1 is entitled to use more sub-channels of the other operator and thus a larger number of UEs can be served. Therefore, the performance of BS1 is limited by the sharing percentage and by the load of cell 2, which needs some of the available resources to satisfy its users. When both cells have the same load, also the respective achievable channel capacity is the same because they are statistically equivalent.

All the results are characterized by a 95% confidence interval with a maximum relative

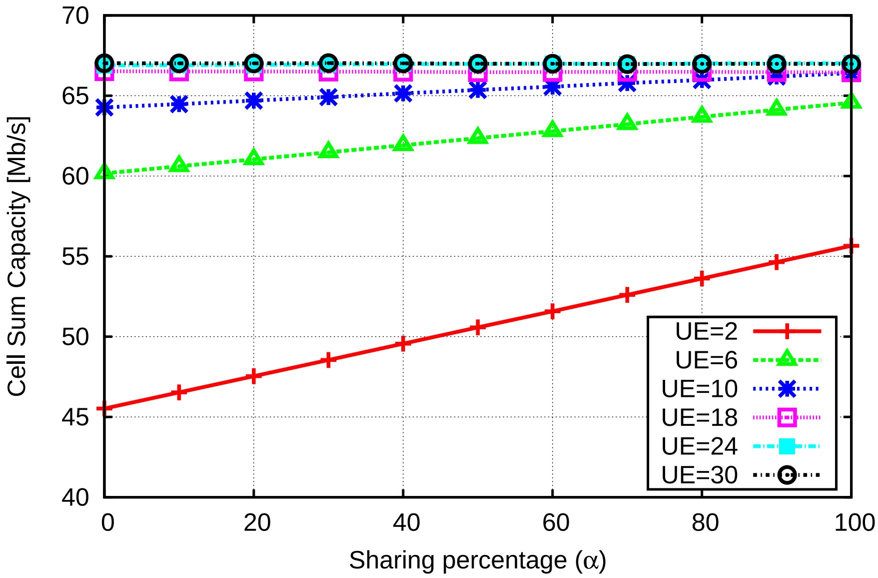

Figure 4.1: Cell sum capacity versus sharing percentage for various number of UEs and symmetric load

error of 5%. Fig. 4.1 shows the sum capacity reached by each cell in the symmetric load scenario, for various numbers of UEs, when the sharing percentage increases. Since both cells are statistically equivalent (for number of UEs, channel state, traffic load), also the respective final outcome is the same. First of all, a clear increment of the capacity with the number of UEs can be noted, which is a direct consequence of the multiuser diversity. The larger the number of UEs, the higher the probability that for each sub-channel there is at least one UE with good channel quality. However, for a cell with higher user density the improvement is quite low because for almost all the sub-channels there is at least one user in good situation. The second important observation that can be done is that there is a neat *sharing gain*. The cell sum capacity increases with the sharing percentage, thus there is an incentive for the network operators to share part of their frequencies.

Fig. 4.2 shows the corresponding *spectrum sharing gain*, i.e. the ratio between the cell capacity achieved with and without spectrum sharing. For small number of UEs and full sharing, 20% gain can be reached over the non-sharing case. As discussed previously, the sharing gain is greater for scenarios with few UEs, and tends to reduce for situations with more users, because additional (due to spectrum sharing) multiuser diversity has a diminishing returns effect. This aspect is more evident in Fig. 4.3, where the total (i.e. over the two cells) sum capacity is depicted as a function of the number of UEs in each cell. In this case, the saturation effect for cells with more users is more evident: after 18 UEs the improvement of the capacity is almost negligible, as discussed above.

Fig. 4.4 depicts the corresponding total sum capacity for the scenario with asymmetric traffic, as a function of the number of UEs in cell 2 (essentially modeling the traffic load

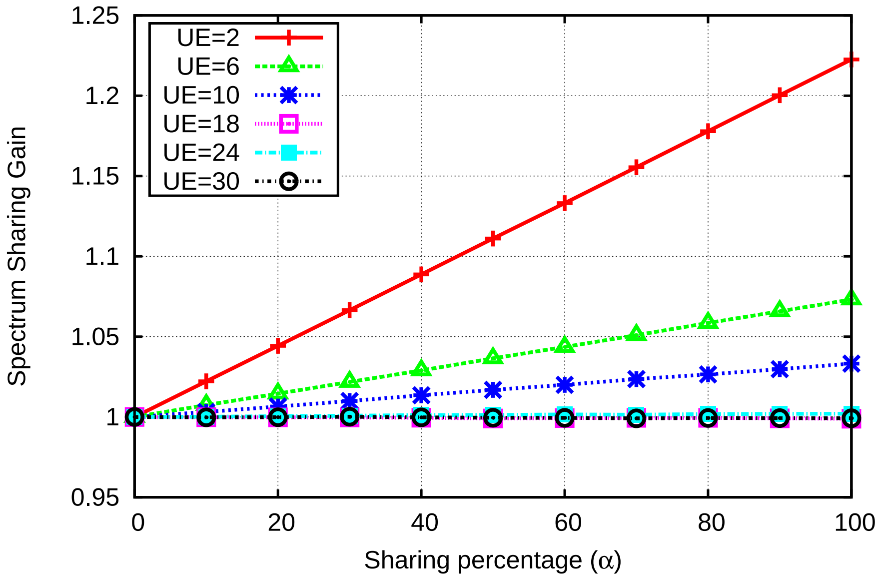

Figure 4.2: Spectrum sharing gain versus sharing percentage for various UEs and symmetric load

of cell 2). Three values for the percentage α of spectrum sharing have been considered, i.e. 0, 50 and 100%. For the reference case of no sharing ($\alpha = 0\%$), the curve increases due to the increasing number of UEs that are served in cell 2 (remember that for cell 1 the capacity is constant since all the 50 resources are always allocated). For the sharing case ($\alpha > 0$), the total capacity still increases with the number of UEs for the same reason. Moreover, there is additional increase, due to the fact that BS1 is entitled to use a higher portion of the spectrum and thus can serve more UEs. It should be noted that after a certain number of UEs in cell 2 has been reached (in this case 25 since the number of initial sub-channels per cell is 50), there are no free resources that BS1 can exploit from cell 2. Any additional increment in the total sum capacity is only due to multiuser diversity, which still yields some benefits from the increase in the number of users, even though the additional improvement decreases for networks with more users.

The capacity results and sharing gains depend on the radio propagation environment. The larger the frequency diversity, the higher the interest that an operator might have in accessing the spectrum of its competitors because it might contain sub-carriers of better quality for it. Also, the allocation policy adopted by the BS plays an important role. In this numerical study, we are not considering fairness among the flows and for each sub-channel the best UE is always chosen. By introducing some fairness, the curves would change significantly, in particular the capacity values will decrease.

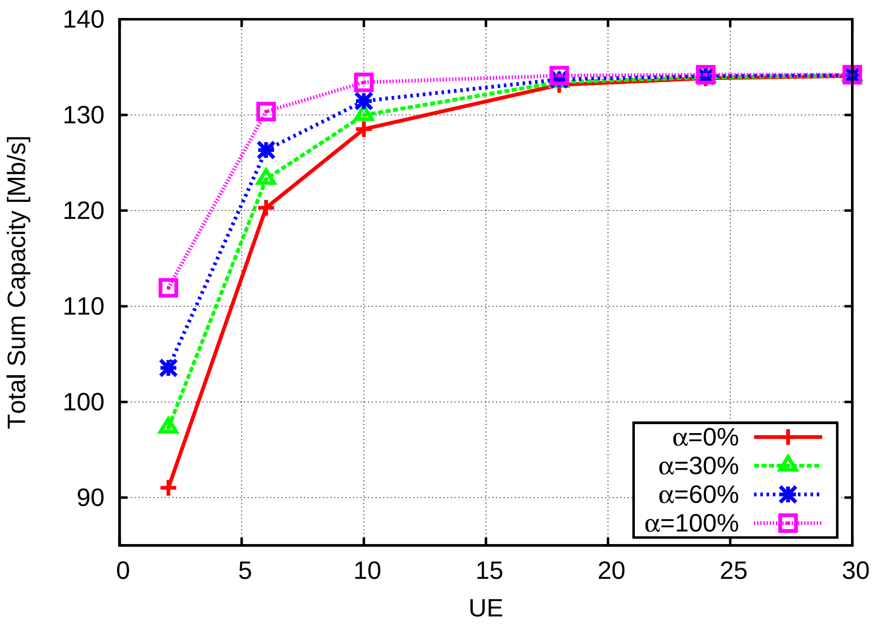

Figure 4.3: Total sum capacity versus the number of UEs per cell for various sharing percentages and symmetric load

4.2 Relaying in Wireless Networks in a Game Theoretic Perspective

Relay networks have been widely studied in information theory [101]. In particular, the relay channel represents one of the most common scenarios studied. Several theoretical results about the capacity of this basic network have been available in the literature since long [102], and others keep being proposed even very recently [103, 104].

At the same time, game theory [105] is being employed more and more every day by wireless telecommunication engineer. From a game theoretic perspective, the relay channel is a natural scenario to evaluate *cooperation* [106], which may improve the communication for users experiencing bad quality on their direct link to the destination. The concept of cooperation has a precise meaning in game theory mostly through the application of *coalitional games*, which we will exploit in this part of our work.

Yet, we do not apply these game theoretic concept to a pure information theory scenario. Instead, we focus on a precise application of the relay channel, with a real network protocol involving packet exchanges and retransmission, through cooperative Automatic Repeat reQuest (ARQ). Our proposal involves a cross-layer solution spanning on the data link (and more specifically, both channel access protocol design and ARQ) and network layers.

The contribution in this sense is two-fold. First, we give an analytical characterization of the performance of the relay channel from a game theoretic perspective. Second, we are also able to define a cooperative protocol where the overall network throughput

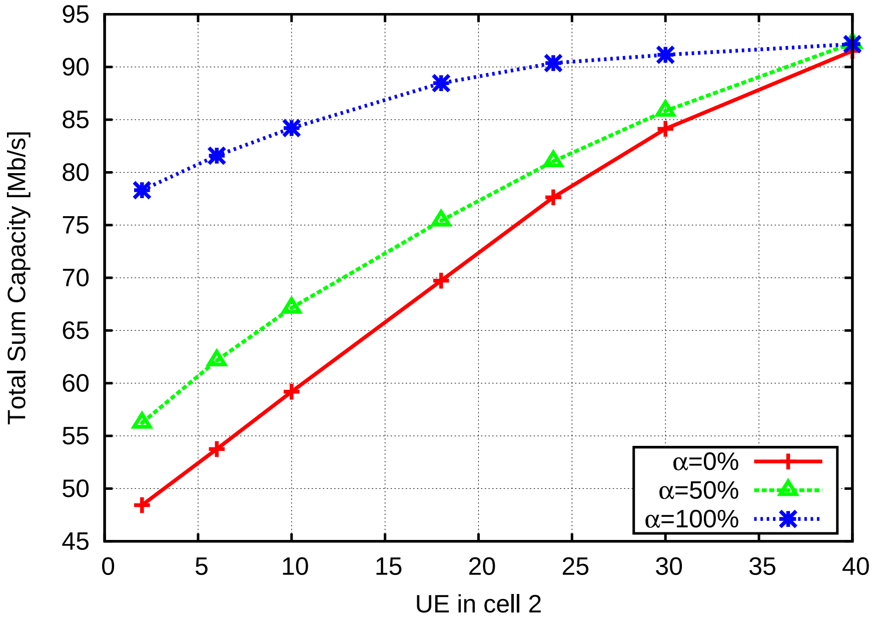

Figure 4.4: Total sum capacity versus the number of UEs in cell 2 for various sharing percentages and asymmetric load

is improved. However, differently from classic results of information theory where the capacity enhancement simply stems from spontaneous cooperation, i.e. the nodes collaborate out of goodwill, we precisely model also the node behavior so that their cooperation is not taken for granted, but rather promoted through a careful design of incentives to all the involved nodes (both those exploiting cooperative relays, and those who aid others, e.g. by forwarding their packets).

In this sense, this throughput enhancement is achievable in practical cases, and is truly beneficial for those nodes who have poor channel conditions, e.g. those at the cell edge, not only because somebody improves their throughput, but also since other nodes are actually willing to cooperate with them, since they see a concrete benefit in it.

4.2.1 Coalitional Game Terminology and Notation

Cooperative game theory [106] is a branch of game theory that provides analytical tools to study the behavior of rational players when they cooperate.

The main area of cooperative games is represented by coalitional games [107], defined as a pair $(\mathcal{N}, v)$, where $\mathcal{N} = \{1, ..., N\}$ is a discrete set of players and v is a function that quantifies the *value* of a coalition in a game. Each coalition $S \subseteq \mathcal{N}$ behaves as a single player, competing against other coalitions in order to obtain a higher value of v. A coalitional game may have the following properties:

Property 1. *(Characteristic form)* The value of a coalition S depends only on who are the members of that coalition, regardless of other coalitions

Property 2. *(Transferable utility)* The value of a coalition is a real number, representing the total utility achieved by the coalition, and it can be arbitrarily divided among its members

For coalitional games satisfying properties 1 and 2, the value $v : 2^{\mathcal{N}} \rightarrow \mathbb{R}$ is a function that assigns to each coalition S the total utility achieved by it. The utility value can be arbitrarily divided among the coalition members and the amount of utility that a player $i \in S$ receives, x_i, is the player's payoff. A payoff allocation is a vector **x** $\in \mathbb{R}^{|S|}$ (where $|S|$ is the cardinality of the set S) whose elements are the payoffs of players belonging to the coalition; in other words, it represents a redistribution of the total utility.

Another interesting property that a coalitional game may have is super-additivity, that for a game with properties 1 and 2 assumes the following form:

Property 3. *(Super-additivity)*
$v(S_1 \cup S_2) \geq v(S_1) + v(S_2) \;\; \forall S_1, S_2 \subset \mathcal{N}$ s.t. $S_1 \cap S_2 = \emptyset$

The super-additivity property expresses in mathematical terms that formation of a larger coalition is always beneficial. Hence, for those games where it holds, the players are encouraged to stick together, forming the grand coalition $\mathcal{N}$.

For a game having all properties listed before, the main aspects to analyze are:

- finding a redistribution of the total utility $v(\mathcal{N})$ such that the grand coalition is stable, i.e. no group of players has an incentive to leave the grand coalition
- finding fairness criteria for the redistribution of the total utility
- quantifying the gain that the grand coalition can obtain with respect to non cooperative behaviors

A payoff allocation is *group rational* if $\sum_{i=1}^{N} x_i = v(\mathcal{N})$ and it is *individually rational* if $x_i \geq v(\{i\}) \; \forall i$, i.e. if every player does not obtain a lower utility by cooperating than by acting alone. A payoff allocation having both properties is said to be an *imputation*.

The concept of *core* is also very important. It is defined as the set of imputations that guarantee that the grand coalition is stable, i.e. all payoff allocations where no group of players $S \subset N$ have an incentive to refuse the proposed payoff allocation, leaving the grand coalition and forming coalition S instead. Mathematically speaking,

$$\mathcal{C}=\left\{\mathbf{x} \text{ s.t. } \sum_{i=1}^{N} x_i = v(\mathcal{N}) \; , \; \sum_{i \in S} x_i \geq v(S) \; \forall S \subset \mathcal{N}\right\} \tag{4.5}$$

Indeed, the core may be empty, in which case the grand coalition is not stable. The existence of the core ought to be checked case by case, possibly exploiting some categories of games where the existence is guaranteed [105, Ch. 13].

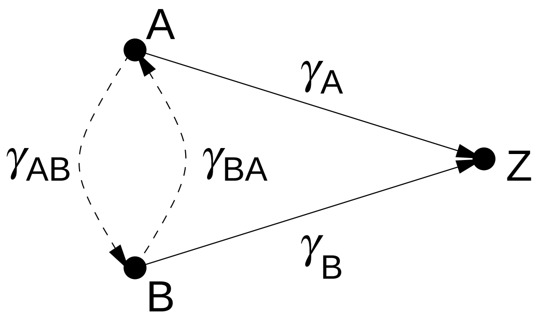

Figure 4.5: The reference scenario.

4.2.2 Problem Statement

We consider the scenario of two nodes, A and B, which want to communicate with an Access Point, Z, as represented in Fig. 4.5. γ_A, γ_B, γ_{AB} and γ_{BA} are the signal to noise ratios (SNRs) between A and Z, B and Z, A and B and B and A respectively. We suppose that:

- γ_A, γ_B, γ_{AB} and γ_{BA} are constant over time, i.e. time invariant channels and fixed transmission powers of A and B. Actually, also slow time-varying channels can be included in this analysis. Moreover, without losing generality, we suppose $\gamma_B > \gamma_A$
- node A and B always have packets to transmit to Z
- time division multiple access (TDMA) is adopted, assuming that the access point Z manages it in a centralized manner. The process that assigns a slot to a new packet is independent identically distributed (i.i.d.) with P_A and $P_B = 1 - P_A$ the probabilities to assign the slot to node A or node B respectively
- we consider an ARQ retransmission scheme with at most 1 retransmission (maximum total number of transmission $F = 2$). In subsection 4.2.2 we will see how the analysis can be generalized for multiple retransmissions
- we focus on the uplink connection from the users to the access point, therefore we neglect the traffic from Z to nodes A and B

Once a new packet for node A is scheduled, the non cooperative transmission process of this packet can be represented by the Markov Chain in Fig. 4.6. Absorbing states R_A and N_A represent the events that the packet is received or not received by Z. Other states represent the actual number of packet transmissions performed by user A, so the initial state is state 1_A. We define $q(\gamma)$ as the probability that a packet is correctly received when the SNR is γ. This function depends on the modulation scheme used and on the packet length. We define $P_{R_A}^{NC}$ as the probability to be absorbed in state R_A in the non cooperative case.

$$P_{R_A}^{NC} = q(\gamma_A) + (1 - q(\gamma_A))q(\gamma_A) \tag{4.6}$$

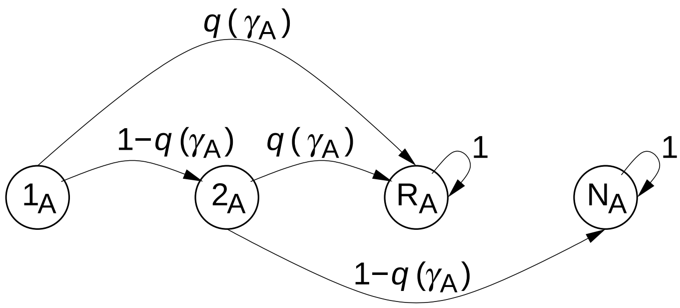

Figure 4.6: Non-cooperative transmission process of a packet of user A

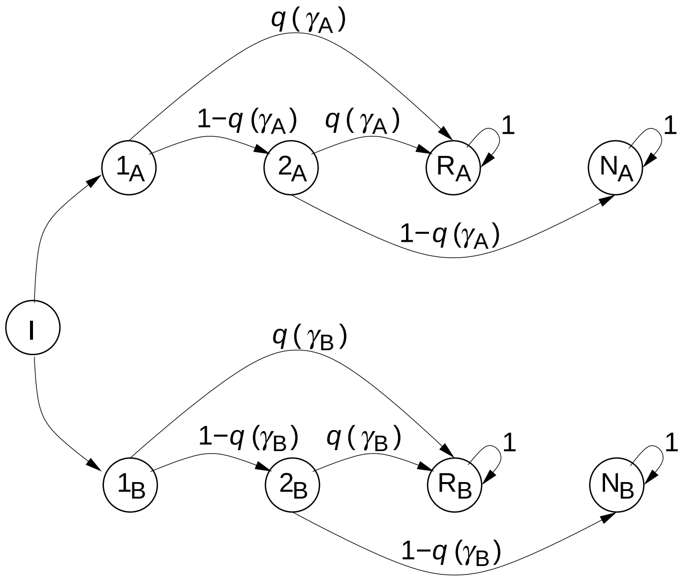

Figure 4.7: Non-cooperative transmission process of a packet in the network

We define N_A^{NC} as the average number of transmissions of the packet in the non cooperative case.

$$N_A^{NC} = q(\gamma_A) + 2(1 - q(\gamma_A)) = 2 - q(\gamma_A) \tag{4.7}$$

The transmission of a packet in the non cooperative case, from the choice of the user to packet reception (or to the maximum number of transmissions), is represented in Fig. 4.7. Initial state I represents the selection of the user that can transmit the packet. Users A and B are selected with probabilities P_A and P_B, respectively. Once either user is selected, the transmission process evolves analogously to the Markov Chain shown in Fig. 4.6. When the packet is correctly received by Z or the maximum number of transmissions is reached (i.e. an absorbing state is entered), another new packet is considered, again starting from state I: a new user is selected, a new packet is transmitted, and so on. Renewal theory [108] allows to study this kind of situations. The beginning of each renewal cycle constitutes a regenerative epoch of the Markov process. The asymptotic metrics of the network can be obtained by studying the average behavior of the Markov process. The asymptotic bit rate of each user is calculated by considering the average number of transmitted bits and dividing it by the average time to absorption:

$$BR_A^{NC} = \frac{P_A P_{R_A}^{NC}}{N^{NC}} \frac{N_{bit}}{T_{pkt}} \tag{4.8}$$

where $N^{NC} = P_A N_A^{NC} + P_N N_B^{NC}$ is the average number of transmissions for packet, N_{bit} is the number of bits in a packet and T_{pkt} is the time needed for a single packet transmission.

Finally, the asymptotic bit rate of the network for the non cooperative scenario is given by:

$$BR^{NC} = BR_A^{NC} + BR_B^{NC} \tag{4.9}$$

Cooperative ARQ

Now the performance of the network is evaluated for the case where cooperation is active, by means of the coalitional game framework. Nodes can cooperate, helping other nodes to retransmit a packet not correctly received by the access point.

We assume that the game satisfies properties 1 and 2. Note that in the two-user case, the former property is automatically satisfied. However, the property still holds true even if the analysis is extended to a network with more than two users, since the TDMA approach guarantees that different coalitions do not interact: each coalition tries to obtain the maximum throughput by using the slots assigned exclusively to it. For what concerns property 2, the problem of the throughput redistribution is addressed in subsection 4.2.2.

The value $v(\cdot)$ of the coalitional game is the throughput obtained by each coalition. In a two-user case, three coalitions are possible: the two coalitions formed by the single users, A and B, and the coalition formed by both users, i.e. the grand coalition $\mathcal{N} = \{A, B\}$. The value of each coalition is:

$$\begin{aligned} cv(\{A\}) = BR_A^{NC} \qquad & v(\{B\}) = BR_B^{NC} \\ v(\mathcal{N}) = BR^C = & BR_A^C + BR_B^C \end{aligned} \tag{4.10}$$

where BR_A^C and BR_B^C are the respective asymptotic bit rates for user A and B in the cooperative scenario.

During a cooperative transmission, the packet transmitted by a node is heard by Z and the other user who actively cooperates. In our formulation, cooperation implies that, if a packet is not correctly received by Z, its retransmission is carried out by the user who has the better signal to noise ratio, provided that it received the packet correctly. Thus, the transmission process for user A can be represented by the Markov Chain in Fig. 4.8. State 2_B (2_A) represents the second transmission in the case that user B has (not) correctly

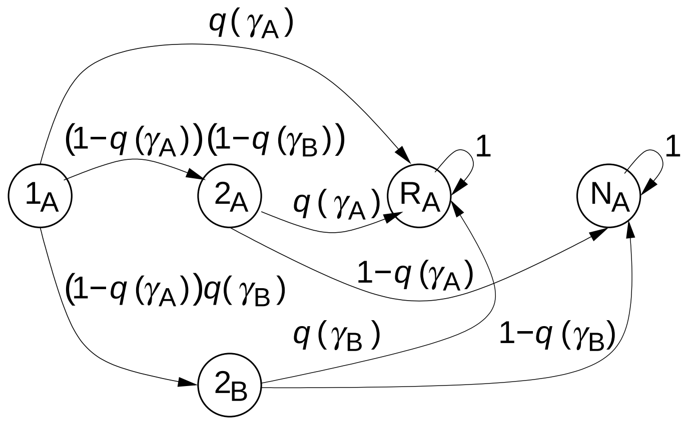

Figure 4.8: Cooperative transmission process of a packet of user A

received the packet during the first transmission; remember that we assume $\gamma_B > \gamma_A$. We obtain:

$$\begin{aligned} RCLP^C_{R_A} &= q(\gamma_A) + (1-q(\gamma_A))((1-q(\gamma_{AB}))q(\gamma_A) \\ &+ q(\gamma_{AB})q(\gamma_B)) \\ N^C_A &= q(\gamma_A) + 2(1-q(\gamma_A)) = 2 - q(\gamma_A) \end{aligned} \tag{4.11}$$

Note that $P^C_{R_A} > P^{NC}_{R_A}$ thanks to the cooperation of B, while $N^C_A = N^{NC}_A$ because we are considering at most 2 transmissions. If more transmissions are considered, we obtain also $N^C_A < N^{NC}_A$. Note that $P^C_{R_B} = P^{NC}_{R_B}$ because B is not helped by anybody.

Along the same lines of (4.8), the asymptotic bit rate of user A in the cooperative scenario is given by:

$$BR^C_A = \frac{P_A P^C_{R_A}}{N^C} \frac{N_{bit}}{T_{pkt}} > BR^{NC}_A \tag{4.12}$$

Finally:

$$v(\mathcal{N}) = BR^C = BR^C_A + BR^C_B > BR^{NC}_A + BR^{NC}_B \tag{4.13}$$

Therefore the game satisfies also property 3.

Throughput subdivision

Now we want to find a payoff allocation that belongs to the core and is fair under certain parameters. Note that, for a super-additive two player game, the core is not empty and coincides with the set of imputations.

In the considered game, the set of imputations is given by:

$$\begin{aligned} x_A =& BR^{NC}_A + (1-w)(BR^C_A - BR^{NC}_A) \\ x_B =& BR^{NC}_B + w(BR^C_A - BR^{NC}_A) \end{aligned} \tag{4.14}$$

Here a *cooperation weight* denoted as w is introduced to determine the throughput share that each user gets. Note that w is introduced to give a proper incentive to both users to cooperate. In fact, only user A, whose channel quality to Z is worse, can directly benefit from being helped by user B's cooperative relaying. However, user B can get an incentive to cooperate if this results in a larger throughput share.

By varying the value of the cooperation weight w in the interval $[0, 1]$, we can obtain all the imputations. It is immediate to see that $x_A + x_B = v(\mathcal{N})\ \forall w$. Moreover, for $w = 0$ we obtain $x_B = BR_B^{NC} = v(\{B\})$ and increasing w we have $x_B > v(\{B\})$. For $w = 1$ we obtain $x_A = BR_A^{NC} = v(\{A\})$ and decreasing w we have $x_A > v(\{A\})$. On the other hand, if $w < 0$ then $x_B < v(\{B\})$ and if $w > 1$ then $x_A < v(\{A\})$. Thus, by setting the value of w we decide the right level of fairness of the subdivision. This takes into account that, in order to cooperate with A, user B has to consume more power, retransmitting packets instead of A, that can in turn save power. We can assign a cost to the power, depending on the application/scenario we are considering. If the cost of the power increases, we have to increase also the value of the cooperation weight w (i.e. to further increase the payoff of cooperative users) in order to keep the same level of fairness.

So far we have supposed that the total throughput can be divided by users rather arbitrarily. From a practical point of view, the only thing that can be controlled is the allocation policy, P_A and P_B. We suppose therefore that the allocation policy is changed from P_A and P_B to P_A^C and P_B^C in order to satisfy the subdivision proposed. Is the new allocation policy feasible? That is, is $P_A^C + P_B^C \leq 1$? It is easy to show that the new allocation policy is feasible. In fact, we have to increase the allocation probability of the cooperating user B and decreasing the allocation probability of A, while keeping constant the total bit rate $v(\mathcal{N})$. Since B has a better SNR, it results that the increase $P_B^C - P_B$ is greater that the decrease $P_A - P_A^C$ in order to keep the total bit rate constant. Therefore:

$$P_B^C - P_B < P_A - P_A^C \Rightarrow P_B^C + P_A^C < P_A + P_B = 1 \tag{4.15}$$

This means that the allocation is feasible and that there is a positive probability that some slots are not assigned to anybody, which would not be meaningful. Therefore, the quantity $P' = 1 - P_A^C - P_B^C$ can be divided between users, increasing for example both P_A^C and P_B^C by the same amount, or increasing them by a weighted amount of P', where we can use again the cooperation weight w. Finally, this means that both users have a further benefit in obtaining an even higher bit rate compared to the subdivision proposed.

Multiple retransmission generalization

Previously, we have found the mathematical expressions for the asymptotic bit rates in the non-cooperative and cooperative cases for $F = 2$. This can be generalized to $F > 2$

as follows. For the non-cooperative case we obtain:

$$
\begin{aligned}
P_{R_A}^{NC} =& q(\gamma_A) + \sum_{i=2}^{F} q(\gamma_A)(1-q(\gamma_A))^{i-1} \\
N_A^{NC} =& q(\gamma_A) + \sum_{i=2}^{F-1} i q(\gamma_A)(1-q(\gamma_A))^{i-1} + \\
& + F(1-q(\gamma_A))^{F-1}
\end{aligned}
\tag{4.16}
$$

For the cooperative case, let $T_{A,Z}$ and $T_{A,B}$ be the times required by Z and B, respectively, to correctly receive the packet from A. Then, we have

$$
\begin{aligned}
rClP_{R_A}^{C} &= q(\gamma_A) + \sum_{i=2}^{F} P(T_{A,Z} = i) \\
&= q(\gamma_A) + \sum_{i=2}^{F} \Bigg[\sum_{k=1}^{i-1} P(T_{A,B}{=}k) P(T_{A,Z}{=}i | T_{A,B}{=}k) \\
&+ P(T_{A,B} > i-1) P(T_{A,Z} = i | T_{A,B} > i-1) \Bigg] \\
&= q(\gamma_A) + \sum_{i=2}^{F} \Bigg[\sum_{k=1}^{i-1} q(\gamma_{AB})(1-q(\gamma_{AB}))^{k-1} q(\gamma_B) \\
&\cdot (1-q(\gamma_A))^{k} (1-q(\gamma_B))^{i-k-1} \\
&+ (1-q(\gamma_{AB}))^{i-1} q(\gamma_A)(1-q(\gamma_A))^{i-1} \Bigg]
\end{aligned}
\tag{4.17}
$$

$$\begin{aligned}
rClN_A^C &= q(\gamma_A) + \sum_{i=2}^{F-1} iP(T_{A,Z} = i) + FP(T_{A,Z} > F-1) \\
&= q(\gamma_A) + \sum_{i=2}^{F-1} i\Bigg[\sum_{k=1}^{i-1} P(T_{A,B} = k) \\
&\cdot \quad P(T_{A,Z} = i|T_{A,B} = k) + P(T_{A,B} > i-1) \\
&\cdot \quad P(T_{A,Z} = i|T_{A,B} > i-1)\Bigg] + F\Bigg[\sum_{k=1}^{F-2} P(T_{A,B} = k) \\
&\cdot \quad P(T_{A,Z} > F-1|T_{A,B} = k) + P(T_{A,B} > F-2) \\
&\cdot \quad P(T_{A,Z} > F-1|T_{A,B} > F-2)\Bigg] \\
&= q(\gamma_A) + \sum_{i=2}^{F-1} i\Bigg[\sum_{k=1}^{i-1} q(\gamma_{AB})(1-q(\gamma_{AB}))^{k-1} q(\gamma_B) \\
&\cdot \quad (1-q(\gamma_A))^k (1-q(\gamma_B))^{i-k-1} + (1-q(\gamma_{AB}))^{i-1} \\
&\cdot \quad q(\gamma_A)(1-q(\gamma_A))^{i-1}\Bigg] + F\Bigg[\sum_{k=1}^{F-2} q(\gamma_{AB}) \\
&\cdot \quad (1-q(\gamma_{AB}))^{k-1}(1-q(\gamma_A))^k (1-q(\gamma_B))^{F-k-1} \\
&+ \quad (1-q(\gamma_{AB}))^{F-2}(1-q(\gamma_A))^{F-1}\Bigg]
\end{aligned} \tag{4.18}$$

It is easy to see that $P_{R_A}^C > P_{R_A}^{NC}$ and $N_A^C \leq N_A^{NC}$, therefore $BR_A^C > BR_A^{NC}$. Thus, the reasonings done in subsection 4.2.2 are still valid even for $F > 2$.

4.2.3 Results

The throughput of the proposed scenario with a time invariant channel has been evaluated through a Matlab simulator. An example of results is shown in Fig. 4.9, where the throughput for both nodes in the cooperative and non-cooperative scenario is plotted. We consider a $16-QAM$ modulation with packet length of 12000 bits. P_A and P_B are equal to 0.5, γ_{AB} and γ_{BA} are both set to 110. The SNR of B, γ_B, is constant and equal to 70, while γ_A varies from 0 to $2\gamma_B = 140$. For very low values of γ_A, in the non-cooperative case, the capacity of node A is close to 0, while that of B is around 2700 bits/s. In this case, most of the slots assigned to A are wasted to retransmit the same packet. When B cooperates, it allows to save most of this wasted slots without decreasing the throughput of A. This saved slots are then assigned to B which doubles its throughput with respect to the non-cooperative case. Increasing γ_A reflects in an enhanced non-cooperative throughput of A. Cooperation from B allows to save fewer slots than the previous case, but it is

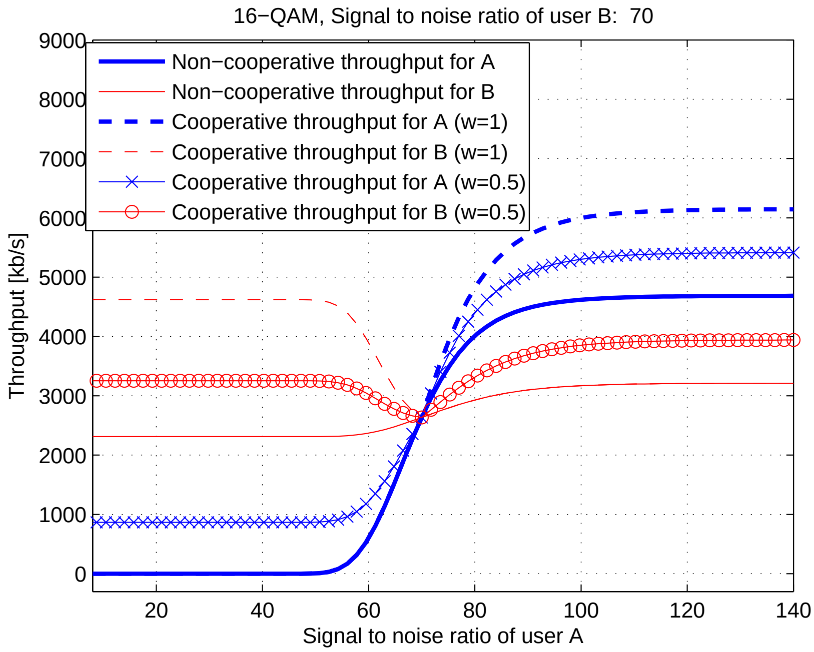

Figure 4.9: Throughput for the time invariant channel case

still beneficial and B experiences a higher throughput compared to the non-cooperative case. The closer γ_A to γ_B, the lower the gain of B. When the SNRs of both nodes are equal, cooperation is useless and the nodes have the same throughput. From this point on the situation is reversed: when $\gamma_A > \gamma_B$, A helps B and the cooperative throughput of node A is higher than its non-cooperative throughput. The larger γ_A, the higher the (absolute) throughput gain.

It is worth noting that, while not rewarding cooperation, i.e. $w = 0$, determines low throughput for the node with worse channel quality, cooperating with weight equal to 1 does not represent an improvement for this user either. In such a case, the whole throughput gain harvested from the cooperation is assigned to the better user to encourage its cooperation. Thus, the case where $w = 0.5$ represents an interesting trade-off, where both users get a cooperation gain, and have the proper incentives to cooperate from a game theoretic standpoint.

4.2.4 Discussion and Possible Extensions

The numerical results show an actual gain for both nodes when they decide to cooperate. Many papers which discuss cooperation in wireless networks often fail to determine the actual benefit for the nodes in cooperating; therefore, they often resort to some form of side payment. However, the viability of such solutions, in both economic and legal terms, is arguable at best. Instead, our proposed approach simply addresses the benefit of cooperation as a throughput enhancement for *both* kinds of users, i.e. those relaying packets and those exploiting a relay. Also, focusing on a time-division multiplexing channel access is not restrictive; this assumption may easily be translated to other forms, e.g. random-based, medium access control.

Possible extensions of the present work, which are currently ongoing, involve the introduction of hybrid ARQ [109], as opposed to plain ARQ as employed here, and the addition of time-varying channels in the analysis. This latter point may require to extend the problem so as to include dynamic games and further negotiation among the cooperating users. In a time invariant channel, cooperation is always triggered in the same manner, i.e. the user with better channel relays the packets for the other; conversely, the main challenge of time varying gains would be that the users can sometimes have reversed cooperation roles. This complicates the set of strategies that can be played by each player; especially, the access point needs to give proper cooperation incentives so as to avoid the "ungrateful" situation (a user that most of the time enjoys cooperation by a relay does not reciprocate when it is its turn to relay a packet). Nevertheless, our preliminary evaluations hinted that a cooperation gain still holds also in this scenario.

Finally, given the promising results found for a simple two-node network, it is surely worth investigating an extension to larger networks, possibly with multi-hop relaying. This development, currently under evaluation, implies both an evaluation on a larger scale and also the definition of a proper negotiation protocol to establish the cooperation roles [110].

4.3 Cooperation in Relay Sharing Scenarios

Cooperation is one of the most promising enabling technologies to meet the increasing rate demands and quality of service requirements in wireless networks. For the case of infrastructure sharing, when a network operator decides to cooperate, it shares some or all of its nodes, that become relays for another network. In such a scenario, cooperation can leverage the benefits of diversity, obtaining a considerable gain in the efficiency of shared resources. Usually, sharing the whole set of nodes can grant the biggest advantage to both networks. However, this clearly comes at the cost of additional traffic that should be handled by some of the shared nodes. In addition, in a realistic environment, an operator may not be willing to share too many nodes to improve the traffic of another operator, e.g. for security or privacy reasons. Therefore, the operator may decide to share only a limited number of nodes, receiving the same treatment from the operator of the other network. If this is the case, an optimal choice of the shared nodes, according to certain criteria, is needed. Indeed, some nodes may be deployed in crucial positions, and they may be particularly suited for helping the other network; on the contrary, nodes placed close to the network border are likely to be less useful or even useless. Furthermore, sharing a node implies that a higher amount of traffic will be routed through it, which may result in a higher latency for the traffic of its own network.

We consider two wireless networks deployed in the same region but operated by different entities. In the first scenario, the two coexisting networks perform their operations separately: each network only uses its own resources to deliver the data packets generated by its nodes. Clearly, since they are assumed to share the same spectrum resources, cross–network interference may limit the overall performance. For such a scenario, we select a set of local parameters: some of them are directly observable and depend only on the topology of the network, like the number of neighbors at a given node, and some others are not observable and depend on the link characteristics and on the traffic load. We exploit Bayesian Network (BN) [111] analysis to estimate the joint probability distribution of this set of parameters, and we use BN also to predict, given the evaluation of the observable parameters, the values of the other parameters that will be used to calculate a performance metric. The use of this probabilistic tool is very promising for wireless network optimisation, and it has been recently exploited, e.g. for predicting the occurrence of congestion in a multi-hop wireless network [112].

Such an approach can also be used to improve the performance of both networks in our scenario. The key idea is that a higher node density may help both networks to augment the available diversity, and thus to find shorter routes for multi-hop communications. It is straightforward that this may be obtained if each network can take advantage of some of the nodes of the other one. We model the interaction between the two networks through Game Theory. In spite of the considerable theoretical gain that a cooperative transmission allows, modeling the involved agents as smart selfish decision-makers usually leads to inefficient non-cooperative equilibria. For example, [113] shows that the IEEE 802.11 distributed Medium Access Control (MAC) protocol leads to an inefficient Nash Equilibrium (NE) and in [114] a situation similar to the Prisoner's Dilemma occurs in a slotted Aloha MAC protocol. To improve the performance of the network, cooperation among the players is often desirable. In our work, we achieve this goal by formulating the problem as a repeated game, which consists in a number of repetitions of a base game. It captures the idea that a player has to take into account the consequences of his current action on the future actions of other players. Cooperation is in fact obtained by punishing deviating users in subsequent stages. Similar approaches have been used for example in [115–117].

In brief, the main contributions of this part of our study are:

- the definition of the cooperation problem of two networks sharing the same spectrum resources as a strategic game;
- the use of BN theory to learn the probabilistic relationships among a set of parameters of interest in the network, in order to infer the performance metric parameters from some observable topological parameters;
- the proposal of a game theoretic algorithm to choose the best nodes to share;
- the implementation of the strategic game and the BN predictor in an actual wireless network simulator, that simulates the network behavior at the Physical, the MAC and the network layers of the protocol stack;
- to numerically show the effectiveness of our algorithm in improving the perfor-

mance of the wireless networks by accurately selecting the nodes to be shared.

4.3.1 Mathematical Preliminaries

In this section we briefly describe the mathematical tool, i.e. Bayesian Networks, that we adopt to identify techniques for the selection of the best cooperating nodes in the network. It is a method to learn an approximate joint probability distribution among a set of random variables from a set of instances of such variables.

A BN is a probabilistic graphical model [111] describing conditional independence relations among a set of M random variables through a Directed Acyclic Graph (DAG). This graph is used to efficiently compute the marginal and conditional probabilities of the M variables. A node i in the DAG represents a random variable x_i, while an arrow that connects two nodes i and j represents a direct probabilistic relation between the corresponding variables x_i and x_j. The absence of a direct arrow between two variables implies that the variables are independent, given certain conditions on the other variables of the graph. The orientation of the arrow is also relevant to describe the relationship between the two variables. If the arrow is directed from node i to node j, we say that i is a parent of j, and we write $x_i \in \text{pa}(x_j)$. To clarify this concept, consider the following example. If nodes h, i, and j are represented in a BN such that h is a parent of i and i is a parent of j, the joint probability of the corresponding variables is

$$P(x_h, x_i, x_j) = P(x_h)P(x_i|x_h)P(x_j|x_i) \ , \tag{4.19}$$

that is simpler than a general joint probability among three variables. See [111] for further details on the BN properties.

Learning the structure The technique to learn the approximate joint probability distribution through a BN is divided into two phases, structure learning and parameter learning. The former is a procedure to define the DAG that represents the qualitative relationships between the random variables, i.e. the presence of a direct connection between a couple of variables, not conditioned by other variables. According to the score based method, e.g. see [118], we do not assume any a priori knowledge on the data, but we just analyze the realizations of the variables and we score each possible DAG with the Bayesian Information Criterion (BIC) [119], that we have chosen as a score function. The BIC is easy to compute and is based on the Maximum Likelihood (ML) criterion, i.e. how well the data suits a given structure, and penalizes DAGs with a higher number of edges. If each variable is distributed according to a multinomial distribution, i.e. it has a finite number of possible outcomes, then the BIC becomes very simple to compute, involving only summations for all possible outcomes of the variables and all possible outcomes of the parents of each variable, see [118].

Learning the parameters The parameter learning phase consists in estimating the parameters of the simplified joint distribution according to the probability structure defined by the DAG chosen in the structure learning phase. In order to have the joint distribution, it suffices to estimate the probability of each variable conditioned by the variables that correspond to its parent nodes in the graph. Coherently with the choice of the BIC as a scoring function, we use the ML estimation technique also to determine all the conditioned probabilities for each variable considered.

4.3.2 System Model

In this section, we describe the network scenario under investigation from the physical up to the routing layer. In our scenario, two ad hoc wireless networks coexist and share the common spectrum resource. Each network consists of N terminals randomly deployed, and each node is a source of traffic, which generates packets according to a Poisson process with intensity λ. The end destination of each packet is another node in the network, chosen at random. Furthermore, time is divided in slots and slot synchronization is assumed across the whole network.

Physical Layer

At the physical layer, Code Division Multiple Access (CDMA) with fixed spreading factor is employed to separate simultaneous transmissions, since both networks share the same spectrum resources, and a training sequence is transmitted at the beginning of each transmission to help channel estimation. The receiving node, $D^{(0)}$, uses a simple iterative interference cancellation scheme to retrieve the desired packet when M simultaneous communications, namely $T^{(1)}, \ldots, T^{(M)}$, are heard. In order to describe this scheme, we need to define the SINR at $D^{(0)}$ for the incoming transmission $T^{(i)}$ from node $D^{(i)}$, i.e.

$$\Gamma^{(i)} = \frac{N_s P^{(i)}}{N_0 + \sum_{j \neq i} P^{(j)}} \ , \tag{4.20}$$

where N_0 is the noise power and N_s is the spreading factor. $P^{(j)}$ indicates the incoming power due to $T^{(j)}$, i.e. for all $j = 1, \ldots, M$:

$$P^{(j)} = \frac{P_T |h_{D^{(j)},D^{(0)}}|^2 d_j^{-\alpha}}{A} \ , \tag{4.21}$$

where P_T is the transmission power, which is considered to be the same for all the nodes in the network, A is a fixed path-loss term, d_j is the distance between the receiving node and the source of $T^{(j)}$, α is the path loss exponent, and $h_{D^{(j)},D^{(0)}}$ is a complex zero mean and unit variance Gaussian random variable, which represents the effect of multi-path fading. More precisely, in our scenario, we consider a time correlated block fading. Therefore, for the channel between nodes $D^{(j)}$ and $D^{(0)}$, the multi-path fading coefficient in time slot t is

$$h_{D^{(j)},D^{(0)}}(t) = \rho \ \ h_{D^{(j)},D^{(0)}}(t-1) + \sqrt{1-\rho^2} \ \xi \ , \tag{4.22}$$

where ρ is the time-correlation factor and ξ is an independent complex Gaussian random variable with zero mean and unit variance. Now we can describe the iterative interference cancellation scheme as follows:

- the destination node $D^{(0)}$sorts the M incoming transmissions according to the received SINR, in decreasing order (for simplicity, we assume $\Gamma^{(1)} \geq \cdots \geq \Gamma^{(M)}$);
- starting from transmission $T^{(1)}$, $D^{(0)}$ tries to decode the corresponding packet, with a decoding probability that is a function of $\Gamma^{(1)}$;
- if the packet is correctly received, its contribution is subtracted from the total incoming signal;
- $D^{(0)}$ attempts to decode the transmission with the next highest SINR, $T^{(2)}$, and goes on until it can try to decode the packet of interest.

MAC Layer

At the MAC layer, we implement a simple transmission protocol based on a Request-To-Send/Clear-To-Send (RTS/CTS) handshake. Every time node $D^{(i)}$ wants to send a packet to node $D^{(j)}$, it checks the destination availability by sending a RTS packet; if $D^{(j)}$ is not busy, it replies with a CTS so that $D^{(i)}$ can start transmitting the packet. Correct reception is acknowledged by means of an ACK packet. In the case of decoding failure, after a random backoff time, node $D^{(i)}$ schedules a new transmission attempt, or discards the packet, if the maximum number of retransmissions has been reached. The signaling packets are very short, i.e. they are transmitted within a single time slot, and are protected by a simple repetition code of rate $1/2$. Instead, data packets may span several time slots, so error detection coding is used to verify their correct reception, i.e. redundancy bits are added at the end of each packet.

Network Layer

The source node and the destination node are not necessarily within coverage range of each other, so we consider multi-hop transmissions. Two nodes are neighbors, i.e. they can communicate directly, if their distance is lower than a threshold value ℓ. In order to transmit to a node that is not within coverage, the nodes use a static routing table, which is built using Optimized Link State Routing (OLSR) [120], a traditional routing protocol, and is available at every node of the network. Each time a node generates a new packet, or receives a packet to be forwarded, the packet is put in the node queue, with First-In First-Out (FIFO) policy. The maximum queue length is fixed and equal for all nodes. If a new packet arrives when the queue is full, it is simply discarded.

4.3.3 Cooperation Strategy

In this section we describe how the two networks that coexist in our scenario can share efficiently the spectrum resources by means of cooperation.

Performance metric

Given the path from $D^{(i)}$ to $D^{(j)}$, we define the delivery delay $\zeta^{(i,j)}$ as the average end-to-end delay of a packet sent along the path, given that the packet is received; and the packet loss probability $p_c^{(i,j)}$ as the probability that a packet is lost along the path. The former depends on the channel and interference conditions, which may require one or more retransmissions, and on the overall traffic level. Indeed, for multi-hop routes, a packet has to wait at each relay node until all the packets it finds in the FIFO queue have been sent. Regarding the latter, the packet loss, there are two main events to be accounted for. One is a high interference level, that may lead to a packet drop due to an excessive number of retransmissions; the other is buffer overflow, i.e. the packet is discarded if the next relay has no room for it in its queue.

We consider a metric to measure the gain offered by the various cooperation strategies, which takes into account the average end-to-end delay of a packet sent along the path from $D^{(i)}$ to $D^{(j)}$. Since no end-to-end packet retransmission mechanism is implemented in our network, the effect of lost packets must also be considered. Ignoring lost packets (i.e. computing the delay statistics only on correctly delivered packets) may lead to an optimistic evaluation of the network performance under heavy traffic, where few packets actually reach the destination. In this case, a high-loss path might end up being considered better than a more reliable path with a slightly higher delivery delay. The other extreme, i.e. defining the delay contribution of a lost packet as infinite, makes the delay evaluation meaningless since the average delay would also be infinite for any positive loss probability. Clearly, neither option is desirable in our case.

Therefore, we propose another definition that gives a finite bias to the average delay in case of a packet loss. In particular, when a packet is lost when going from $D^{(i)}$ to $D^{(j)}$, we increase the delay of the following packet in the same path by the interarrival time between packets routed on that path. [1] This additional delay is given by $(N-1)/\lambda$, i.e. the inverse of the per-path average traffic intensity (recall that each packet generated at $D^{(i)}$ has a randomly chosen destination among the remaining nodes of the network, so that the per-node traffic λ needs to be divided by the number of possible destinations, $N-1$).

According to this reasoning, we recursively define the *weighted delivery delay* of a data

[1] Equivalently, we assign to lost packets a delay contribution equal to the interarrival time, to received packets the actual delay incurred, and then divide the sum of all contributions by the number of correctly received packets only.

packet sent via multi-hop transmission by node $D^{(i)}$ to node $D^{(j)}$ as:

$$\mathcal{W}^{(i,j)} = \left(1 - p_c^{(i,j)}\right) \zeta^{(i,j)} + p_c^{(i,j)} \left(\frac{N-1}{\lambda} + \mathcal{W}^{(i,j)} \right) . \tag{4.23}$$

In this calculation, the channel and interference conditions, and thus the loss probability, are assumed to be independent for different packets. This is due to the fact that the destination for each packet is chosen at random, and the time between two subsequent packet transmissions over the same path is deemed to be long enough.

From Eq. (4.23) we obtain:

$$\mathcal{W}^{(i,j)} = \frac{N-1}{\lambda} \frac{p_c^{(i,j)}}{1 - p_c^{(i,j)}} + \zeta^{(i,j)} . \tag{4.24}$$

The delivery delay $\zeta^{(i,j)}$ and the loss probability $p_c^{(i,j)}$ depend on the nodes that the routing protocol selects as relays. In a static network, it is possible to estimate these values during a training period. Instead, if the network is dynamic (mobile nodes or time-varying traffic statistics), this is not possible. We propose a different way of estimating the delay and the loss probability, based only on instantaneous geographic and routing information. Since a packet sent over a multi-hop path has to traverse a number of nodes before reaching the destination, we make the assumption that both the overall path delivery delay and the overall path loss probability can be decomposed into contributions given by the various traversed nodes. More precisely, the overall delivery delay is given by the sum of the average delays required to traverse every single node (time in queue plus transmission time), whereas the overall loss probability is obtained from the loss probabilities at every node (probability of transmission failure and probability of buffer overflow). If $\mathcal{R}^{(i,j)}$ is the set of nodes belonging to the path between $D^{(i)}$ and $D^{(j)}$ (excluding $D^{(i)}$ and $D^{(j)}$), we have:

$$\zeta^{(i,j)} = \zeta_q^{(i)} + \sum_{h \in \mathcal{R}^{(i,j)}} \zeta_q^{(h)} , \tag{4.25}$$

where $\zeta_q^{(h)}$ is the average time between the arrival of a packet at node $D^{(h)}$ and its reception at the next hop. This delay depends on the next relay; indeed, while the time needed for traversing the queue is the same for all packets, the time required for a successful transmission depends on the channel condition, and hence on the next hop chosen. We estimate $\zeta_q^{(h)}$ averaging over all the possible next-hop relays, thus over all the neighbors of node $D^{(h)}$.

The packet loss in the multi-hop path is calculated in a similar way, i.e.

$$p_c^{(i,j)} = 1 - (1 - p_t^{(i)})(1 - p_q^{(j)}) \prod_{h \in \mathcal{R}^{(i,j)}} (1 - p_t^{(h)})(1 - p_q^{(h)}) , \tag{4.26}$$

where $p_t^{(h)}$ is the probability that a transmission from node h to the next hop fails because the maximum number of retransmissions is reached, and $p_q^{(h)}$ is the probability that a packet correctly received at node $D^{(h)}$ is discarded due to buffer overflow. Furthermore, we notice that $p_q^{(h)}$ depends on the queue of the receiving node $D^{(h)}$, while $p_t^{(h)}$ depends also on which node is used as next hop. For this reason, similarly to what we have done for $\zeta_q^{(h)}$, we consider a value averaged over all the neighbors of $D^{(h)}$.

With (4.25) and (4.26) we can calculate the weighted delivery delay $\mathcal{W}^{(i,j)}$, defined in (4.24). This parameter should be estimated for each couple of nodes, with a sufficiently long training period. From (4.24), we define $\overline{\mathcal{W}}$ as the average over all the couples of nodes belonging to the network. This will be used in the following as the performance metric of the whole network.

Stochastic estimation of local parameters

In a real network, the values of the parameters $\zeta_q^{(i)}$, $p_t^{(i)}$, and $p_q^{(i)}$ should be estimated based on local information. Our idea is to use some parameters that can be easily calculated at each node $D^{(i)}$. We consider in particular the number of flows $\mathcal{F}^{(i)}$, that can be easily calculated from the routing table, and the number of neighbors, $\mathcal{N}^{(i)}$. We have estimated the probabilistic relationships among ζ_q, p_t, p_q, $\mathcal{F}$, and $\mathcal{N}$. Notice that we removed the dependence on the specific node. In fact, the Bayesian Network approach exploits the collected data, which are specific for each node, to find out the correlation between the local parameters and the values of $\mathcal{N}$ and $\mathcal{F}$. The result is a set of general conditional distributions (one per each local parameter) which can be therefore applied to any node of the network. It follows that once the number of flows or neighbors of a given node is known, the distributions of $\zeta_q^{(i)}$, $p_t^{(i)}$, and $p_q^{(i)}$ for that node are also known.

We first collected the measures of these parameters in our scenario as a function of the traffic load λ, for different topologies. Then we calculated the structure of the BN. We should notice that this procedure is different from using a training period to directly derive the local parameters. In fact, in this case a training period would be needed every time the topology changes, so as to evaluate their value for each specific node or path. On the contrary, with our procedure we can estimate the general joint probability among these parameters, that does not depend on the specific topology.

The structure of the BN is reported in Fig. 4.10. The structure of this BN is the same for all the values of λ, while quantitatively the probabilistic relationships change with λ. We notice that $\mathcal{N}$ does not influence, to a first approximation, the values of the three performance parameters, once the value of $\mathcal{F}$ is observed. In other words, once we calculate from the routing table the value of $\mathcal{F}$, we can have an estimate of the probability distribution of the three performance parameters. From these estimated parameters, we can calculate also the overall network performance $\overline{\mathcal{W}}$.

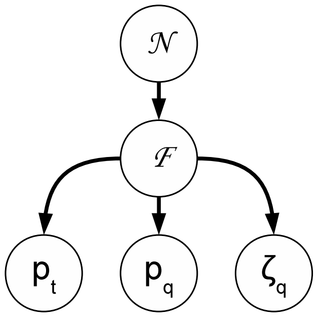

Figure 4.10: Bayesian Network showing the probabilistic relationships among the 5 parameters of interest: ζ_q, p_t, p_q $\mathcal{F}$, and $\mathcal{N}$.

Cooperation

When cooperation is exploited, some nodes are shared between the two networks, and the routing tables calculated via OLSR change accordingly. By using the framework introduced above, we can estimate the overall performance of the two networks with and without cooperation. We denote with $\overline{\mathcal{W}}_k(\mathcal{D}_1, \mathcal{D}_2)$ the weighted delivery delay of network k, with $k = 1, 2$, when the two networks share the set of nodes $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. In particular, $\overline{\mathcal{W}}_k(\emptyset, \emptyset)$ is the performance metric of network k when no nodes are shared. Thus, for any choice of the nodes shared we can calculate the variation in $\overline{\mathcal{W}}_k$ for the two networks. Then, we can model the cooperation strategy by means of Game Theory, by considering each network as a selfish agent whose utility function can be any decreasing function of $\overline{\mathcal{W}}_k$. To sum up, the following steps are followed in our framework:

- we learn the network behavior by measuring the parameters of interest over several random topologies with fixed setup;
- we use the BN method to infer the joint distribution among ζ_q, p_t, p_q, $\mathcal{F}$, and $\mathcal{N}$;
- we evaluate the utility functions $\overline{\mathcal{W}}_k(\mathcal{D}_1, \mathcal{D}_2)$, for the two networks $k \in \{1, 2\}$, for all the possible choices of the sets $\mathcal{D}_1$ and $\mathcal{D}_2$.
- we select the two subsets $\mathcal{D}_1$ and $\mathcal{D}_2$ to be shared, based on the game theoretic approach described in Section 4.3.3.

Game theoretic approach

The problem is formulated as a repeated 2-player game, where the players are the two networks. We name the nodes of the networks from 1 to $2N$, where the nodes in the sets $\mathcal{Q}_1 = \{1, ..., N\}$ and $\mathcal{Q}_2 = \{N+1, ..., 2N\}$ belong to network 1 and 2, respectively. The strategy of each network is represented by the set of nodes $\mathcal{D}_1$ and $\mathcal{D}_2$ they decide to share, therefore in the most general formulation the strategy sets are the power sets $\mathcal{S}_1 = 2^{\mathcal{Q}_1}$ and $\mathcal{S}_2 = 2^{\mathcal{Q}_2}$. The utility function of each network, $u_k : 2^{\mathcal{Q}_1} \times 2^{\mathcal{Q}_2} \to \mathbb{R},\ k =$

Table 4.2: Simulation parameters

Number of nodes per network	10
Transmission power [dBm]	24
Noise floor [dBm]	-103
Modulation used	BPSK
Time slot duration [ms]	1
Packet length [bit]	4096
λ [pkt/s/node]	1 to 5
Spreading factor	16
Fading correlation factor ρ	0.9

$1, 2$, is the reciprocal of the average weighted delay per path for that network, that is, $\overline{\mathcal{W}}_k^{-1}(\mathcal{D}_1, \mathcal{D}_2)$. Each of these metrics jointly depends on the strategies of both players: if a network decides to share a given node, that node is loaded by the traffic of the other network that passes through it. On the other hand, an additional shared node decreases the overall amount of traffic that passes through the other nodes.

We assume for simplicity that the networks do not have the freedom to choose the number of nodes to share. They can share either no nodes or exactly 2 nodes, therefore the cardinality of each strategy space is $\binom{N}{2} + 1$. Although our approach can be extended to a larger number of cooperating nodes, our numerical results show that a large fraction of the available cooperation gain is already achieved with this simple choice.

If we consider a single stage of this game it is immediate to see that the unique NE is the strategy profile $s = (\emptyset, \emptyset)$, i.e. no network cooperates. In fact, given the strategy of the other, each network prefers to share no nodes in order not to increase the total traffic through its nodes. However, in the repeated formulation it can be shown that each strategy profile that allows to reach a better utility for both players is a NE. A player deviating from that strategy profile can be punished by the other player during subsequent stages. The duration of this punishment can be set so that the gain obtained during the deviating stage does not compensate the loss during the subsequent stages. Punishment strategies in repeated games allow multiple equilibria with varying utilities for each player.

Inspired by the Nash bargaining solution [121], we decide to maximise the product

$$\left(u_1 - u_1^{NC}\right)\left(u_2 - u_2^{NC}\right), \tag{4.27}$$

where u_1^{NC} and u_2^{NC} are the status quo utilities, i.e. the utilities $\overline{\mathcal{W}}_1^{-1}(\emptyset, \emptyset)$ and $\overline{\mathcal{W}}_2^{-1}(\emptyset, \emptyset)$ obtained when networks do not cooperate. We additionally impose the mathematical constraint $u_k - u_k^{NC} \geq 0$, $k = 1, 2$, to avoid the situation where the maximum corresponds to a decrease in the utilities of both networks. The solution found results in increased utilities for both networks compared to the non cooperative case, therefore it is

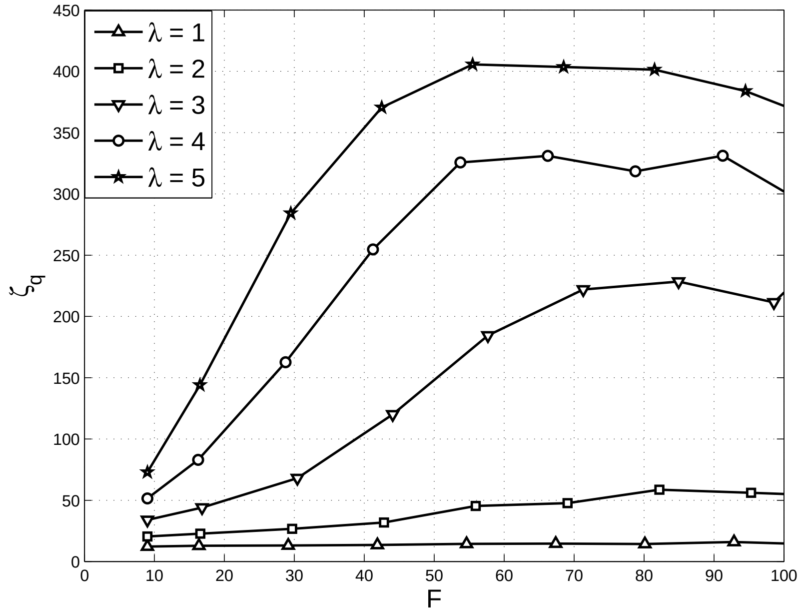

Figure 4.11: BN estimation of the average delivery delay ζ_q as a function of the number of flows $\mathcal{F}$ passing through the node.

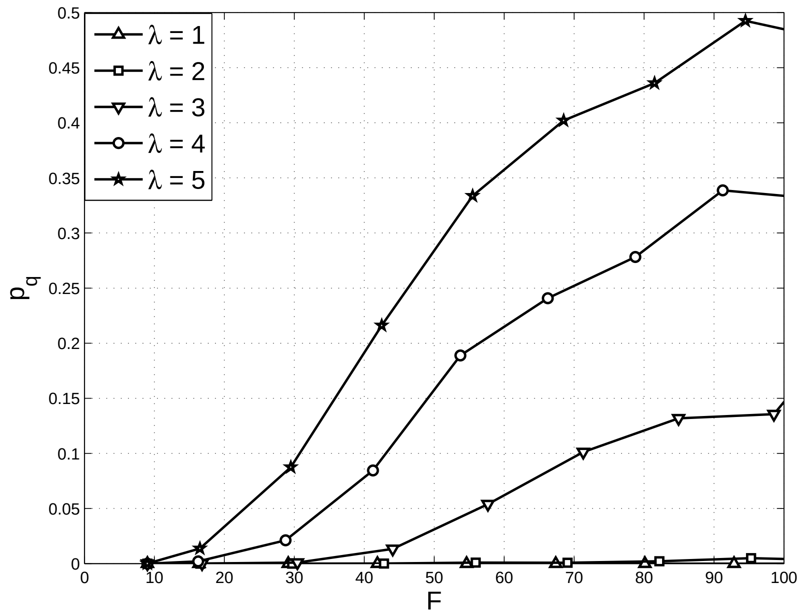

Figure 4.12: BN estimation of the probability of buffer overflow p_q as a function of the number of flows $\mathcal{F}$ passing through the node.

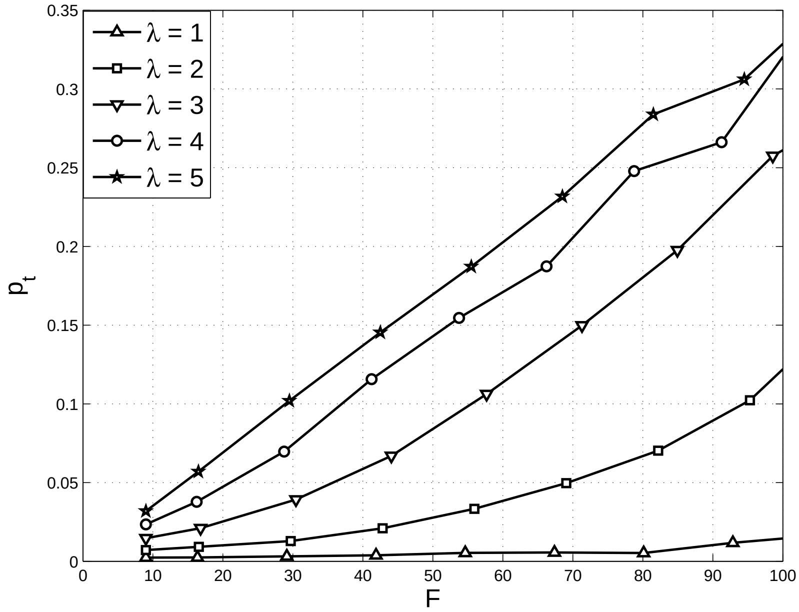

Figure 4.13: BN estimation of the probability of transmission failure p_t as a function of the number of flows $\mathcal{F}$ passing through the node.

a NE for the repeated game formulation.

4.3.4 Results

In this section we present the simulation setup and the main results of our approach for cooperation.

Simulation Setup

In order to prove the effectiveness of our cooperation strategy, we developed a network simulator which encompasses the layers from physical to routing, as described in Section 4.3.2. The system parameters are reported in Table 4.2. Each simulation run is performed with randomly generated connected networks, and lasts for 10000 time slots, including an initial transient phase. Different values of the traffic generation intensity λ were considered, from 1 packet/s, corresponding to a lightly loaded network, up to 5 packet/s, which is instead the case of an overloaded network. In each scenario, 500 simulation runs were performed to collect the data required for the BN inference. Based on this information, the empirical distributions and the average values of ζ_q, p_t and p_q, conditioned on $\mathcal{F}$, were derived.

In the subsequent steps, a new set of 500 simulation runs was performed for each value of λ. In each run, two networks are again randomly deployed; the overall system performance is theoretically evaluated by computing the values of $\overline{\mathcal{W}_k}$, based on the routing tables, and the values of $\overline{\mathcal{W}_k}(\mathcal{D}_1, \mathcal{D}_2)$, with $k \in \{1, 2\}$, where $\mathcal{D}_1$ and $\mathcal{D}_2$ are the optimal sets of nodes to be shared, according to the game theoretic framework proposed in Section 4.3.3.

The aim is to verify how much gain is achievable with our approach with respect to a random selection of the nodes shared and a fully cooperative strategy. Therefore, the network performance obtained by using our Game Theoretic node selection strategy is compared to those achieved by the following strategies: 1) no cooperation, 2) two nodes shared, randomly chosen by each network, and 3) all nodes shared.

Bayesian Network estimation

Exploiting the stochastic estimation of local parameters through the BN approach proposed in Section 4.3.1, we can evaluate the expected value of the three parameters of interest, namely the average delivery delay ζ_q, the probability of buffer overflow p_q and the probability of transmission failure p_t, as a function of the number of flows $\mathcal{F}$ passing through the node and of the traffic intensity λ. The expected values of ζ_q, p_q, and p_t are shown in Figs. 4.11, 4.12 and 4.13, respectively. We notice that the highest number of flows through a single node is reached when that node becomes the only connection among three separate clusters of nodes. If these groups have similar cardinalities, and the number of nodes in each network is N, we can rise up to a maximum of about $4(N-1)^2/3$ flows through a single node, that is close to the maximum value of $\mathcal{F}$ represented in the figures. We also observe in Fig. 4.11 that for very high values of $\mathcal{F}$ and λ, the average delivery delay decreases. We conjecture that this happens for two reasons: (1) the queue of these nodes are always almost full, so that the time to traverse them cannot grow much further, whereas (2) a node traversed by a high number of flows is often chosen as receiver by most of his neighbors. For these reasons, when it transmits, a lower number of communications can interfere, thus leading to a lower time needed to deliver a packet to the next hop.

Cooperation performance

In Fig. 4.14, we present the actual gain, in terms of delay reduction, offered by the considered cooperation strategy. The curves are obtained by averaging over 500 random topologies, each consisting of two networks of $N = 10$ nodes. The other system parameters are reported in Tab. 4.2. We plot the average weighted delay of each network (due to the symmetry of the scenario, it is not necessary to distinguish between the networks) in four different cases, that is: (1) when no nodes are shared, namely No Coop; (2) when 2 nodes randomly chosen are shared, namely 2 Rand; (3) when 2 nodes, selected through the proposed Game-theoretic approach, are shared, namely 2 GT; (4) when all nodes are shared, namely Full Coop.

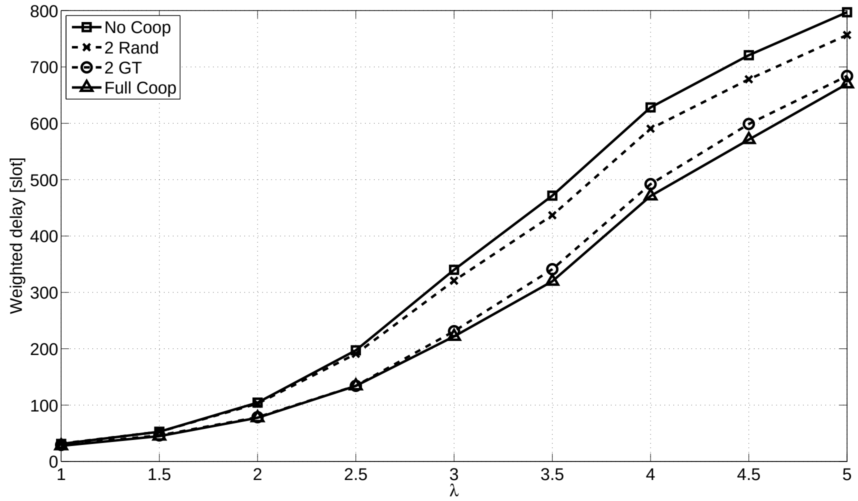

Figure 4.14: Weighted delay as a function of the packet generation intensity λ, for the four compared scenarios: with no nodes shared (No Coop); with two nodes shared, randomly chosen (2 Rand); with two nodes shared, chosen via Game Theory (2 GT); and with all the nodes shared (Full Coop).

It can be observed that, as intuition suggests, full cooperation grants the highest benefits, due to the higher diversity. Hence, this is the maximum achievable gain for the scenario investigated. This gain is more pronounced when the networks are heavily loaded, since congested paths are more frequent, and spatial diversity becomes more advantageous.

When only two nodes can be shared, the choice of the shared nodes makes the difference. In fact, Fig. 4.14 shows that a careful selection of the resources to be shared can significantly increase the achievable gain when compared to a blind random selection. A random selection can not offer a significant gain for lightly loaded networks, while, for heavily loaded networks, it can offer only one third of the gain granted by full cooperation. On the contrary, if the same number of nodes are shared, but chosen by means of our game-theoretic approach, the maximum achievable gain is fully obtained for lightly loaded networks and closely approached for heavily loaded networks.

5 Conclusions

In Section 1.1, a multi-user network was analysed, where multiple secondary TX-RX pairs coexist with multiple primary systems, in the same geographical area and utilizing the same spectral band. The TXs employ multiple antennas. The main results are:

- The characterization of the Pareto boundary for the utility region of the secondary users where fixed prices weights are given. For each secondary user, the utility function is defined as its achievable rate minus the sum of weighted interference rates this user produces at the primary users. The characterization does not hold for the rate region under soft- or peak-power-shaping constraints. Under such conditions, each Pareto rate tuple is achieved under different prices weights. Acquiring these weights would require an iterative process.
- Given a set of existing primary users, the Pareto boundary of the secondary users' achievable rate region is characterized under the null-shaping constraints. The Pareto efficient strategies can be performed by the secondary users if they cooperate.
- Motivated by distributed (noncooperative) operation of the secondary systems, null-shaping constraints are designed to improve the efficiency of the noncooperative systems. The null-shaping constraints, corresponding to virtual primary users, are characterized such that all points on the Pareto boundary of the rate region without constraints are achieved by the noncooperative secondary systems. This result shows that imposing null-shaping constraints on the secondary users can be sufficient to improve their noncooperative outcome.
- Following the previous result, the noncooperative secondary user selection problem is investigated to increase the achievable sum rate. The selection should activate only a subset of the existing secondary systems for operation. Assuming that the secondary users are noncooperative and null-shaping constraints corresponding to existing primary users exist, a secondary user selection algorithm is proposed that improves the sum performance of the systems. The algorithm is greedy such that in each iteration step, activating a secondary system has to increase the sum rate of the selected secondary systems set.

In Section 1.2, a cellular setup including two multi-antennas BSs was studied, where each BS has a protected band which it uses exclusively and also a shared band which is shared with the other BS. The main contributions are:

- Multi-antenna technology is applied over the protected bands through the ZFB strategy, so that each BS guarantee zero interference to its customers, while the two BSs interfere on each other in the shared band.

- In order to avoid the large amount of overhead that cooperation induces, the two BSs are assumed non-cooperative and the situation between the two BSs is formulated as a non-cooperative game in which the objective of each BS is to maximize the system data rate.
- The existence of an NE is proven and sufficient conditions are provided for its uniqueness. By proving that the non-cooperative game is a supermodular game, the uniqueness of NE implies the global convergence of best response dynamics (iterative waterfilling).
- The efficiency of the NE is compared to the cooperative maximum sum rate by extensive simulations. It is shown that the setup of protected and shared bands provides significant efficiency gains to the non-cooperative systems.

In Section 2.1, it was assumed that not only the TXs, but also the RXs have multiple antennas. The distributed design of beamforming vectors was considered under the practical assumption that the TXs and RXs only have local CSI, i.e. from the links directly connected to them. The main contributions are:

- The distributed beamforming design in the MIMO IC with partial CSI is modelled as a Bayesian game.
- Equilibria are derived for two versions of the Bayesian games; the egoistic (which maximizes the beamforming gain at the intended receiver) and the altruistic (which minimizes the interference created towards other receivers).

In Section 2.2, the autonomous and decentralized allocation of rate and power was studied in a two-user single-antenna IC with partial CSI. First, the practical case was considered where the rate is fixed and each source has to determine the power which maximizes its utility. Then, the general case was considered where each source has to select its strategy defined in terms of the power and the transmitting rate, jointly. Finally, the optimization approach was also considered in the two asymptotic regimes of interference limited and noise limited systems. The main achievements are:

- NE are determined with low complexity via best response algorithms of an equivalent game (interference-limited regime and general case) or closed-form expressions (noise limited regime).
- NE are analyzed in terms of existence, multiplicity, and convergence of the best response. Conditions for uniqueness and convergence are provided.
- Only for the interference-limited and noise-limited regimes, optimum can be determined via closed-form expressions. In the general case, it requires exhaustive search.
- NE leads to fairer resource allocation than optimization.

In Section 3.1, zero-sum stochastic games with two players and perfect information were studied. The contributions are the following:

- A stochastic game model is introduced. It is proven that, for all discounted factors close enough to 1, the discounted value belongs to the field of rational functions with real coefficients.

- Two algorithms are proposed which compute a pair of uniform discount optimal strategies $(\mathbf{f}^*, \mathbf{g}^*)$, which are optimal in the long run average criterion as well.
- The convergence in a finite time of the first algorithm, based on policy improvement, is proven. A simple method is shown to find the range of discount factors in which $(\mathbf{f}^*, \mathbf{g}^*)$ are discount optimal.
- It is shown by simulations that the second algorithm has a lower complexity than the first one, in terms of number of pivot operations.
- For transient stochastic games, it is proven that $(\mathbf{f}^*, \mathbf{g}^*)$ are optimal under the undiscounted criterion as well.

In Section 3.2, discount cooperative MDPs were considered, in which the payoffs at each stage are multiplied by a discount factor and summed up over time. The outcomes of the study are:

- Non-cooperative and cooperative multi-agent MDPs are shortly surveyed and a stationary stage-wise CPDP for cooperative discounted MDPs (MDP-CPDP) is proposed.
- It is proven that the proposed MDP-CPDP satisfies 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.
- It is shown that the proposed MDP-CPDP fulfills the time consistency property, which is a crucial one in repeated games theory: it suggests that a CPDP should respect the terminal fairness property in a subgame starting from any time step.
- It is shown that, under some conditions, for all discount factors small enough, also the greedy players having a myopic perspective of the game are satisfied with the proposed MDP-CPDP.
- The n-tuple step cooperation maintenance property of a CPDP is investigated. 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 step n 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. It is found that the single step cooperation maintenance property is the strongest one among all n's. Furthermore, a necessary and sufficient condition is given for the proposed MDP-CPDP to satisfy the n-tuple step cooperation maintenance property, for all integers n.

In Section 3.2, dynamic cooperative game theory was introduced, where the game is not played one-shot but rather over an infinite horizon. Two criteria were taken into account to sum over time the payoffs earned in each single stage game, specifically the average and the discount criterion. The main results are:

- It is proven that an exponential number of queries is necessary for any deterministic algorithm even to approximate SSM with polynomial accuracy.
- Three randomized algorithms are proposed to compute a confidence interval for SSM. The first one, SCI, assumes that the coalition values in each state are available off-line to the estimator agent. SCI can be seen as a benchmark for the performance of the other two methods, DCI1 and DCI2. The last two methods can be utilized also if we pragmatically assume that the estimator learns the coalition values in each static game while the Markov chain process unfolds. DCI2 reveals the most natural connection between confidence intervals of Shapley value in static games and in Markovian games.
- As a by-product of the study of DCI2, confidence intervals are provided for the Shapley-Shubik index in static games. Also, a straightforward way to optimize the tightness of DCI1 is proposed.
- The proposed three approaches are compared in terms of tightness of the confidence interval. It is proven that DCI1 is tighter than SCI, with an equal number of queries and for a suitable choice of the number of queries on coalition values in each state. This occurs essentially because DCI1 allows us to tune the number of samples according to the weight of the state. The simulations confirmed that DCI2 is more accurate than the SCI and DCI1 when both the confidence probability is close to 1 and a tight confidence interval for the Shapley-Shubik index of static games is available, like the Clopper-Pearson interval.
- Finally, it is shown that a polynomial number of queries is sufficient to achieve a polynomial accuracy for the proposed algorithms. Hence, in order to compute SSM, the proposed randomized approaches are more accurate than any deterministic approach for a number of players sufficiently high. The three proposed randomized approaches also produce confidence intervals for the Shapley value in *any* cooperative Markovian game.

In Section 3.4, a scenario was considered where several providers share a network to provide connection towards a unique common destination to their customers. A coalition game framework was provide to facilitate the design of the available network links and their costs such that there exists an optimum routing strategy and a cost sharing satisfying all the subsets of providers. The main results are:

- By using the framework of stochastic games, algorithms are provided to compute the minimum costs that each coalition of providers can ensure for itself. This helps the optimum design of a network, which should guarantee the existence of an efficient and stable costs partition among the providers.
- Also situations are modeled in which there are two players with conflicting interests, like a hacker against a service provider, or in which a service provider wants to reduce the damages to the network caused by a natural disaster.
- An epidemic spread network model is shown as well. From a theoretical perspective, some results on uniform optimal strategies in stochastic game are extended to

the case of undiscounted criterion.

In Section 3.5, distributed networks were considered, in which users can cooperate and pursue their own interest. Cooperative game theory with NTU was used, to derive efficient and stable allocations in a setting in which the users can cooperate to reach a common goal. The main achievements are:

- The relation between the static and the Markovian game is studied. The Core of the Markovian game is still the most attractive set of rate allocations, both from a centralized point of view and for the single users. Both the discounted and the average criterion to sum the rate over time are considered.
- The Core of the Markovian game is found nonempty and its connections with the Core of the single stage game are studied. Also the possibility that coalitions can change over time, along the Markov process, is considered and it is found that the allocation is still stable, for any subgame.
- It is shown that, under the discounted criterion, the procedure of joint rate allocations in the Markovian game for each starting state of the HMC is a delicate procedure. Indeed, the associated single stage allocations may not be feasible. Thus, a way is proposed to ensure their feasibility.
- The α-fair allocation procedures are analyzed, with particular attention to the max-min fair ($\alpha \to \infty$) and to the proportional fair ($\alpha \to 1$). A condition is found on the single stage games ensuring that, if the single stage allocations all satisfy such criteria, they also do it in the long run game. Moreover, the Nash bargaining solution is defined.
- The situation that an agreement among the users is not found, and everybody threatens to jam is investigated and it is proven that, if the number of players P increases, the probability that some user can still communicate tends to 0 with exponential rate. As a by product of this analysis, it is found that the Nash bargaining solution tends to all the three fair criteria when P tends to infinity.
- It is proven that in the Markovian channel the expected sum rate in the long run game tends to 0 when P tends to infinity.

In Chapter 4, game theory was used to address resource allocation problems arising from spectrum and infrastructure sharing.

- In Section 4.1, considering only spectrum sharing, an upper bound on the system performance in terms of aggregate capacity was calculated and the existence of a possible sharing gain is shown. The main factors impacting it are the number of users in the network, the frequency diversity, and the traffic load for each operator.
- In Section 4.2, the theory of coalitional games was exploited to analyse cooperative relaying. The proper cooperation mechanism is derived so that a gain is obtained by both those who have their transmissions relayed to the final destination and also those who act as relays. This gain can be directly related to a throughput improvement if the users follow specific access procedures.

- In Section 4.3, the focus moved to a scenario where two cellular network operators are willing to share some of their relays to gain benefits in terms of lower packet delivery delay and reduced loss probability. Bayesian Network analysis is exploited to compute the correlation between local parameters and overall performance, whereas the selection of the nodes to share is made by means of a game theoretic approach. The results show that an accurate selection of the shared nodes can significantly increase the performance gain with respect to a random selection scheme.

Bibliography

[1] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[2] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, “Breaking spectrum gridlock with cognitive radios: An information theoretic perspective,” *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[3] G. Scutari, D. Palomar, J.-S. Pang, and F. Facchinei, “Flexible design of cognitive radio wireless systems,” *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 107–123, 2009.

[4] X. Shang, B. Chen, and H. V. Poor, “On the optimality of beamforming for multi-user MISO interference channels with single-user detection,” in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, 2009.

[5] R. Zhang and Y.-C. Liang, “Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 88–102, 2008.

[6] R. Zhang and S. Cui, “Cooperative interference management in multi-cell down-link beamforming,” in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[7] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, “Distributed resource allocation schemes,” *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 53–63, 2009.

[8] J.-S. Pang, G. Scutari, D. Palomar, and F. Facchinei, “Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3251–3271, 2010.

[9] S. Vishwanath and S. A. Jafar, “On the capacity of vector Gaussian interference channels,” in *Proc. of Information Theory Workshop (ITW)*, 2004, pp. 365–369.

[10] R. Mochaourab and E. A. Jorswieck, “Optimal beamforming in interference networks with perfect local channel information,” *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1128–1141, 2011.

[11] E. A. Jorswieck, “Beamforming in interference networks: Multicast, MISO IFC and secrecy capacity,” in *Proc. of International Zurich Seminar on Communications (IZS)*, 2010, invited.

[12] G. Scutari, D. Palomar, and S. Barbarossa, “Cognitive MIMO radio: A competitive optimality design based on subspace projections,” *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 46–59, 2008.

[13] E. A. Jorswieck and R. Mochaourab, “Beamforming in underlay cognitive radio: Null-shaping design for efficient Nash equilibrium,” in *Proc. of International Workshop on Cognitive Information Processing (CIP)*, 2010, invited.

[14] M. J. Osborne and A. Rubinstein, *A course in Game Theory*. MIT Press, 1994.

[15] E. G. Larsson, D. Danev, and E. A. Jorswieck, “Asymptotically optimal transmit strategies for the multiple antenna interference channel,” in *Proc. of Allerton Conference on Communications, Control and Computing*, 2008.

[16] R. B. Myerson, “An introduction to game theory,” Northwestern University, Center for Mathematical Studies in Economics and Management Science, Discussion Papers 623, 1984.

[17] G. Dimic and N. Sidiropoulos, “On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm,” *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, 2005.

[18] E. A. Jorswieck, P. Svedman, and B. Ottersten, “Performance of TDMA and SDMA based opportunistic beamforming,” *IEEE Transactions on Communications*, vol. 7, no. 11, pp. 4058–4063, 2008.

[19] 3GPP TS 36.213 V10.1.0,, “Technical specification group radio access network (E-UTRA); physical layer procedures,” Tech. Rep., Mar. 2011.

[20] T. Yoo and A. Goldsmith, “On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[21] R. Mochaourab and E. A. Jorswieck, “Resource allocation in protected and shared bands: Uniqueness and efficiency of Nash equilibria,” in *Proc. ICST/ACM International Workshop on Game Theory in Communication Networks (Gamecomm)*, Oct. 2009, pp. 1–10.

[22] E. A. Jorswieck and R. Mochaourab, “Power control game in protected and shared bands: Manipulability of Nash equilibrium,” in *Proc. International Conference on Game Theory for Networks (GameNets)*, May 2009, pp. 428–437, invited.

[23] R. Amir, “Supermodularity and complementarity in economics: An elementary survey,” *Southern Economic Journal*, vol. 71, no. 3, pp. 636–660, Jan. 2005.

[24] D. M. Topkis, *Supermodularity and Complementarity*. Princeton University Press, 1998.

[25] X. Vives, “Complementarities and games: New developments,” *Journal of Economic Literature*, vol. 43, no. 2, pp. 437–479, June 2005.

[26] C. Saraydar, N. Mandayam, and D. Goodman, “Efficient power control via pricing in wireless data networks,” *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291–303, Feb. 2002.

[27] E. Altman and Z. Altman, “S-modular games and power control in wireless networks,” *IEEE Trans. Autom. Control*, vol. 48, no. 5, pp. 839–842, May 2003.

[28] R. Mochaourab, N. Zorba, and E. Jorswieck, “Nash equilibrium in multiple antennas protected and shared bands,” in *International Symposium on Wireless Communication Systems (ISWCS)*, 2012.

[29] E. V. Belmega, B. Djeumou, and S. Lasaulce, “Resource allocation games in interference relay channels,” in *Proc. International Conference on Game Theory for Networks (GameNets)*, May 2009, pp. 575–584.

[30] G. Scutari, D. Palomar, and S. Barbarossa, “Competitive design of multiuser MIMO systems based on game theory: A unified view,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1089–1103, Sep. 2008.

[31] E. A. Jorswieck, E. G. Larsson, and D. Danev, “Complete characterization of the pareto boundary for the MISO interference channel,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5292–5296, 2008.

[32] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

[33] J. C. Harsanyi, “Games with incomplete information played by Bayesian players, I-III. Part I. The basic model,” *Management Science*, vol. 14, no. 3, pp. 159–182, 1967.

[34] R. Zakhour and D. Gesbert, “Coordination on the MISO interference channel using the virtual SINR framework,” in *Proc. of ITG International Workshop on Smart Antennas (WSA)*, 2009.

[35] M. Y. Ku and D. W. Kim, “Tx-Rx beamforming with multiuser MIMO channels in multiple-cell systems,” in *Proc. of International Conference on Advanced Communication Technology (ICACT)*, 2008.

[36] W. Choi and J. Andrews, “The capacity gain from intercell scheduling in multi-anetnna systems,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, pp. 714–725, 2008.

[37] K. S. Gomadam, V. R. Cadambe, and S. A. Jafar, “Approaching the capacity of wireless networks through distributed interference alignment,” in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, 2008.

[38] S. Y. Shi, M. Schubert, and H. Boche, “Rate optimization for multiuser MIMO systems with linear processing,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020–4030, 2008.

[39] F. R. Farrokhi, K. J. R. Liu, and L. Tassiulas, “Transmit beamforming and power control for cellular wireless systems,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, 1998.

[40] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge University Press, 2003.

[41] A. F. Molisch, *Wireless Communications*. IEEE, 2005.

[42] G. N. He, M. Debbah, and S. Lasaulce, “K-player bayesian waterfilling game for fading multiple access channels,” in *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2009.

[43] Scutari, Palomar, and Barbarossa, “Asynchronous iterative water-filling for gaussian frequency-selective interference channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2868–2878, 2008.

[44] R. Etkin, A. Parekh, and D. Tse, “Spectrum sharing for unlicenced bands,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 517–528, 2007.

[45] J. Rosen, “Existence and uniqueness of equilibrium points for concave N-person games,” *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.

[46] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Springer Verlag, 1997.

[47] J. Filar, “Ordered field property for stochastic games when the player who controls transitions changes from state to state,” *Journal of Optimization Theory and Applications*, vol. 34, no. 4, pp. 503–513, 1981.

[48] T. Raghavan and Z. Syed, “A policy-improvement type algorithm for solving zero-sum two-person stochastic games of perfect information,” *Mathematical Programming*, vol. 95, no. 3, pp. 513–532, 2003.

[49] R. Jeroslow, “Asymptotic linear programming,” *Operations Research*, vol. 21, pp. 1128–1141, 1973.

[50] A. Hordijk, R. Dekker, and L. Kallenberg, “Sensitivity analysis in discounted markov decision processes,” *OR Spektrum*, vol. 7, no. 3, pp. 143–151, 1985.

[51] E. Altman, K. Avrachenkov, and J. Filar, “Asymptotic linear programming and policy improvement for singularly perturbed markov decision processes,” *ZOR: Mathematical Methods of Operations Research*, vol. 49, no. 1, pp. 97–109, 1999.

[52] J. Filar, E. Altman, and K. Avrachenkov, “An asymptotic simplex method for singularly perturbed linear programs,” *Operations Research Letters*, vol. 30, no. 5, pp. 295–307, 2002.

[53] K. Knopp, *Theory and Application of Infinite Series*. Springer, 1990.

[54] L. Shapley, “A value for n-person games,” *Contributions to the Theory of Games*, vol. 2, pp. 31–40, 1953.

[55] Aumann, “Economic Applications of the Shapley Value,” *In: S.S. J.-F. Mertens (ed.) Game-Theoretic Methods in General Equilibrium Analysis, Kluwer Academic Publisher*, pp. 121–133, 1994.

[56] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein, “Internet Economics: The use of Shapley value for ISP settlement,” *Proceedings of CoNEXT 2007*, pp. 1–5, 2007.

[57] R. Stanojevic, N. Laoutaris, and P. Rodriguez, “On economic heavy hitters: Shapley value analysis of 95th-percentile pricing,” *In: Proceedings of the 10th annual conference on Internet measurement ACM*, vol. 48, pp. 75–80, 2010.

[58] K. Avrachenkov, J. Elias, F. Martignon, and L. Neglia, G.and Petrosyan, “A Nash bargaining solution for Cooperative Network Formation Games,” *NETWORKING*, pp. 307–318, 2011.

[59] R. Aumann and S. Hart, *Handbook of Game Theory with Economic Applications*. Elsevier, 1994, vol. 2.

[60] F. Javadi, M. Kibria, and A. Jamalipour, “Bilateral Shapley Value Based Cooperative Gateway Selection in Congested Wireless Mesh Networks,” *IEEE GLOBECOM 2008*, pp. 1–5, 2008.

[61] L. Shapley and M. Shubik, “A method for evaluating the distribution of power in a committee system,” *The American Political Science Review*, vol. 48, no. 3, pp. 787–792, 1954.

[62] A. Taylor and A. Pacelli, *Mathematics and Politics: Strategy, Voting, Power, and Proof*. Springer Verlag, 2008.

[63] A. Predtetchinski, “The strong sequential core for stationary cooperative games,” *Games and economic behavior*, vol. 61, no. 1, pp. 50–66, 2007.

[64] L. Petrosjan, “Cooperative stochastic games,” *Advances in Dynamic Games*, pp. 139–145, 2006.

[65] Y. Bachrach, E. Markakis, E. Resnick, A. Procaccia, J. Rosenschein, and A. Saberi, “Approximating power indices: theoretical and empirical analysis,” *Autonomous Agents and Multi-Agent Systems*, vol. 2, pp. 105–122, 2010.

[66] J. Banzhaf III, “Weighted voting doesn’t work: A mathematical analysis,” *Rutgers L. Rev.*, vol. 19, pp. 317–343, 1964.

[67] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

[68] K. Avrachenkov, L. Cottatellucci, and L. Maggi, “Cooperative Markov decision processes: Time consistency, greedy players satisfaction, and cooperation maintenance,” Eurecom, Tech. Rep. EURECOM+3601, 01 2012. [Online]. Available: http://www.eurecom.fr/publication/3601

[69] K. Athreya and C. Fuh, “Bootstrapping Markov chains: countable case,” *Journal of Statistical Planning and Inference*, vol. 33, pp. 311–331, 1992.

[70] K. Avrachenkov, L. Cottatellucci, and L. Maggi, “Confidence intervals for Shapley value in Markovian dynamic games,” Eurecom, Tech. Rep. EURECOM+3602, 01 2012. [Online]. Available: http://www.eurecom.fr/publication/3602

[71] H. Chernoff, “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations,” *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, 1952.

[72] E. Wilson, "Probable inference, the law of succession, and statistical inference," *JASA*, pp. 209–212, 1927.

[73] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Mathematical Statistics*, vol. 10, pp. 105–118, 1939.

[74] A. Agresti and B. Coull, "Approximate is better than "exact" for interval estimation of binomial proportions," *American Statistician*, pp. 119–126, 1998.

[75] C. Clopper and E. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.

[76] T.Roughgarden, "Routing games," *in N. Nisan, T.Roughgarden, E. Tardos, V. Vazirani, Algorithmic Game Theory, Cambridge University Press*, pp. 461–486, 2007.

[77] L. Maggi, K. Avrachenkov, and L. Cottatellucci, "Stochastic games for cooperative network routing and epidemic spread," in *ICC 2011 Workshop on Game Theory and Resource Allocation for 4G, June 5-9, 2011, Kyoto, Japan*, Kyoto, JAPAN, 06 2011. [Online]. Available: http://www.eurecom.fr/publication/3338

[78] M. Madiman, "Cores of cooperative games in information theory," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, Jan. 2008.

[79] B. Peleg and P. Sudhölter, *Introduction to the Theory of Cooperative Games*. Springer Verlag, 2007.

[80] R. J. La and V. Anantharam, "A game-theoretic look at the Gaussian multiaccess channel," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, pp. 87–105, 2004.

[81] E. Altman, K. Avrachenkov, L. Cottatellucci, M. Debbah, G. He, and A. Suarez, "Operating point selection in multiple access rate regions," in *IEEE, Teletraffic Congress, 2009. ITC 21 2009. 21st International*, Paris, FRANCE, 09 2009, pp. 1–8.

[82] S. V. Hanly and D. N. Tse, "Multi–access fading channels: Part II: Delay–limited capacities," *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[83] J. Herzog and T. Hibi, *Monomial Ideals*. Springer, 2010.

[84] J. Edmonds, "Submodular functions, matroids, and certain polyhedra," in *Combinatorial Optimization - Eureka, You Shrink!*, Paris, FRANCE, 2003, pp. 11–26.

[85] A. Feldman and A. Kirman, "Fairness and envy," *The American Economic Review*, vol. 64, no. 6, pp. 995–1005, 1974.

[86] J. F. N. Jr., "The bargaining problem," *Econometrica: Journal of the Econometric Society*, pp. 155–162, 1950.

[87] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.

[88] J. Filar and L. A. Petrosjan, "Dynamic cooperative games," *International Game Theory Review*, vol. 2, no. 1, pp. 47–65, 2000.

[89] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[90] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[91] K. W. Shum and C. W. Sung, "On the fairness of rate allocation in gaussian multiple access channel and broadcast channel," 2006. [Online]. Available: Arxivpreprintcs/0611015

[92] H. Boche and M. Schubert, "Nash bargaining and proportional fairness for wireless systems," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 5, pp. 1453–1466, 2009.

[93] E. A. Jorswieck, L. Badia, T. Fahldieck, M. Haardt, E. Karipidis, J. Luo, R. Piesiewicz, and C. Scheunert, "Resource sharing improves the network efficiency for network operators," in *Proc. of Wireless World Research Forum Meeting*, 2011.

[94] G. Hardin, "The Tragedy of the Unmanaged Commons," *Trends in Ecology & Evolution*, vol. 9, no. 5, 1994.

[95] L. Anchora, L. Badia, E. Karipidis, and M. Zorzi, "Capacity Gains due to Orthogonal Spectrum Sharing in Multi-Operator LTE Cellular Networks," in *International Symposium on Wireless Communication Systems (ISWCS)*, 2012.

[96] L. Canzian, L. Badia, and M. Zorzi, "Relaying in Wireless Networks Modeled through Cooperative Game Theory," in *Proc. CAMAD*, 2011, pp. 97–101.

[97] G. Quer, F. Librino, L. Canzian, L. Badia, and M. Zorzi, "Using Game Theory and Bayesian Networks to Optimize Cooperation in Ad Hoc Wireless Networks," in *Proc. ICC*, 2012.

[98] L. Anchora, L. Canzian, L. Badia, and M. Zorzi, "A Characterization of Resource Allocation in LTE Systems Aimed at Game Theoretical Approaches," in *Proc. CAMAD*, 2010, pp. 47–51.

[99] ns-3 simulator. [Online]. Available: http://www.nsnam.org/

[100] L. Anchora, M. Mezzavilla, L. Badia, and M. Zorzi, "Simulation Models for the Performance Evaluation of Spectrum Sharing Techniques in OFDMA Networks," in *Proc. ACM MSWiM*, 2011, pp. 249–256.

[101] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[102] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. IT-25, pp. 572–584, Sep. 1979.

[103] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. ISIT*, Seattle, WA, Jul. 2006, pp. 1668–1672.

[104] M. Stein, “Towards optimal schemes for the half-duplex two-way relay channel,” *IEEE Journal on Selected Areas in Communications*, 2011, submitted to, available as arXiv preprint at http://arxiv.org/abs/1101.3198.

[105] M. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT, 1994.

[106] J. Lemaire, *Cooperative Game Theory*. Wiley Online, 2006.

[107] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, “Coalitional game theory for communication networks: a tutorial,” *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 77–97, Sep. 2009.

[108] D. Cox, *Renewal Theory*. London, UK: Methuen & Co., 1970.

[109] L. Badia, M. Levorato, and M. Zorzi, “Markov analysis of selective repeat type II hybrid ARQ using block codes,” *IEEE Transactions on Communications*, vol. 56, no. 9, pp. 1434–1441, Sep. 2008.

[110] L. Badia, M. Levorato, F. Librino, and M. Zorzi, “Cooperation techniques for wireless systems from a networking perspective,” *IEEE Wireless Communications Magazine*, vol. 17, no. 2, pp. 89–96, Apr. 2010.

[111] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[112] G. Quer, H. Meenakshisundaram, B. Tamma, B. S. Manoj, R. Rao, and M. Zorzi, “Using Bayesian Networks for Cognitive Control of Multi-hop Wireless Networks,” in *Proceedings of IEEE MILCOM*, San Jose, CA, US, Nov. 2010.

[113] G. Tan and J. Guttag, “The 802.11 MAC protocol leads to inefficient equilibria,” *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05)*, vol. 1, pp. 1–11, Mar. 2005.

[114] R. Ma, V. Misra, and D. Rubenstein, “Modeling and Analysis of Generalized Slotted-Aloha MAC Protocols in Cooperative, Competitive and Adversarial Environments,” *Proceedings of the 24th IEEE International Conference on Distributed Computing Systems (ICDCS ’06)*, p. 62, Jul. 2006.

[115] L. Lifeng and H. El Gamal, “The Water-Filling Game in Fading Multiple-Access Channels,” *IEEE Trans. on Information Theory*, vol. 54, no. 5, pp. 722–730, May 2008.

[116] M. Cagalj, S. Ganeriwal, I. Aad, and J. P. Hubaux, “On Selfish Behavior in CSMA/CA Networks,” in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05)*, vol. 4, 2005, pp. 2513–2524.

[117] V. Srivastava, J. Neel, A. B. MacKenzie, R. Menon, L. A. DaSilva, J. E. Hicks, J. H. Reed, and R. P. Gilles, “Using Game Theory to Analyze Wireless Ad Hoc Networks,” *IEEE Communications Surveys and Tutorials*, vol. 7, no. 4, pp. 46–56, 2005.

[118] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*. Springer, 2007.

[119] G. Schwarz, “Estimating the Dimension of a Model,” *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[120] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot, “Optimized Link State Routing protocol for ad hoc networks,” in *Proceedingsof the IEEE International Multi Topic Conference IEEE INMIC.*, 2001.

[121] G. Owen, *Game Theory*, 3rd ed. New York: Academic, 2001.