Information Theory in Multi-Agent Learning Systems

By

DAVID ALEJANDRO MARTÍNEZ VÁSQUEZ.

Advisor

EDUARDO MOJICA NAVA, PH.D.

Department of Electrical and Electronics Engineering
UNIVERSIDAD NACIONAL DE COLOMBIA

A dissertation submitted to Universidad Nacional de Colombia, in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Electrical and Electronics Engineering.

JUNE 16, 2020

Abstract

In this work, we propose a multi-agent learning framework based on the mutual information between the agents and their environment. Initially, each agent, based on its neighborhood information, uses the Gaussian process regression (GPR) to infer the environment behavior. Then, a minimization of the mutual information between an agent and the environment is calculated by means of the rate distortion function (RDF). In this way, a border between misunderstanding and redundancy of the environment information is obtained, which is used as a decision rule by the agents. The calculation of the RDF is conveniently performed through the Blahut-Arimoto algorithm, from which, the most important elements for our model are the Lagrange multiplier $\mathfrak{s}$, and the conditional distribution describing the similitude between the agent and the environment. The parameter $\mathfrak{s}$ plays an important role in the rationality level assumed by the agents in the decision making process. On the other hand, due to its Boltzmann distribution form, the conditional probability distribution establishes a *Logit dynamics* pattern, used by the agents as a rule for the action selection. Finally, we include a distributed optimization setting by means of the potential games approach, in which the *Nash equilibrium* convergence is found through a *distortion based potential function*.

The framework, in spite of being mainly implemented in mobile sensor networks, demonstrates applicability in other multi-agent contexts, such as smart grids.

DEDICATION AND ACKNOWLEDGEMENTS

This work is dedicated to the memory of my brother and best fiend Milo, who even from another dimension is still teaching me to live. To my mother Ofe, for all I am. To my sisters, Lulú and Mandis, my nieces Chalo, Sari and Sofi, and my nephew Santi for their constant and unconditional support. To Nata for her help, advices, comprehension and patience in the bad moments, and for being always will to listen to my ideas.

I give especial thanks to mi advisor Eduardo Mojica, for his confidence, valuable advices about researching, and the contribution to my work.

Table of Contents

LIST OF TABLES

TABLE **Page**

List of Figures

Figure **Page**

CHAPTER 1

INTRODUCTION

In the last years, the field of multi-agent systems has gained lots of interest in the research community to develop solutions in areas such as smart grids, conventional power networks, social networks, static and mobile sensor networks, communication networks, among others, in order to provide to the elements of a system abilities to make decisions in a *decentralized form*, since the conditions of isolation in the case of microgrids, or the high data traffic to sink nodes, in the case of mobile sensor networks, and the impossibility to have a continuous connection between all the network nodes in the general case, make the centralized dependency more difficult every day. The most relevant characteristics, which have begun to be intrinsic of multi-agent systems, are the distributed control and optimization, whose implementation has a narrow relationship with game theory. Additionally, the learning capacity on each agent requires a network adaptation to maintain the environment understanding in spite of the continuous interconnection change. In this sense, multiple learning techniques have been applied in multi-agent systems, among which we can find reinforcement learning, neuronal networks, deep learning, just to mention a few. In the game theory context, the learning process is implicit in something known as the dynamics, which depend on the type of game played, and consequently on the application in which they are used. Some examples are the *replicator dynamics*, used in evolutionary games, and the *Logit* or *best response* dynamics, used in strategic games. In general, the dynamics objective is to define the set of learning rules to choose the strategies offering the best payoffs, and in this way, allow the system to reach an optimal state, which is commonly known as the

Nash equilibrium.

Most of the approaches, especially in mobile sensor networks, that involve game theory and distributed optimization, have been focused on agent utility definitions based on energy consumption, sleep and awake modes, consensus based payoffs, and so on. However, these works have not been concerned about the value of the information found in the agent environment, an how it can improve the decision making and the system convergence towards an equilibrium point. In this regard, we propose a multi-agent learning framework that allows the agents to identify the environmental cues offering the highest welfare, which can be focused to follow redundant signals or, on the contrary, to follow the cues offering the highest difference to the current environment state. This proposed model, begins with an environment perception at each agent, obtained through the Gaussian process regression (GPR) approach, in which the neighbor information is used to infer the state in the agent surroundings. After that, by means of the rate distortion function (RDF), the agents can identify a border between redundancy and misunderstanding about the environment information, and in this way, they can choose the strategies to follow. Finally, the potential games approach is used to include a distributed optimization scheme. In this way, we establish a *Logit dynamics* pattern to define the action selection rule for the agents under the rationality levels established by the Lagrange multiplier $\mathfrak{s}$ associated to the Blahut-Arimoto algorithm, which is a computational and straightforward way to calculate the rate distortion function. The convergence towards an *Nash equilibrium* is guaranteed by a *distortion based potential function*, inspired in the expected distortion associated to the mutual information minimization.

The model adaptability to different contexts is shown through an application in smart grids in Appendix A. However, we focus the implementation to mobile sensor networks due to the relevance of the redundancy identification in settings having a high number of agents recovering information in spatial fields, which can cause clustering formation and redundant covering in determined zones, and consequently, redundant transmissions to sink nodes or data centers. Additionally, the continuous change of the network topology and the node connections do not allow the agents to have a full environment knowledge, which can be modeled through the proposed rationality measure.

In this sense, in the next section we describe the most relevant work related to mobile sensor networks in which the environment information is included in the decision making process, and the distributed control is based on game theory.

1.1 Related Work

The agent environment prediction through GPR (also known as Kriging filter) has been combined with information theory in many works, mainly to find informative positions where the agents can move. In [36], authors use a distributed Gaussian process regression (DPGR) in order to infer the agent environment behavior using just its neighborhood information. In this way, locations with the highest uncertainty are determined by means of the entropy maximization to define a utility function used in the central Voronoi tessellation (CVT) algorithm [26]. The entropy maximization is also applied in [76] to design a sampling strategy for mobile robotic wireless sensor networks (MRWSs) focused on the most informative zones within a spatial field that is described using a Gaussian process. The computational cost on each node associated to the environment prediction is addressed in [78] by means of a sparse Gaussian process. The authors in this work compare three strategies to find the most informative locations for the agents: mutual information based measurement selection algorithm (MI), principal feature analysis (PFA) and informative vector machine (IVM). In [119], a Gaussian process is proposed to describe an anisotropic field where mobile nodes find their next position according to a centralized sampling strategy based on the Fisher Information Matrix minimization. In [14], authors use Gaussian processes to model a scalar field in which underwater vehicles (AUVs) define their movement through the entropy minimization between un-sampled and sampled positions. Despite of the fact that these works address the uncertainty level in spatial fields, the redundancy of the environment information is not considered as a learning factor for the agent decision making.

The redundancy in the environment information has also been taken into account in approaches focused on the energy consumption, which have been covered mainly from the data collection and data aggregation perspectives[6]. In the case of the data collection, in [54] a compressed sensing (CS) theory is proposed in order to reduce the sampling points on each sensor, which leads to the reduction of energy consumption and redundancy, since the low sampled information is reconstructed in data centers where the energy is not limited. In [111], authors propose a model based on a random network to reduce energy consumption in zones with redundant information, defining a sleep-awake schedule for nodes in which the number of active terminals necessary for keeping the network connected is minimized. In the case of the data aggregation, even though most of the research is focused on the network lifetime and energy consumption, data redundancy reduction is also taken into account. In [120], a distributed routing

algorithm based on game theory is proposed to reduce the network load compressing the correlated data between nodes. In [91], authors propose a method based on the rate distortion function, in which, under a given distortion condition, agents estimate the measurements of all their peers within the network and detect correlated information to prevent its transmission. These works, in spite of involving data redundancy, do not consider the agent environment perception as a tool for decision making. Furthermore, most of them require a full information configuration, in which, all the network terminals must be linked.

On the other hand, an analysis of game theory and distributed optimization for sensor networks involving redundancy identification, lead us to the network coverage problem, which has been addressed in literature from two sensing contexts, the static and the mobile. In terms of coverage optimization for static networks, i.e., networks lacking of node movement, most of the research has been focused on sleep and awake scheduling for nodes in order to increase the lifetime of the network, and to decrease the redundancy of monitored locations. In this regard, in [121], authors propose an evolutionary game based algorithm named Game-Theoretical Complete Coverage (GCC) that schedules the sleep and awake modes for nodes in a sensor network to reduce the energy consumption, and to improve the coverage. In this work, the nodes monitoring redundant locations are scheduled to waste energy in a distributed way, and the sensing radius is changed depending on the population in the area. In [2], the coverage problem in wireless sensor networks is addressed using the *k-cover problem*, in which k represents the minimum number of sets, named cover sets, required to cover the whole network. A node belongs just to one k-cover set i, and it is activated to sense in a time slot i, which means that the network lifetime is proportional to k. The k-cover set selection for each node follows a potential game approach, in which the payoff increases if the node is the only one belonging to it. Through the two implemented algorithms, called SNECA and ANECA, authors in this work demonstrate convergence to the Nash equilibrium. In [123], authors propose a potential game in which the utility function depends on the energy/processing cost, and the agent assignment of numerical values to the network locations having relevant events. The Nash equilibrium is also proven in this case. From a mobility context, network coverage has been addressed from different perspectives. From a partial perspective, techniques such as sweep, focused, targeted, and barrier coverages, are the most studied. On the other hand, from a full coverage point of view, many techniques have been proposed, being the most relevant the fuzzy and evolutionary computing, virtual force, and geometry based coverage[71]. In [57], through a game

theoretic approach, authors show the improvement in target detection when the network nodes have mobility abilities. Additionally, they compare the coverage area between static and mobile sensor networks, showing the advantages of mobility. In [110], the Voronoi diagrams are the basis to implement three algorithms to calculate the locations where sensors have to move in order to optimize the network coverage. The first two, named VOR and Minimax, are designed to make the nodes move towards uncovered holes, avoiding the new holes generation in the Minimax case. In the third one, named VEC, the sensors are simulated as electronic particles, and in this way, they move away from densely covered areas. In [21], a sweep coverage case is addressed, in which each point of interest (POI) within the monitored region receives a weight according to their relevance level. In this way, the proposed algorithm allows the mobile sensors to visit the most relevant places more frequently. In [112], a particle Swarm Optimization (PSO) approach based on social behavior of flocking birds is formulated to maximize the network coverage and minimize the energy consumption. In this work, a re-sampling process is introduced to improve the performance of PSO, and the exploration of regions having the highest fitness is controlled through an inertia parameter. In [19], authors define a mission space $\Omega \in \mathbb{R}^2$, and a density function $R(x)$, with $x \in \Omega$, to represent the probability that an event $\mathfrak{s}$ at x exceeds a specific threshold. The event detection follows the model proposed in [25], in which the event signal decays as a polynomial of the distance. The joint detection probability $P(x, \mathbf{a})$, i.e., the probability that the event is detected by the set of agents located at $\mathbf{a}$, is used to maximize the expected event detection frequency. This work, is used in [60] to define the potential function $\phi(\mathbf{a}) = \sum_s R(x)P(x, \mathbf{a})$, which is maximized to optimize the coverage in a spatial field. The maximization is performed by each agent using the algorithm RSAP (Restricted Spatial Adaptive Play) and a wonderful life utility function (WLU [106]). The trade-off between energy consumption and coverage is also studied in [84]. This time, authors use a potential game whose utility function depends on the energy expenditure both for sensing and moving. They show that the convergence time to a Nash Equilibrium decreases if the mutual information between observed and unobserved regions is included within the utility function.

1.2 Main Contribution

Although the aforementioned approaches exhibit a successful performance in static and mobile settings, they are not concerned about the redundancy of information between the agent and the environment, and how it can improve the decision making process. Addi-

tionally, the research works involving game theory do not put an especial interest to the definition of rationality levels at which an agent can get the highest or the lowest understanding about the environment. In this regard, the main contribution of our approach is the ability of the agents to detect environment cues having high or low redundancy, whose identification is performed through the mutual information minimization provided by the rate distortion function. In this way, agents can decide to follow redundant signals in event tracking tasks, or to avoid them if a field exploration is desired. On the other hand, the rationality level given by the parameter $\mathfrak{s}$ of the Blahut-Arimoto algorithm, allows the system to define a high or a low environment understanding. At the highest rationality level, the agents find environment positions having more utility in terms of the distortion reduction. On the contrary, at the lowest rationality level, the agents perform in a highly distorted setting, but developing exploratory skills that contribute to the system convergence towards the *Nash equilibrium*.

As we will show in Chapter 4, the model is implemented in mobile sensor networks, outperforming the work in [60] in terms of the number of agents covering a variable spatial field, and in the number of time steps in an invariant field with obstacles. Additionally, in Appendix A, we demonstrate its applicability in other multi-agent contexts, where the reactive power sharing problem of a microgrid is addressed. Below we show a list of the published and submitted work related to our approach.

1.3 List of Publications

1.3.1 International Journals

1. D.A. Martínez, E. Mojica-Nava, K. Watson, and T. Usländer, *"Multi-Agent Self-Redundancy Identification and Tuned Greedy-Exploration"*. Submitted to the IEEE Transactions on Cybernetics.

2. D.A. Martínez, E. Mojica-Nava, *"Distortion Based Potential Game for Distributed Coverage Control"*. Submitted to the IEEE Transactions on Systems, Man and Cybernetics:Systems.

1.3.2 International Conferences

1. D. A. Martínez, E. Mojica-Nava, K. Watson, and T. Usländer, *"Multi-agent Learning Framework for Environment Redundancy Identification for Mobile Sensors in*

an IoT Context", 3rd International Conference on Smart Data and Smart Cities, Delft-The Netherlands, vol. XLII-4/W11. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018, pp. 33–41.

2. D. A. Martínez and E. Mojica-Nava, *"Information Theory and Self Organization in Sensor Networks"*, 9th International Workshop on Optimization in Logistics and Industrial Applications 2018 - 1st German-French Joint Research Workshop on Industrie 4.0 and Industrie Du Futur, Karlsruhe-Germany, May 2018.

3. D. A. Martínez, R. Rincón, E. Mojica-Nava, and A. Pavas, *"Reactive power sharing in microgrids: An information-theoretical approach"*, 2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC), IEEE, Cartagena-Colombia, Oct 2017, pp. 1–6.

4. D. A. Martínez and E. Mojica-Nava, *"Correlation as a measure for fitness in multi-agent learning systems"*, 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE, Cartagena-Colombia, Nov 2016, pp. 1–6.

5. D. A. Martínez and E. Mojica-Nava, *"Graph transfer function representation to measure network robustness"*, Impact and Advances of Automatic Control in Latinamerica, Medellin-Colombia, Oct 2016, pp. 172–176.

6. D. A. Martínez and E. Mojica-Nava, *"Entropy measures in evolving networks"*, Complex Networks: from theory to interdisciplinary applications, Marseilles-France,July 2016.

7. D. A. Martínez, C. Cusgüen, and E. Mojica-Nava, *"Correlation network with stubborn agents in an opinion dynamic model"*, 2016 Conference on Complex Systems, Amsterdam-The Netherlands, September-2016.

8. D. A. Martínez, R. Rincón, E. Mojica-Nava, *"Reactive Power Sharing in Isolated Micogrid Using a Controller Based on Information Theory"*, Latin American Conference on Complex Networks, Puebla-Mexico, September 2017.

9. D. A. Martínez, E. Mojica-Nava, *"Agent-Environment Mutual Information in a Potential Game Context"*, Latin American Conference on Complex Networks, Cartagena-Colombia, Agosto 2019.

1.4 Document Organization

The rest of this paper is organized as follows. In Chapter 2, we describe the main concepts associated to the definition of the proposed learning model. First, we describe the rate distortion function (RDF) and the associated Blahut-Arimoto algorithm, used to find it in a less complex way. Second, we describe the Gaussian process regression (GPR) approach, and how it will be used to infer the agent information. Finally, we show the main concepts about game theory, with especial emphasis in potential games, the *Logit* and the *replicator dynamics*, which are used in the model implementation for mobile sensor networks and smart grids, respectively.

In Chapter 3, by means of the parameter $\mathfrak{s}$ of the Blahut-Arimoto algorithm, we define the highest and the lowest rationality values for the agent learning process, which determine the maximum and minimum environment understanding, according to the borders established by the rate distortion function.

In Chapter 4, we include the potential game approach in the model, defining a *distortion based potential function* that allows the system to find a Nash equilibrium. Finally, we show the model performance in a mobile sensor network in order to address the network coverage problem. Additionally, the applicability of the model in other multi-agent system context is demonstrated in Appendix A.

CHAPTER 2

PRELIMINARIES

In this chapter, we describe the main concepts used to formulate the proposed multi-agent learning framework. Firstly, we describe the fundamental notions related to information theory, putting especial interest in the rate distortion function and the Blahut-Arimoto algorithm. Secondly, we expose the most relevant definitions about the Gaussian process regression approach, which is used in our model to infer the agent environment behavior. Finally, we describe the game theory definitions associated to our framework formulation, emphasizing in potential games and the *Logit dynamics* model, which is used as a rule for agents to choose their strategies towards a desired system configuration, known as the *Nash equilibrium*.

2.1 Entropy and Mutual Information

The entropy is a measure of the uncertainty of a random variable X, which can take any value x belonging to the alphabet $\mathcal{X}$. This is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x), \tag{2.1}$$

where $p(x)$ is the probability of $X = x$, and the logarithm base is 2, hence, the entropy is given in bits[1] [27] [93]. The above quantity is always positive, since $0 \leq p(x) \leq 1$ and $\log p(x) \leq 0$.

[1] Although we have chosen 2 as the logarithm base, it could take any base value. So, for instance, a logarithm base e results in an entropy measured in nats.

Example 2.1.1. Consider a random variable X, with alphabet $\mathcal{X} = [a,b,c,d,e]$. Let us suppose that there is no information about the frequency of each symbol in a communication process, therefore, we assume the uniform probability distribution $p(x) = \frac{1}{5}$. Then, the entropy is given by

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x) = -5\left(\frac{1}{5}\log\frac{1}{5}\right) = 2.3219\ bits. \tag{2.2}$$

Now, suppose that the receptor of a transmitted message receives the string $\{aaabcccdee\}$. In this case, the probability of each $x \in \mathcal{X}$ is $p(a) = 0.4, p(b) = 0.1, p(c) = 0.3, p(d) = 0.1$, and $p(e) = 0.1$. In consequence, the entropy value is

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x) = -0.4\log 0.4 - 3(0.1\log 0.1) - 0.3\log 0.3 = 2.0464\ bits, \tag{2.3}$$

which reflects the uncertainty reduction when we have previous information about the symbol probabilities in a message.

In general terms, the entropy is maximum when the probability distribution is uniform, and decreases as previous information is provided. This can be observed in Figure 2.1, which shows the entropy curve for a random variable X having two possible values $a, b \in \mathcal{X}$, each one with probabilities q and $1-q$, respectively. Observe how the entropy value is the highest when $q = \frac{1}{2}$, i.e., the distribution is uniform, and it is zero when $q = 0$ or $q = 1$, since at these points there is no uncertainty about the value of X.

The above description lead us to define an expression for the conditional entropy between two random variables.

Definition 2.1. The conditional entropy of two random variables X and Y is defined by

$$H(Y|X) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(y|x). \tag{2.4}$$

The expression in (2.4) allows us to formulate the next theorem.

Theorem 2.1. *The joint entropy of two random variables X and Y is defined as*

$$H(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(x,y) = H(X) + H(Y|X). \tag{2.5}$$

Proof.

$$H(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(x,y) \tag{2.6}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x)p(y|x) \tag{2.7}$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x), \tag{2.8}$$

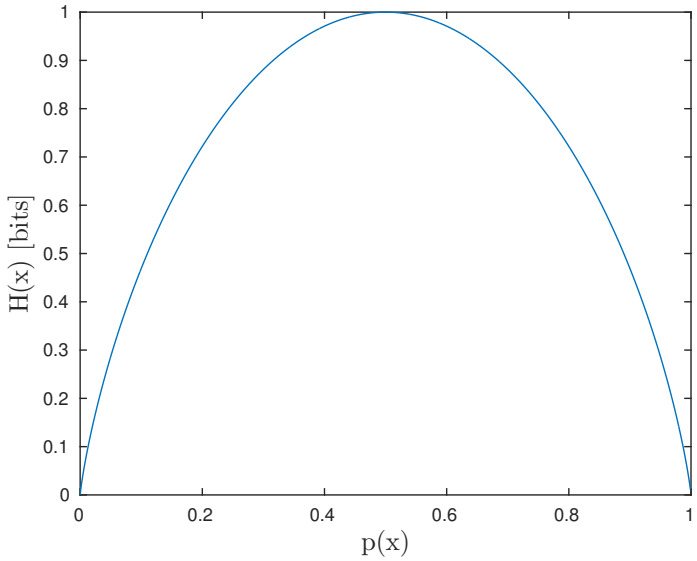


Figure 2.1: Entropy curve for a random variable X with two possible values $a, b \in \mathcal{X}$ having probabilities $p(a) = q$, and $p(b) = 1 - q$.

since $\sum_{y \in \mathcal{Y}} p(x,y) = p(x)$,

$$H(X,Y) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y|x), \tag{2.9}$$

and from (2.4), we have

$$H(X,Y) = H(X) + H(Y|X). \tag{2.10}$$

■

The above theorem tell us that the joint entropy is equal to the entropy of X plus the entropy of Y reduced because of the previous knowledge of X. Now, having into account the above definitions, let us define the mutual information between two random variables, which is one of the most relevant concepts of our research work.

Definition 2.2. The mutual information between two random variables X and Y is defined by

$$I(X;Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{2.11}$$

and is given in bits.

The expression in (2.11) can also be described through the Kullback-Leibler divergence, which is defined as a measure of the distance between the probability distributions

$p(x)$ and $q(x)$, and is given by

$$KL(p(x)||q(x)) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}. \tag{2.12}$$

Equation (2.12) is interpreted as the loss of information when a random variable X, originally described with a probability distribution $p(x)$, is described using a different distribution $q(x)$. Then, for the mutual information case, we have

$$I(X;Y) = KL(p(x,y)||p(x)p(y)). \tag{2.13}$$

In an entropy context, the mutual information represents the uncertainty reduction of Y when some cue about X is previously known. This is established through the next theorem.

Theorem 2.2 (Entropy and mutual information)**.** *For two random variables X and Y, we have that*

$$I(X;Y) = H(Y) - H(Y|X). \tag{2.14}$$

Proof. From (2.11) we have

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} && (2.15)\\ &= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{p(y|x)}{p(y)} && (2.16)\\ &= -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(y) + \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(y|x) && (2.17)\\ & && (2.18) \end{aligned}$$

since $\sum_{x\in\mathcal{X}} p(x,y) = p(y)$, then

$$I(X;Y) = -\sum_{y\in\mathcal{Y}} p(y)\log p(y) - \left(-\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(y|x)\right), \tag{2.19}$$

and using (2.4), we obtain

$$I(X;Y) = H(Y) - H(Y|X). \tag{2.20}$$

■

Once the concept of mutual information has been defined, we focus our interest in the minimum mutual information necessary to represent X by means of Y when they are separated by a distance commonly know as the *distortion measure*. This is described in the next section.

2.2 The Rate Distortion Function

The rate distortion function represents the minimum mutual information between two random variables involved in a setting having distortion. In order to describe this concept, let us assume a message interchange between a sender and a receiver through a communication channel, as depicted in Figure 2.2. The random variable X represents the sent message, while the random variable Y the received one. The noise source, intrinsic to the communication channel, generates the distortion that avoids to have X in the reception point [93]. The difference between the sent and the received messages is described through a distortion measure, which in our case is the *squared error distortion measure* defined by

$$\mathcal{L}(x,y) = (y-x)^2,\ x \in \mathcal{X},\ y \in \mathcal{Y}. \tag{2.21}$$

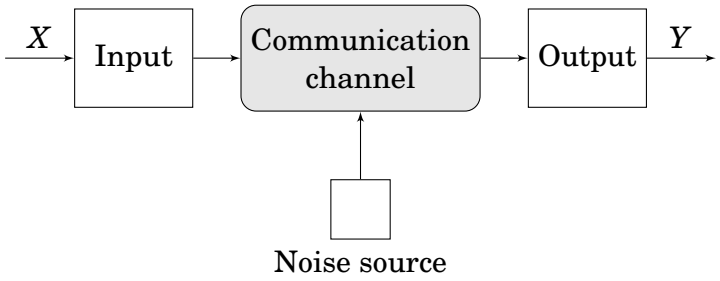


Figure 2.2: Message interchange between a sender and a receiver.

The expression in (2.21) determines an expected distortion value, named D, associated to a rate in bits, named R, necessary to have legibility of the sent message in the reception point. In this sense, the rate distortion function is defined as the next minimization problem[2]

$$\begin{aligned} R(D) = \quad & \underset{p(y|x)}{\text{minimize}} I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(y|x)}{p(y)} \\ & \text{subject to} \sum_{x,y} p(y|x)p(x)\mathcal{L}(x,y) \leq D \\ & \sum_{x,y} p(y|x) = 1 \\ & p(y|x) \geq 0, \end{aligned} \tag{2.22}$$

[2]Henceforth, for simplicity, we are going to use x and y to refer the $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively.

which, represented in the Lagrange multipliers form[3], lead us to the expression

$$\begin{aligned} R(D) = \underset{p(y|x)}{\text{minimize}} \Bigg[& \sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} \\ & - s(\sum_{x,y} p(y|x)p(x)\mathcal{L}(x,y) - D) \\ & + \sum_{x} \lambda_x(\sum_{y} p(y|x) - 1)\Bigg], \end{aligned} \tag{2.23}$$

where the inequality restriction $p(y|x) \geq 0$, is temporarily ignored. Now, making

$$\begin{aligned} J[p(y|x), p(y)] = & \sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} \\ & - s\sum_{x,y} p(y|x)p(x)\mathcal{L}(x,y) + \sum_{x} \lambda_x \sum_{y} p(y|x), \end{aligned} \tag{2.24}$$

we have

$$R(D) = sD - \lambda + \min_{p(y|x)} \min_{p(y)} J[p(y|x), p(y)], \tag{2.25}$$

where $\sum_x \lambda_x = \lambda$. Equation (2.25) is a double minimization problem tackled in two steps described in Appendix C, whose solutions are given by

$$p^*(y|x) = \frac{p(y)e^{s\mathcal{L}(x,y)}}{\sum_y p(y)e^{s\mathcal{L}(x,y)}}, \tag{2.26}$$

and

$$p^*(y) = \sum_x p(x)p(y|x), \tag{2.27}$$

which are the basis for the Blahut-Arimoto algorithm formulation [15], described here below.

2.2.1 The Blahut-Arimoto Algorithm

The Blahut-Arimoto algorithm calculates iteratively the $p^*(y|x)$ and $p^*(y)$ of (2.25), until a convergence condition is found, as depicted in Figure 2.3. This receives as input parameters an initial uniform distribution $p_o(y) = \frac{1}{|\mathcal{Y}|}$, the Lagrange multiplier $\mathfrak{s} \in \mathbb{R}^-$, and the previously known source distribution $p(x)$. The resulting outputs are the conditional distribution $p(y|x)$, i.e., the probability of a $y \in \mathcal{Y}$ for a given $x \in \mathcal{X}$, the expected distortion D, and its associated rate R, whose values are determined by the value of $\mathfrak{s}$, which we describe in detail hereafter.

[3]We use the Lagrange multiplier representation $\mathfrak{s}$ as an approximation to the symbol used in the literature related to the Blahut-Arimoto algorithm.

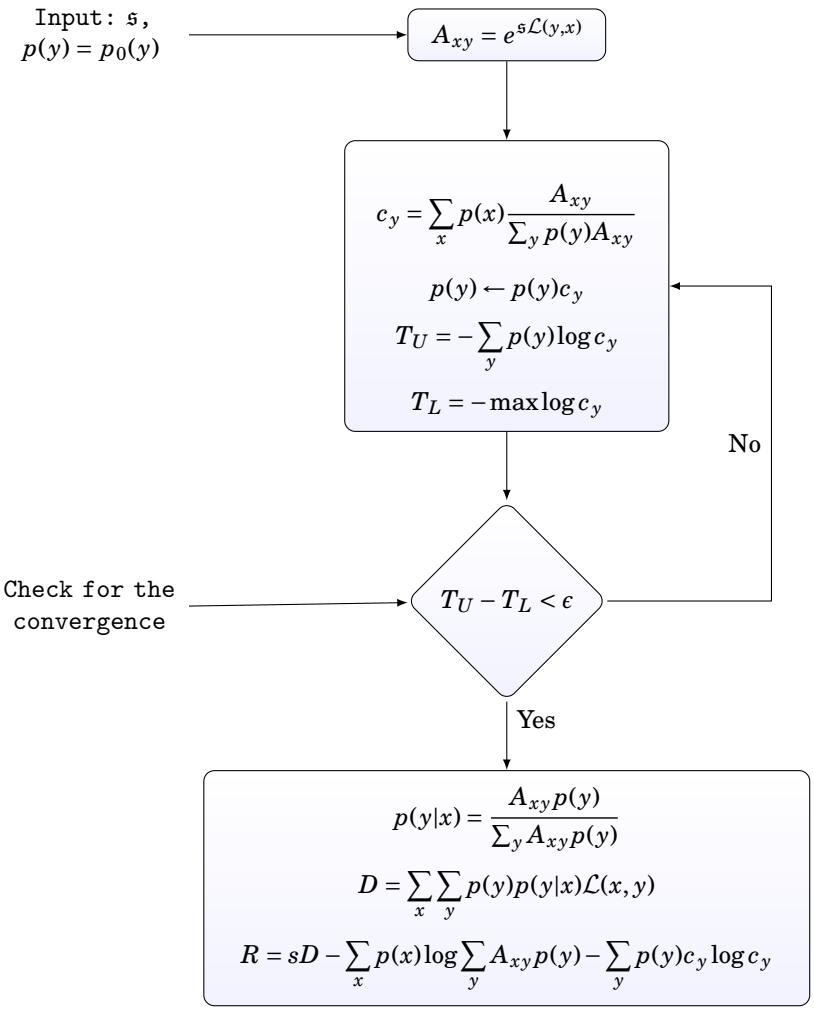


Figure 2.3: The Blahut-Arimoto algorithm.

2.2.1.1 The Parameter $\mathfrak{s}$

For simplicity, let us explain the relevance of the parameter $\mathfrak{s}$ by means of the following example. Consider a random variable X with $\mathcal{X} = \{x \in \mathbb{R} : -10 \leq x \leq 10\}$, and a Gaussian source distribution $p(x) = \mathcal{N}\left(\mu = 0,\ \sigma^2 = 3\right)$. For $-20 \leq \mathfrak{s} \leq -2$, the rate distortion function is the one shown in Figure 2.4. The allowed region is composed of the set of points $\{R, D\}$ that guarantee message legibility in a reception point, in spite of receiving a symbol $y \in \mathcal{Y}$ when the sent symbol in the transmission point was $x \in \mathcal{X}$. On the other hand, the set of points $\{R, D\}$ on the $R(D)$ curve determine the boundary at which the mutual information is minimum but enough to have legibility between the emitted and the received message, in other words, this set of points establish the limit at which the information is neither redundant nor misunderstood.

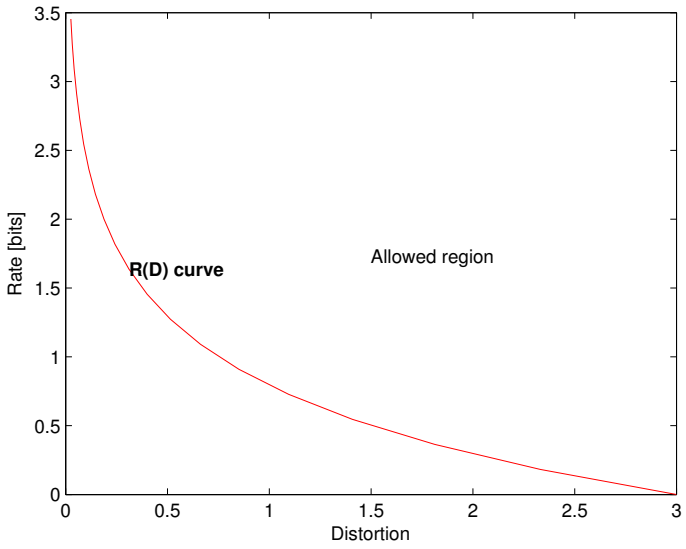


Figure 2.4: R(D) curve for $-20 \leq \mathfrak{s} \leq -2$ and $p(x) = \mathcal{N}\left(\mu = 0,\ \sigma^2 = 3\right)$

The relationship between the parameter $\mathfrak{s}$ and the $R(D)$ curve is visible if we recall (2.25), which can be rewritten in the linear form

$$R(D) = sD + b, \tag{2.28}$$

in other words, $\mathfrak{s}$ determines a negative slope in a $\{D,R\}$ point of the curve, being the lowest value associated to the highest rate and the lowest distortion, as shown in Figure 2.5, in which we can observe a lower distortion for a slope $s = \frac{-2}{\sigma^2}$ than for a slope $s = \frac{-1}{\sigma^2}$.

The parameter $\mathfrak{s}$ also affects the certainty of the conditional distribution $p(y|x)$. Figure 2.6 shows how the variance of $p(y|x = -5.15)$ changes depending on the value of $\mathfrak{s}$. Note how the variance decreases when $\mathfrak{s}$ is more negative, i.e., for a low distortion. In contrast, the variance increases for values of $\mathfrak{s}$ close to zero, i.e., when the distortion is higher.

At this point, it is noticeable that the selection of $\mathfrak{s}$ can define the level of distortion that we want to have for a specific application. In our approach, this parameter is important to establish an agent *rationality* level for the learning process of a multi-agent system. This will be described in detail in Chapter 3.

2.2.1.2 Calculation of the Source Probability p(x)

The source distribution $p(x)$ used as an input in the Blahut-Arimoto algorithm is not always known, which means that it must be obtained through a statistical procedure.

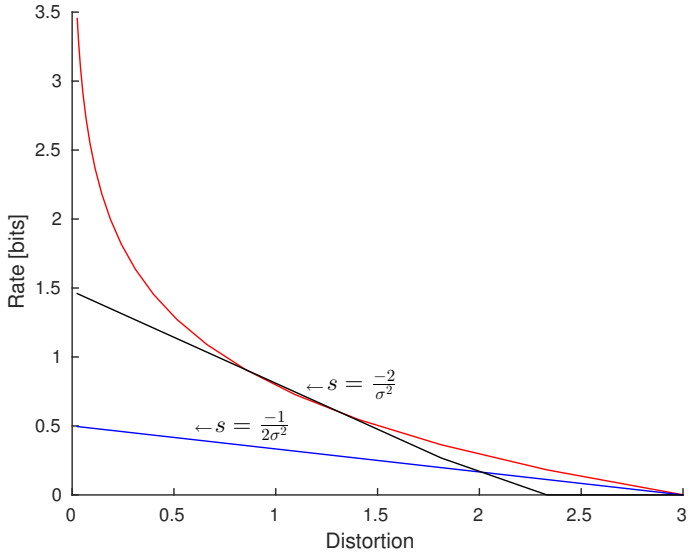


Figure 2.5: Relationship between the parameter s and the $R(D)$ curve.

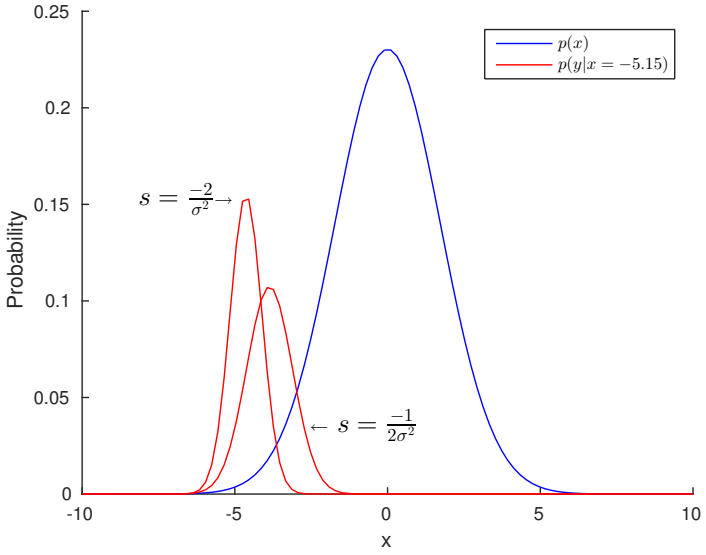


Figure 2.6: Relationship between the parameter s and $p(y|x)$.

In this sense, one of the most popular approaches to calculate unbiased probability distributions from previous knowledge exposed in the form of *distribution moments*, is the maximum entropy principle, whose expression is defined by

$$\begin{aligned} \underset{p(x)}{\text{maximize}} \quad & \sum_x p(x)\log p(x) \\ \text{subject to} \quad & \sum_x p(x)g_k(x) = F_k \\ & \sum_x p(x) = 1 \\ & p(x) \geq 0, \end{aligned} \tag{2.29}$$

where F_k is the expected value of a function $g_k(x)$. Using the Lagrange multipliers, the solution of this maximization problem is given by

$$J(p,\lambda) = -\sum_x p(x)\log p(x) + \lambda_0\left(1 - \sum_x p(x)\right) + \sum_k \lambda_k\left(F_k - \sum_x p(x)g_k(x)\right), \tag{2.30}$$

where

$$\frac{\partial J(p,\lambda)}{\partial p(x)} = -1 - \log p(x) - \lambda_0 - \sum_k \lambda_k g_k(x). \tag{2.31}$$

Setting the above term to zero we have

$$p(x) = \frac{1}{Z}e^{-\sum_k \lambda_k g_k(x)}, \tag{2.32}$$

where $Z = e^{1+\lambda_0}$. From the first constraint

$$1 = p(x) = \sum_x p(x) = \frac{1}{Z}\sum_x e^{-\sum_k \lambda_k g_k(x)}, \tag{2.33}$$

therefore

$$Z = \sum_x e^{-\sum_k \lambda_k g_k(x)}. \tag{2.34}$$

The above approach is highly useful to find an unbiased probability distribution $p(x)$ when the previous knowledge is present in form of distribution moments, such as the expected value. However, when this previous information is not present, the $p(x)$ distribution having maximum entropy can be found in the following equivalent Gaussian distribution with mean ($\mu = 0$), and covariance Σ [74]

$$p(x) = \frac{1}{(2\pi)^2|\Sigma|^{\frac{1}{2}}}e^{\frac{1}{2}x^T\Sigma^{-1}x}. \tag{2.35}$$

Then, assuming a zero mean, our problem is reduced to the calculation of the covariance Σ. Fortunately, we can take advantage of the popular kernel RBF (Radial Basis Function), which is defined as

$$c(x_n, x_m) = \alpha e^{-\gamma(x_n - x_m)^2}, \tag{2.36}$$

where $c(x_n, x_m)$ is the covariance between a pair of points x_n and x_m belonging to the observed data, named *training data*, while the parameters α and γ define the smoothness of the resultant distribution. This lead us to describe the Gaussian process regression approach, used in this work as a method to infer the *source distribution* describing the agent environment.

2.3 Non Parametric Learning

One of the most simple cases of a learning process is the fitting of a function $f(x)$ based on a set of known values, named the *training data*, to predict unknown values, named the *test data*, in a set of positions in x. In this regard, the use of linear parametric regressions with the form

$$y_n = f(x_n; w_0, w_1) = w_0 + w_1 x, \tag{2.37}$$

has been widely used, in which, the parameters w_n are calculated from an error minimization. In this sense, for a two parameters regression, and using the *squared loss function*[4], this minimization takes the form

$$\underset{w_o, w_1}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} (y_n - f(x_n; w_0, w_1))^2, \tag{2.38}$$

where the set $\mathbf{y} = [y_1, ..., y_N]$ represents the *training values* corresponding to the $\mathbf{x} = [x_1, ..., x_N]$ *training points*. This kind of model, in spite of being extended to non linear cases by means of the addition of parameters, becomes unmanageable as the parameter number increases. In this regard, the Gaussian Processes Regression (GPR), offers a non parametric alternative to address this kind of problems.

2.3.1 Gaussian Process Regression

The GPR consists of the definition of a posterior function f^*, from a prior function f satisfying a previously set of observed data, named the *training data*. In order to explain this, let us assume that we have a set of N *training points*, given by the set $\mathbf{x} = [x_1, ..., x_N]$, and their corresponding *training values*, given by $\mathbf{f} = [f_1, ..., f_N]$. In the same way, we have a set of M *testing points*, given by the set $\mathbf{x}^* = [x_1^*, ..., x_M^*]$, at which we want to predict the corresponding *testing values*, given by the set $\mathbf{f}^* = [f_1^*, ..., f_M^*]$.

[4]The squared loss function measures how close is a particular prediction model to the *training data*.

Since the GPR assumes that the function values at all the points (training and testing) follow a Gaussian density, the vectors $\mathbf{f}$ and $\mathbf{f}^*$ can be combined in a single vector given by

$$\mathbf{f_t} = \begin{pmatrix} \mathbf{f} \\ \mathbf{f}^* \end{pmatrix}, \tag{2.39}$$

which, also follows a Gaussian density [89] [86]. Assuming a zero mean, the prior distribution describing the whole model, is defined by

$$p(\mathbf{f_t}) = \mathcal{N}\left(0, \begin{bmatrix} C & C_* \\ C_*^T & C_{**} \end{bmatrix}\right), \tag{2.40}$$

where C is the covariance matrix for the training points, C_* is the cross-covariance between the training and test points, while C_{**} is the covariance matrix for the testing points, which are obtained by applying the kernel function defined in (2.36). On the other hand, since in a multivariate Gaussian context a subset of variables conditioned on the others is also Gaussian distributed, the expression

$$p(\mathbf{f}^*|\mathbf{f}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.41}$$

also represents a Gaussian distribution, where

$$\boldsymbol{\mu} = C_* C^{-1} \mathbf{f}, \tag{2.42}$$

and

$$\boldsymbol{\Sigma} = C_{**} - C_*^T C^{-1} C_*. \tag{2.43}$$

Finally, if we want to find the set of *testing values* $\mathbf{f}^*$, we use the Cholesky decomposition to find a J such that $\boldsymbol{\Sigma} = JJ^T$, and thus, $\mathbf{f}^* \approx \boldsymbol{\mu} + J\mathcal{N}(0, I)$. Let us to illustrate the above concepts with a simple example.

Example 2.3.1. Consider the Figure 2.7, which shows the set of training points

$$\mathbf{x} = [-8, -6, -4, -2, 0, 2, 4, 6, 8],$$

and their corresponding training values

$$\mathbf{f} = [0.54, 1.9, -2, 0.54, 0.43, -1.25, -0.6, 0.28, 3.6],$$

represented as red circles. Suppose that we want to predict the *testing values* at the set of *testing points* given by $\mathbf{x}^* = [-7, -5, -1, 3, 5]$, depicted as the dashed vertical lines.

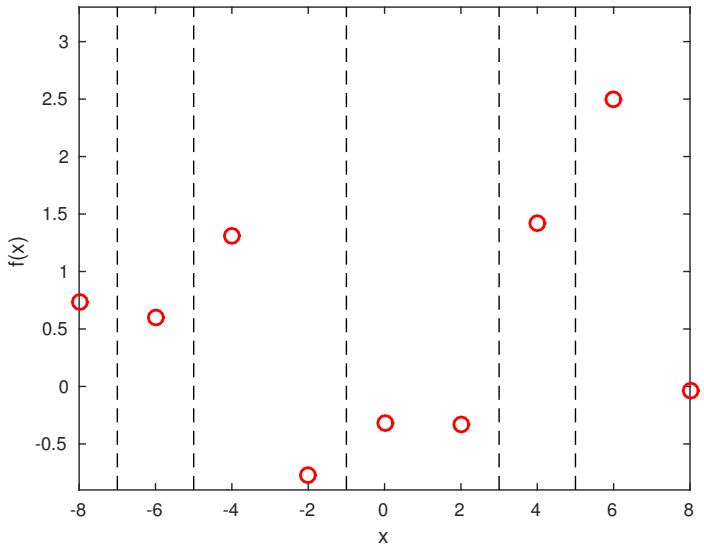


Figure 2.7: The sets of training points, training values, and the test points.

By means of a MATLAB script, we calculate the set of mean values

$$\boldsymbol{\mu} = [-0.17, 0.6, 0.84, 1.12, 0.52],$$

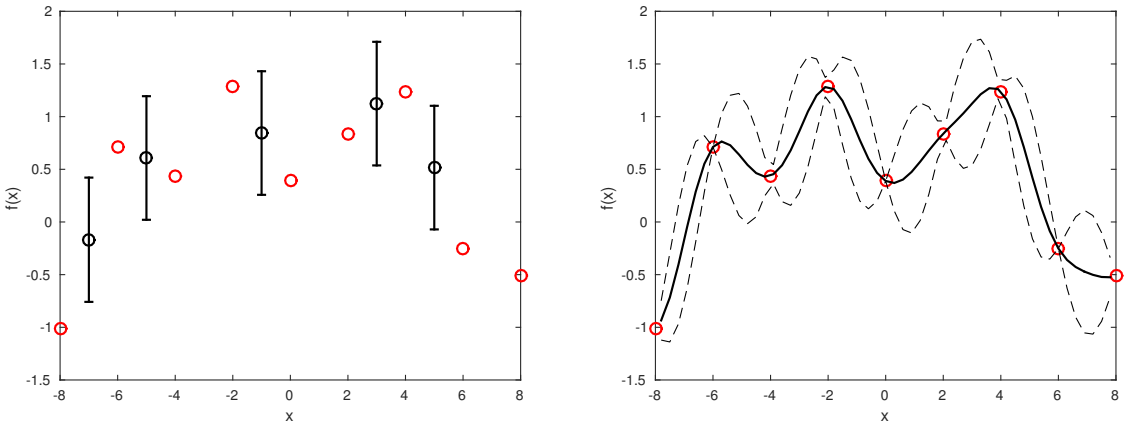
and the set of covariances

$$\boldsymbol{\Sigma} = [0.35, 0.34, 0.34, 0.34, 0.34].$$

These results are shown in Figure 2.8. In Figure 2.8a, we observe the mean values, represented as black dots, and their corresponding standard deviation. In Figure 2.8b we show the resulting function f^*, depicted as a black curve.

2.4 Game Theory

Game theory has become a powerful tool for multi agent learning systems, since its inclusion in distributed optimization approaches, has permitted the maximization or minimization of objective functions in order to reach a Nash equilibrium, at which each agent has the best utility, while the others remain in their current state. In this section, we describe the most relevant concepts related to game theory, emphasizing in potential games, which is one of the key elements to formulate our learning model.



(a) The testing values and their corresponding standard deviations.

(b) The resultant predicted function.

Figure 2.8: The GPR prediction result.

2.4.1 Strategic Games

The concept of strategic games was introduced by Neumann and Morgenstern in their seminal work *"Theory of Games and Economic Behavior"* [108], whose initial economic focus, has been derived to others sciences such as mathematics, biology, sociology, physics, and engineering, among others. In this theory, the game players, henceforth referred to as agents, interact in a setting at which the individual welfare does not only depend on the individual selected strategy, but also on the strategies assumed by the others.

In a formal way, a *strategic game* defined by $\mathcal{G}$, is composed of a set of N selfish agents, in which an agent i, $i \in \{1,...,N\}$, has a finite set of strategies or actions S_i. In this way, the set of all the system strategy configurations, or strategy profiles, is given by $S = \prod_1^N S_i = S_1 \times,...,\times S_N$. This means that in a determined time step, each agent can choose a strategy $s_i \in S_i$, which results in a system strategy profile $\mathbf{s} = (s_1,...,s_N), \mathbf{s} \in S$, at which the utility of the agent i is $U_i(\mathbf{s}) : S_1 \times,...,\times S_N \rightarrow \mathbb{R}$. Adopting the standard game theoretic notation, we use s_{-i} to refer the set of actions assumed by the agents different to the agent i. With the above concepts in mind, let us define the *Nash equilibrium*.

Definition 2.3 (The Nash equilibrium)**.** An action profile $s^* = (s_i^*, s_{-i}^*)$ is a Nash equilibrium, if $\forall i \in N$ and $\forall s_i \in S_i$

$$U_i(s^*) \geq U_i(s_i, s_{-i}^*). \tag{2.44}$$

This means that in the strategy profile s^*, an agent has no incentive to adopt a unilateral deviation [123] [8].

As we have mentioned, the agent strategy selection depends on its utility, and the actions taken by the others. This process, commonly known as the *dynamics*, is an evolving mechanism that defines the action selection of the agents, through which the multi-agent system eventually finds the Nash equilibrium. In this sense, many types of dynamics have been proposed, being the most popular the *Logit dynamics*, *replicator dynamics*, *Smith dynamics*, among others [90]. In this work, we are especially concerned with the *Logit dynamics*, which is a noisy version of the best response dynamics. However, the replicator dynamics is also considered in one of the applications of the proposed learning model, as we will see in Appendix A.

2.4.2 The Logit Dynamics

In a simple form, in a best response dynamics setting, the agents take turns to assume the most profitable action against the selected actions by the other agents. According to [17], for a given set of best responses denoted as $\mathcal{M}(s_i, s_{-i})$, where (s_i, s_{-i}) is the current strategy profile, the best response choice for an agent i, is determined by the probability

$$p(s_k|s_{-i}) = \begin{cases} \frac{1}{|\mathcal{M}(s_i,s_{-i})|} & \text{if } s_k \in \mathcal{M}(s_i, s_{-i}) \\ 0 & \text{otherwise.} \end{cases} \tag{2.45}$$

This kind of dynamics assume agent rationality, i.e., agents have complete knowledge about the strategies followed by the others, which is not always possible in a realistic context. In this regard, the *Logit dynamics* tackles the knowledge limitation through a rationality measure β, used to define the rules followed by an agent in order to choose its strategies [17], according to the expression

$$p(s_k|s_{-i}) = \frac{e^{\beta U_i(s_k, s_{-i})}}{\sum_{s_i \in \mathcal{S}_i} e^{\beta U_i(s_i, s_{-i})}}, \tag{2.46}$$

where $s_k \in S_i$.

The above expression has the form of a Boltzmann distribution, in which the rationality measure β is similar to the inverse temperature. According to [8], the *Logit dynamics* is a *noisy* best-response dynamics, where the noise level is determined by β. In this sense, for $\beta = 0$, the decisions are made under the highest noise condition, and the strategy selection follows a uniform distribution, i.e., each strategy has the same probability to be chosen. On the other hand, when $\beta \to \infty$, the agent tends to choose the strategy that corresponds to its best response. Independently of the initial strategy profile, the *Logit* dynamics converges to a stationary distribution after a number of steps given by the

rationality level. This means that, after a sufficiently large time, the probability of find the system in a specific strategy profile remains unchanged, and there is no a strategy that improves an agent benefit when the others remain static, which constitutes a *Nash equilibrium*.

2.4.3 Replicator Dynamics

Before to explain the concept of replicator dynamics, let us explain some relevant concepts related to evolutionary games.

2.4.3.1 Evolutionary Game Theory

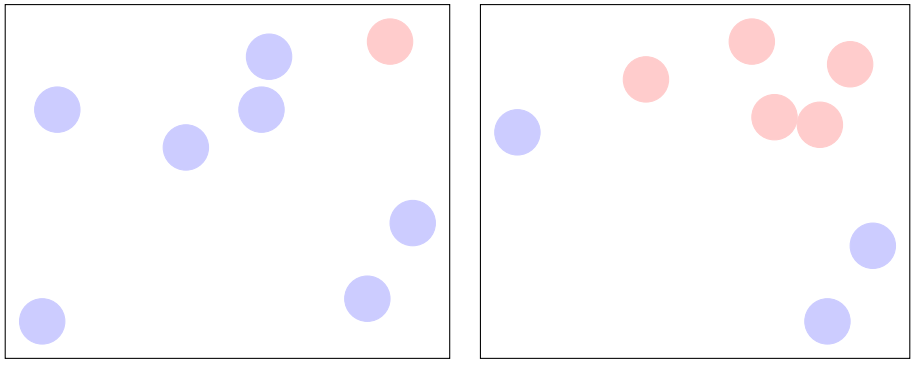
In evolutionary games the utility of each individual or player is interpreted as a fitness value that depends on the frequency or proportion of a phenotype in a population [77]. In contrast with strategic games, in evolutionary games the players do not make decisions based on rationality, but in the acquired information through the interaction. In this sense, the individuals find out the payoff of their peers, and emulate the strategies followed by the ones having the highest rewards. This process is similar to the natural selection, in which the strategies having good rewards reproduce faster, whereas those strategies having the poorest incentives tend to disappear. This is illustrated in Figure 2.9. First, in Figure 2.9a we show a player in red to represent the population proportion following the strategy with the highest payoff. Second, in Figure 2.9b we show how the players that initially follow the blue strategy having lower payoff, begin to change to the red one after the interaction with the first player. The increment of the number of players following the red strategy causes its payoff decrease. Conversely, the strategy in blue begins to be worthy.

In a formal way, for a pair of strategies A and B, the respective frequencies are denoted by x_A and x_B. Then, the population composition is given by $x = (x_A, x_B)$, and the corresponding fitness are $f_A(x)$ and $f_B(x)$, which lead us to express the selection dynamics as

$$\dot{x}_A = x_A \left[f_A(x) - \overline{f} \right] \tag{2.47}$$

$$\dot{x}_B = x_B \left[f_B(x) - \overline{f} \right], \tag{2.48}$$

where $\overline{f} = x_A f_A(x) + x_B f_B(x)$ is the average fitness, and $x_A + x_B = 1$. This last condition,



(a) The fitness of the strategy A (in red) is higher.

(b) The fitness of the strategy B (in blue) begins to increase.

Figure 2.9: Population dependency of fitness.

allows us to make $x = x_A$ and $x_B = 1 - x$, therefore we have that

$$\overline{f} = xf_A(x) + (1-x)f_B(x) \tag{2.49}$$

$$= xf_A(x) + f_B - xf_B. \tag{2.50}$$

Replacing (2.50) in (2.47), we obtain

$$\dot{x} = x(1-x)[f_A(x) - f_B(x)], \tag{2.51}$$

which is a differential equation with equilibrium at $x = 0$, $x = 1$, and all the values of x satisfying $f_A(x) = f_B(x)$.

2.4.3.2 The Replicator Dynamics Equation

Consider a set of n strategies and a $n \times n$ matrix, denoted as the payoff matrix, whose component a_{ij} represents the payoff associated to the interaction between the strategies $i, j \in n$ [77]. Then, the expected payoff of the strategy i is given by

$$f_i = \sum_{j=1}^{n} x_j a_{ij}. \tag{2.52}$$

Therefore the average fitness is

$$\overline{f} = \sum_{i=1}^{n} x_i f_i. \tag{2.53}$$

Then, the resulting replicator dynamics equation is given by

$$\dot{x} = x_i \left[f_i(x) - \overline{f} \right]. \tag{2.54}$$

Example 2.4.1 (The prisoners dilemma)**.** In this popular example, two crooks have been captured by the police, who has offered two options. The first one is to confess, and the second one is to keep quiet. If one of the crooks confess that both committed the crime, he will be set free and the other will spend 5 years in jail. If both confess, they will get a sentence of 3 years. If neither confess, they have to spend 1 year in jail. Let us denote by A the strategy of keep quiet, and by B the strategy of confess. The corresponding payoff matrix is shown in Table 2.1. The pairs (x, y) on each cell, represent the payoff of player X and Y, respectively, for a given combination of strategies A and B. Observing the payoff values, we can deduce that when both players choose A, just the player X improves its payoff for changing to strategy B. On the other hand, if both payers have chosen B, neither player will improve its utility for changing to A. Therefore, the strategy B is a *Nash equilibrium* [113] [77]. In terms of evolutionary games, this Nash equilibrium implies an evolutionarily stable strategy (ESS), i.e, a strategy that will offer the best payoff in spite of the appearance of other strategies.

		Player Y	
		C	K
Player X	C	$(3,3)$	$(0,5)$
	K	$(5,0)$	$(1,1)$

Table 2.1: Payoff matrix for the prisoner dilemma game.

If we apply the replicator dynamics equation of (2.54), we can observe the evolution of both strategies. This is shown in Figure 2.10. Observe how the frequency of strategy B (x_B) overcomes the frequency of A (x_A), no matter the initial conditions.

2.4.4 Potential Games

A strategic game is a potential game if exists a potential function $\phi(\mathbf{s}) : S \to \mathbb{R}, \mathbf{s} \in S$ that reflects the individual utility change when each agent unilaterally assume a new strategy, no matter which one caused it. In this way, the local optima of $\phi(\mathbf{s})$ can be used to find the set of pure Nash equilibrium of the whole system. We can find many types of potential games in literature [73] [109]. Hereafter, let us describe some of them.

Definition 2.4 (Ordinal potential game)**.** An ordinal potential game is defined by

$$U_i(s_k, s_{-i}) - U_i(s_l, s_{-i}) > 0 \Longleftrightarrow \phi(s_k, s_{-i}) - \phi(s_l, s_{-i}) > 0, \tag{2.55}$$

where $s_k, s_l \in S_i$.

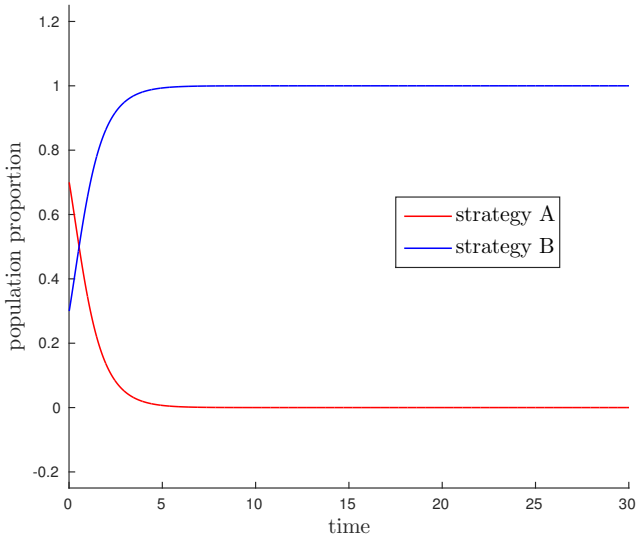


Figure 2.10: The evolution of strategies A and B.

Definition 2.5 (Exact potential game)**.** An exact potential game is defined by

$$U_i(s_k, s_{-i}) - U_i(s_l, s_{-i}) = \phi(s_k, s_{-i}) - \phi(s_l, s_{-i}), \tag{2.56}$$

where $s_k, s_l \in S_i$.

Definition 2.6 (Weighted potential game)**.** A weighted potential game is defined by

$$U_i(s_k, s_{-i}) - U_i(s_l, s_{-i}) = \frac{\phi(s_k, s_{-i}) - \phi(s_l, s_{-i})}{w_i}, \tag{2.57}$$

where $s_k, s_l \in S_i$, and $w_i \in \mathbb{R}^+$.

Example 2.4.2. Considering the prisoners dilemma of example 2.4.1, we can define the potential function shown in Table 2.2. This is an exact potential function since for player X we have

$$U_X(A, A) - U_X(B, A) = 5 - 3 = 2, \tag{2.58}$$

which is equal to

$$\phi(A, A) - \phi(B, A) = 2 - 0 = 2. \tag{2.59}$$

The above proof can be demonstrated for any strategy combination and player.

In this regard, the agents constituting a multi-agent system have as common goal the maximization (minimization) problem

$$\begin{aligned} \max_{\mathbf{s}} \quad & \phi(\mathbf{s}) \\ \text{s.t.} \quad & \mathbf{s} \in S. \end{aligned} \tag{2.60}$$

		Player Y	
		A	B
Player X	A	0	2
	B	2	1

Table 2.2: Potential function for the prisoner dilemma game.

Additionally to the exposed variety and relative simplicity, one of the main advantages of potential games is the existence of at least one action profile that guarantees a Nash equilibrium, which is reached in our case, through the *Logit dynamics*. However, in most of the cases, the potential function definition could become a challenging affair. The most straightforward method is to make $U_i(\mathbf{s}) = \phi(\mathbf{s}), s \in S$, which is applicable in systems having small number of agents, since it requires that each agent has a complete information about the payoffs obtained by the others, due to its utility depends directly on the potential function. In this sense, approaches such as wonderful life utility and the Shapley value, have been proposed in [116] and [7] respectively, which fit well in incomplete information contexts [101]. In our work, we define a *distortion based potential function*, that exploits the rate distortion function characteristics, defined in Section 2.22. We will describe it in detail in Chapter 4.

CHAPTER 3

INFORMATION-BASED RATIONALITY

The multi-agent learning model proposed in this work combines the predictive capabilities of GPR, the minimization of the mutual information, obtained through the rate distortion function, and the qualities of potential games in terms of distributed optimization. In this chapter, we firstly establish a multi-agent setting, in which each agent infers its environment by means of the GPR approach, using as *training data* the information provided by the agents belonging to its neighborhood. Once this environment is modeled through a probability distribution, it is used as the source distribution of the Blahut-Arimoto algorithm, which allows us to obtain the rate and distortion values associated to the parameter $\mathfrak{s}$, which in our case, acts as a *rationality measure* having a maximum and a minimum determined by the rate distortion function. In this regard, the *maximum rationality* value establishes the point from which there is no an improvement for the agent in terms of environment understanding, while the *minimum rationality* value defines the border at which the distortion about the environment is maximum but enough to understand it. Both rationality levels determine the agent behavior, since they can define the equilibrium deviation, and the convergence time, as we will show in Chapter 4, where the potential game approach is included in our model.

3.1 The Multi-Agent Environment

In order to describe the multi-agent environment, let us consider the configuration shown in Figure 3.1, which contains a set of N mobile sensing agents indexed by $i \in$

$\{1,...,N\}$, deployed in a spatial field $\Omega \in \mathbb{R}^2$, having the positions $S = \{s_1,...,s_N\}, S \in \Omega$, and sensing measurements $V = \{v_1,...,v_N\}$[1]. The *training set* for an agent i combines its own position s_i and measurement v_i, with the positions $s_j's$ and measurements $v_j's$ of its neighborhood, defined by $\mathcal{N}_i = \{j \in N : ||s_i - s_j|| \leq r_c\}$, where r_c is the connection radius. The environment of the agent i is composed of the set of *testing points* $Z_i = \{z_1,...,z_M\}$, and the corresponding set of *testing values* $W_i = \{w_1,...,w_M\}$, which are inferred using GPR. As we have mentioned in Section 2.3.1, the GPR prediction process give us a pair (μ_m, Σ_m) for each point $m = \{1,...,M\}$ of the *testing set*, in other words, the environment for an agent i, can be described by a set of Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This lead us to the next definition.

Definition 3.1 (The agent environment)**.** The environment of the agent i is given by

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} \mathcal{N}(\mu_1, \Sigma_1) \\ . \\ . \\ . \\ \mathcal{N}(\mu_M, \Sigma_M) \end{bmatrix}, \tag{3.1}$$

where μ_m and Σ_m are the predicted mean and variance in the testing point $m \in \{1,...,M\}$, respectively.

The set of locations Z_i constituting the environment, becomes a set of possible actions to take for the agent, and defines its movement within the spatial field.

Definition 3.2 (The agent action set)**.** The set of locations $Z_i = \{z_1,...,z_M\}$ determines the set of possible actions to be taken for an agent i, named *the agent action set*.

In this sense, each agent chooses the action offering the best utility at each time step, and in this way, it moves within the spatial field towards an equilibrium point, as we will describe in Chapter 4.

The set of distributions given in (3.1) determines the source distribution $p(x)$ necessary to calculate the rate distortion function through the Blahut-Arimoto algorithm.

Remark 3.1 (Blahut-Arimoto source distribution)**.** *The source distribution $p(x)$, necessary to calculate the rate distortion function $R(D)$, and consequently, the minimum*

[1]Although the model is defined in $\mathbb{R}^2$, it can be extended to $\in \mathbb{R}^3$

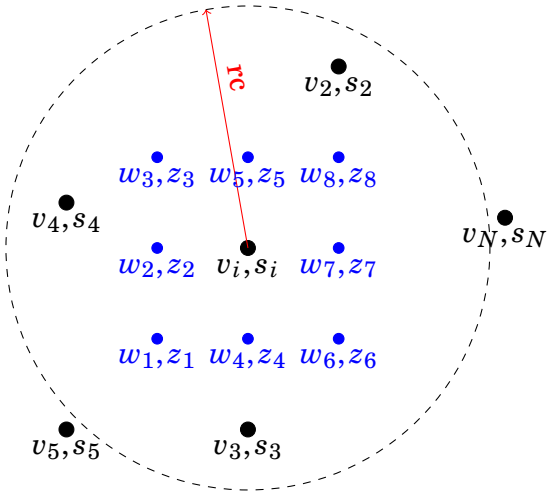


Figure 3.1: Multi-agent model setting.

mutual information between the agent i and its environment, is given by

$$\mathbf{p}(\mathbf{x}) = \begin{bmatrix} p(x_1) \\ . \\ . \\ . \\ p(x_M) \end{bmatrix} = \begin{bmatrix} \mathcal{N}(\mu_1, \Sigma_1) \\ . \\ . \\ . \\ \mathcal{N}(\mu_M, \Sigma_M) \end{bmatrix}. \tag{3.2}$$

With this in mind, we can proceed to apply the Blahut-Arimoto algorithm, which provides the pair rate and distortion (R, D), for a given value of the parameter $\mathfrak{s}$, which, as we have mentioned, is the factor that set the rationality level of or approach. This is described below.

3.1.1 The Lowest Rationality

In Section 2.2.1.1, we have shown the relationship between the parameter $\mathfrak{s}$ of the Blahut-Arimoto algorithm, with the rate distortion function, and how it determines the points (R, D) on it. Now, let us find the value of $\mathfrak{s}$ at which an agent has the maximum allowed distortion about its environment, but enough information to comprehend it. At this point, we have the lowest rationality, which is described by means of the next theorem.

Theorem 3.1 (The lowest rationality)**.** *Consider the environment for the agent i, described by the vector of Gaussian distributions* $\mathbf{p}(\mathbf{x})$ *given by (3.2), and the corresponding set of values for the parameter* $\mathfrak{s}$ *expressed as*

$$\mathfrak{s} = \left[\mathfrak{s}_1, \quad . \quad . \quad . \quad , \mathfrak{s}_M\right]^T. \tag{3.3}$$

Then, the upper boundary condition for $\mathfrak{s}$ *is*

$$\mathfrak{s} \leq -\frac{1}{2\boldsymbol{\Sigma}}. \tag{3.4}$$

Proof. According to [27], the rate distortion function for a Gaussian source $\mathcal{N}(\mu, \sigma^2)$ and squared-error distortion is given by the expression

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2. \end{cases} \tag{3.5}$$

For the case of our work, this Gaussian source is given by $\mathbf{p}(\mathbf{x})$. Then, the rate distortion function between an agent i and the environment point $m \in \{1, ..., M\}$, for $0 \leq D_m \leq \Sigma_m$, is defined by

$$R(D_m) = \frac{1}{2}\log\left(\frac{\Sigma_m}{D_m}\right). \tag{3.6}$$

Recalling that the slope in a point (D, R) of the function $R(D)$ is given by the parameter $\mathfrak{s}$, we have from (3.6) that

$$\mathfrak{s}_m = \frac{dR(D_m)}{dD_m} = -\frac{1}{2D_m}. \tag{3.7}$$

Then, for the upper distortion limit $D_m = \Sigma_m$, we have

$$\mathfrak{s}_m \leq -\frac{1}{2\Sigma_m}. \tag{3.8}$$

Therefore, considering all the points belonging to the agent environment, we have that

$$\mathfrak{s} \leq -\frac{1}{2\boldsymbol{\Sigma}}. \tag{3.9}$$

■

Remark 3.2 (Exploratory behavior)**.** *For the lowest rationality defined in (3.4), the agents behave in a exploratory form, since they tend to avoid the environment cues determined by their neighbors.*

The above result gains relevance in a setting where the field exploration is required, since the agent movement through uncorrelated regions, could promote the discovering of wealthy locations. Hereafter, we describe the set of steps necessary to establish the *maximum rationality* value.

3.1.2 The Highest Rationality

In the previous section, we have found the value of the parameter $\mathfrak{s}$ at which an agent has the minimum information about the environment, but enough to understand it, which is determined by the limits imposed by the rate distortion function. Now, we are going to find the value of $\mathfrak{s}$ at which an agent has the maximum information about its environment, in other words, *the highest rationality*. If we think in the distortion reduction as a measure of utility, the highest rationality point defines a limit at which an agent does not improve its benefit for decreasing the distortion or increasing the rate, as we will shown in Chapter 4. In other words, there is a distortion level at which the agent understands the environment in the same way as if it were zero. In this approach, we are going to employ the conditional distribution $p(y|x)$, resulting from the mutual information minimization, as a measure of the understanding about the environment for each agent. This lead us to the next definition.

Definition 3.3 (Agent environment understanding)**.** Let $p(y)$ be the distribution describing the agent behavior, and let $\mathbf{p(x)}$, defined in (3.2), be the distribution describing the agent environment. Then, the agent understanding about the environment is given by the set of conditional distributions

$$\mathbf{p(y|x)} = \begin{bmatrix} p(y|x_1) \\ . \\ . \\ . \\ p(y|x_M) \end{bmatrix}, \tag{3.10}$$

where $p(y|x_m) = \frac{A_{xy}p(y)}{\sum_y A_{xy}p(y)}$, and $A_{xy} = e^{\mathfrak{s}\mathcal{L}(y,x_m)}$, as stated in the Blahut-Arimoto algorithm described in Section 2.2.1.

In this sense, there is a value for the parameter $\mathfrak{s}$ at which the agent information has the highest similitude to the environment, and it occurs when the distributions $\mathbf{p(x)}$ and $\mathbf{p(y|x)}$ have the lowest distance or, in other words, the minimum Kullback-Leibler divergence. To find it, let us begin describing, through Lemmas 3.1 and 3.2, respectively, the matrix A_{xy}, and the distribution $p(y)$ in terms of the distortion, which, for the set of points constituting the agent action set, is given by the vector

$$\mathbf{D} = \begin{bmatrix} D_1, & . & . & . & ,D_M \end{bmatrix}^T. \tag{3.11}$$

Lemma 3.1. *Let $0 \leq D_m \leq \Sigma_m$, with $m \in \{1,...,M\}$. Then, for a given $x = x_m$ and $y \in \mathcal{Y}$, A_{xy} is exponentially decreasing in D_m, and is given by*[2,3]

$$A_{xy} = e^{-\frac{(y-x_m)^2}{2D_m}}. \tag{3.12}$$

Proof. From the Blahut-Arimoto algorithm we have $A_{xy} = e^{\mathfrak{s}_m(y-x_m)^2}$, and from Theorem 3.1 we have $\mathfrak{s}_m = \frac{-1}{2D_m}$.

Then

$$A_{xy} = e^{-\frac{(y-x_m)^2}{2D_m}}. \tag{3.13}$$

■

Lemma 3.2. *Let $0 \leq D_m \leq \Sigma_m$, with $m \in \{1,...,M\}$. Then, for a given $x = x_m$ and $y \in \mathcal{Y}$, $p(y)$ is an exponentially increasing function in D_m, and is given by*

$$p(y) = e^{-\frac{(y-x_m)^2}{2(\Sigma_m - D_m)}}. \tag{3.14}$$

Proof. According to [27], the mutual information of two correlated Gaussian random variables is given by

$$I(X,Y) = -\frac{1}{2}\log(1-\rho^2), \tag{3.15}$$

where ρ is the correlation coefficient between X and Y. Equating this expression with (3.6), we have

$$\frac{1}{2}\log\frac{\Sigma_m}{D_m} = -\frac{1}{2}\log(1-\rho_m^2). \tag{3.16}$$

Therefore

$$\Sigma_m = \frac{D_m}{1-\rho_m^2}. \tag{3.17}$$

On the other hand, from the Blahut-Arimoto algorithm we have[4]

$$p(y|x) = A_{xy}p(y). \tag{3.18}$$

Since, $p(x)$ and $p(y|x)$ are Gaussian, then $p(y)$ is also Gaussian. Therefore

$$p(y|x) = e^{-\frac{(y-x_m)^2}{2D_m}} e^{-\frac{(y-x_m)^2}{2\sigma_{y_m}^2}}. \tag{3.19}$$

[2]The alphabet $\mathcal{Y}$ represents a set of possible measurement values taken by the agent.

[3]To facilitate the calculation we ignore the normalization term $\frac{1}{\sum_y A_{xy}}$.

[4]To facilitate the calculation we ignore the normalization term $\frac{1}{\sum_y A_{xy}p(y)}$.

The conditional Gaussian distribution $p(y|x)$ can be expressed as

$$p(y|x) = \mathcal{N}\left(\mu_{y_m} + \frac{\rho_m \sqrt{\Sigma_m \sigma_{y_m}^2}}{\Sigma_m}(x_m - \mu_m), \sigma_{y_m}^2 - \frac{\rho_m^2 \sigma_{y_m}^2 \Sigma_m}{\Sigma_m}\right). \tag{3.20}$$

Hence

$$e^{-\frac{(y-x_m)^2}{2\left(\sigma_{y_m}^2 - \frac{\rho_m^2 \sigma_{y_m}^2 \Sigma_m}{\Sigma_m}\right)}} = e^{-\frac{(y-x_m)^2}{2D_m}} e^{-\frac{(y-x_m)^2}{2\sigma_{y_m}^2}} \tag{3.21}$$

and $\sigma_{y_m}^2 = \frac{D_m \rho_m^2}{1-\rho_m^2}$.

Now, using (3.17), we have

$$\Sigma_m - D_m = \frac{D_m}{1-\rho_m^2} - D_m = \frac{D_m \rho_m^2}{1-\rho_m^2} = \sigma_{y_m}^2. \tag{3.22}$$

Then

$$p(y) = e^{-\frac{(y-x_m)^2}{2(\Sigma_m - D_m)}}. \tag{3.23}$$

■

So far, we have defined $p(y)$ and A_{xy} as functions of the distortion. Now, using (3.18), Lemma 3.1 and 3.2, the conditional distribution $p(y|x)$ can also be given in terms of the distortion. This is established in the following Lemma.

Lemma 3.3. *Let $0 \leq D_m \leq \Sigma_m$, with $m \in \{1,...,M\}$. Then, for a given $x = x_m$ and $y \in \mathcal{Y}$, $p(y|x)$ has the expression*

$$p(y|x) = e^{\frac{-(y-x_m)^2}{\frac{2D_m}{\Sigma_m}(\Sigma_m - D_m)}}. \tag{3.24}$$

Proof. According to the Blahut-Arimoto algorithm, $p(y|x) = A_{xy} p(y)$. Then, using Lemma 3.1 and Lemma 3.2 we have

$$p(y|x) = e^{\frac{-(y-x_m)^2}{2D_m}} e^{\frac{-(y-x_m)^2}{2(\Sigma_m - D_m)}}, \tag{3.25}$$

so that

$$p(y|x) = e^{\frac{-(y-x_m)^2}{\frac{2D_m}{\Sigma_m}(\Sigma_m - D_m)}}. \tag{3.26}$$

Then, the resultant covariance of the Gaussian distribution $p(y|x)$ is given by

$$\sigma_{(y|x)_m}^2 = \frac{D_m}{\Sigma_m}(\Sigma_m - D_m). \tag{3.27}$$

■

Now, using (3.27), we can obtain a relationship between the distribution $p(x)$ describing the environment, and the distribution $p(y|x)$, which, as we have defined, describes the agent understanding about the environment. This is formulated by means of the next theorem.

Theorem 3.2 (The highest rationality)**.** *Let $KL[p(x)||p(y|x)]$ be the distance between the distribution describing the environment and the distribution describing the agent understanding about the environment. Then, the agent rationality is the highest when $KL[p(x)||p(y|x)]$ is minimized, and this occurs for*

$$\mathfrak{s} = \frac{-1}{\mathbf{\Sigma}}. \tag{3.28}$$

Proof. The similitude between the conditional distribution $p(y|x)$ and the distribution $p(x)$, which describes the environment, can be measured through the Kullback-Leibler divergence, which for a pair of Gaussian distributions as in this case, has the expression (see Appendix B)[5]

$$KL[p(x)||p(y|x)] = \frac{1}{2}\log\frac{\sigma_{(y|x)_m^2}}{\Sigma_m} + \frac{\Sigma_m}{2\sigma_{(y|x)_m^2}} - \frac{1}{2}. \tag{3.29}$$

From Lemma 3.3 we have

$$\sigma_{(y|x)_m^2} = \frac{D_m}{\Sigma_m}(\Sigma_m - D_m). \tag{3.30}$$

Then

$$\begin{aligned} KL[p(x)||p(y|x)] = \frac{1}{2}\log\frac{\frac{D_m}{\Sigma_m}(\Sigma_m - D_m)}{\Sigma_m} \\ + \frac{\Sigma_m}{2\frac{D_m}{\Sigma_m}(\Sigma_m - D_m)} - \frac{1}{2}. \end{aligned} \tag{3.31}$$

Now, the value of D_m for a minimum distance between both distributions is found solving the minimization problem

$$\begin{aligned} \underset{D_m}{\text{minimize}} \quad & -KL[p(x)||p(y|x)] \\ \text{subject to} \quad & D_m \leq \Sigma_m \\ & -D_m \leq 0. \end{aligned} \tag{3.32}$$

Using the Lagrange multipliers method we have

$$\frac{(\Sigma_m - 2D_m)}{2}\left[\frac{D_m(\Sigma_m^* - D_m) - (\Sigma_m^*)^2}{D_m^2(\Sigma_m - D_m)}\right] - \lambda_1\Sigma_m - \lambda_2 = 0, \tag{3.33}$$

[5]We use $\mu_{y_m} = \mu_m$, i.e., the Blahut-Arimoto algorithm finds the conditional distribution $p(y|x = \mu_m)$ for each point $m \in \{1,...,M\}$ of the agent environment.

where $\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_1(D_m - \Sigma_m) = 0, D_m\lambda_2 = 0$.

From (3.33), it is evident that $\Sigma_m \neq D_m$ and $D_m \neq 0$, then $\lambda_1 = \lambda_2 = 0$, and the only possible solution is $D_m = \frac{\Sigma_m}{2}$.

Since from Theorem 3.1 $\mathfrak{s}_m = -\frac{1}{2D_m}$, then we have that

$$\mathfrak{s}_m = -\frac{1}{\Sigma_m}. \tag{3.34}$$

Therefore, considering all the points belonging to the agent environment, we have

$$\mathfrak{s} = -\frac{1}{\mathbf{\Sigma}}. \tag{3.35}$$

■

Corollary 3.1. *Let $\mathfrak{s}_m = -\frac{1}{\Sigma_m}$ be the value of $\mathfrak{s}$ at the environment point $m \in \{1,...,M\}$ for an agent i. Then, the amount of information at this point is equal to 0.5 $bits$.*

Proof. From (3.6) we have $\mathfrak{s}_m = -\frac{1}{\Sigma_m}$.

From Theorem 3.1, if $\mathfrak{s}_m = -\frac{1}{\Sigma_m}$ then $D_m = \frac{\Sigma_m}{2}$.

So that

$$R(D_m) = \frac{1}{2}\log\frac{\Sigma_m}{\frac{\Sigma_m}{2}} = 0.5\ bits. \tag{3.36}$$

■

The results in (3.35) and (3.36), show the limits for the distortion and the rate at which the agent rationality is maximum. In other words, a decrease of the distortion below $\frac{D_m}{2}$, or an increase of the rate above 0.5 bits does not improve the agent understanding about its environment. This is demonstrated in the next section.

3.1.3 The Rationality Effect

As we have mentioned, the highest rationality occurs when $\mathfrak{s} = -\frac{1}{\mathbf{\Sigma}}$, whose equivalent distortion is $\mathbf{D} = \frac{\mathbf{\Sigma}}{2}$. This can be observed in Figure 3.2, which shows a minimum for the Kullback-Leibler value at this point, and a maximum in the borders. This effect is also visualized in Figures 3.3, 3.4, and 3.5, which show in the left side, the trajectories followed by a set of 10 agents moving through a variable spatial field, and the corresponding comparison between the distributions $p(x)$ and $p(y|x)$ in the right side. In Figure 3.3a, we observe the trajectories of the agents when $\mathfrak{s} = -\frac{100}{\mathbf{\Sigma}}$, i.e., a very low distortion value. In this case, the similitude between the distributions $p(x)$ and $p(y|x)$ is low, as depicted

in Figure 3.3b. On the other hand, Figures 3.3a and 3.3b show the behavior when $\mathfrak{s} = -\frac{1}{\Sigma}$, i.e., the highest rationality value. In this case, the trajectories of the agents are almost the same as in the previous case, and the similitude between the probability distributions is the highest. Finally, in Figures 3.5a and 3.5b, we show the case of $\mathfrak{s} = -\frac{1}{\Sigma}$, i.e., the lowest rationality. Here, the agents follow trajectories that cover more locations within the spatial field, making the system more exploratory. On the other hand, the similitude between the probability distributions also decreases.

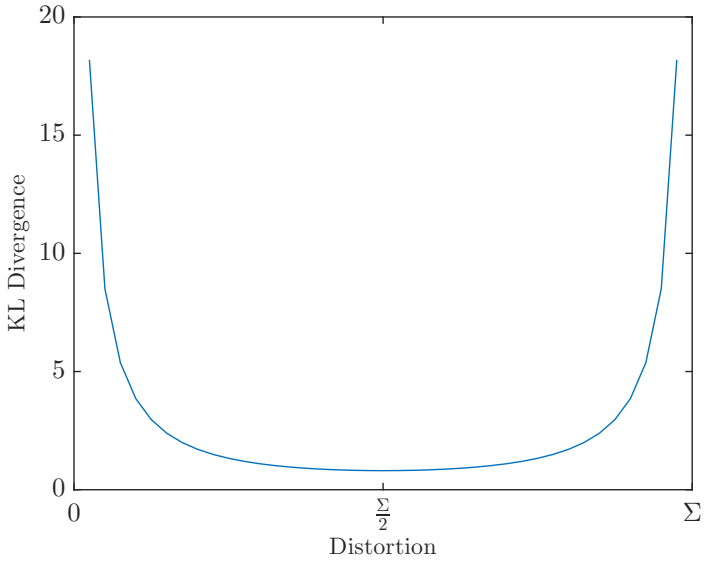


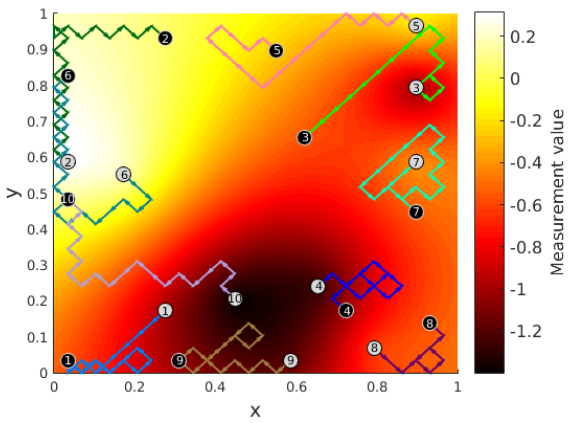
Figure 3.2: $KL[p(x)||p(y|x)]$ for different distortion values.

3.1.4 Agent Redundancy Tracking

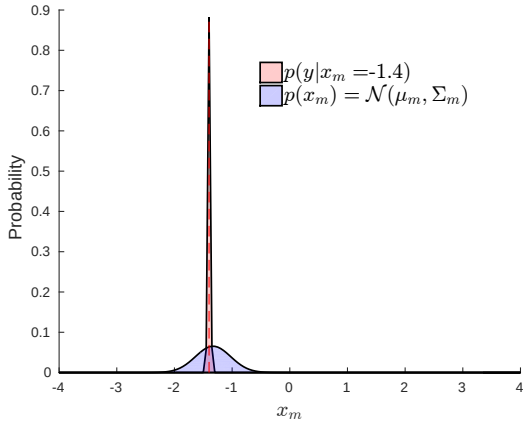
The agent understanding about the environment, given by the conditional $p(y|x)$, allows the agent to choose the action $z_i \in Z_i$ to be performed depending on the desired kind of exploration within the spatial field. In this sense, an agent selects the action having the highest conditional probability when it expects to follow redundant environment cues. On the other hand, an agent selects the action having the lowest conditional probability if it wants to avoid redundant cues coming from the environment. This lead us to the next pair of definitions.

Definition 3.4 (The less redundant environment location)**.** For an agent i, with action set Z_i, the action exhibiting the lowest redundancy about the environment is given by

$$z_m \in Z_i : \min \mathbf{p}(\mathbf{y}|\mathbf{x}) = p(y|x_m), \text{ for } m \in \{1,...,M\}. \tag{3.37}$$

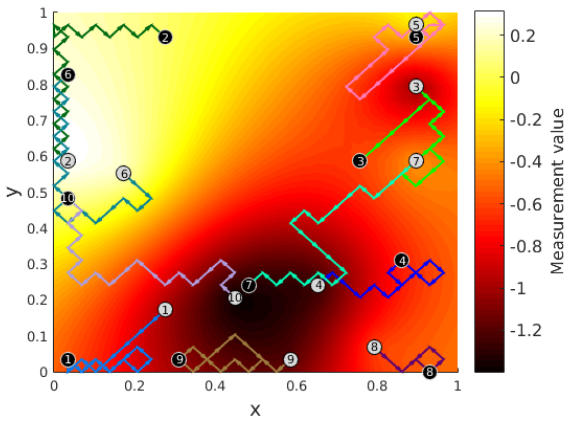


(a) Trajectories of agents when $\mathfrak{s} = -\frac{100}{\Sigma}$.

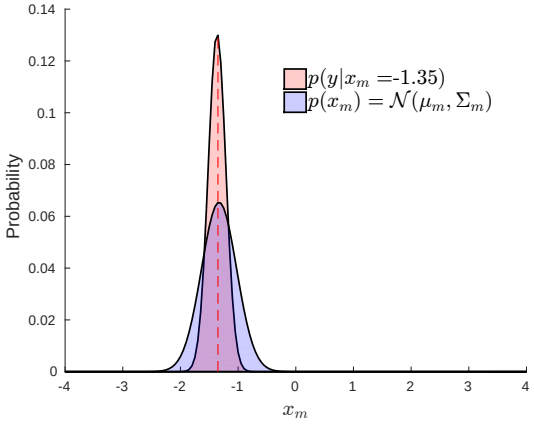


(b) The difference between $p(x)$ and $p(y|x)$ when $\mathfrak{s} = -\frac{100}{\Sigma}$.

Figure 3.3: Multi-agent system behavior for a very low distortion value.



(a) Trajectories of agents when $\mathfrak{s} = -\frac{1}{\Sigma}$.



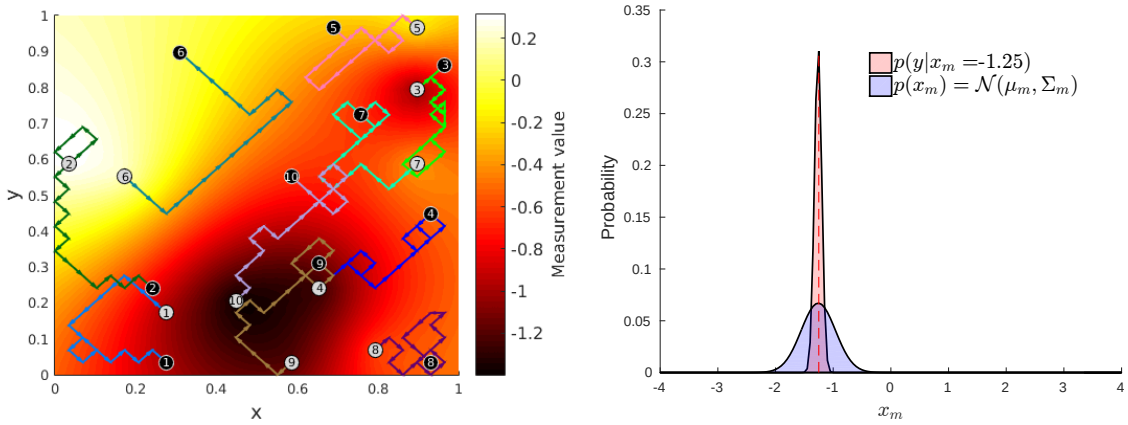
(b) The difference between $p(x)$ and $p(y|x)$ when $\mathfrak{s} = -\frac{1}{\Sigma}$.

Figure 3.4: Multi-agent system behavior for the highest rationality.

Definition 3.5 (The most redundant environment location)**.** For an agent i, with action set Z_i, the action exhibiting the highest redundancy about the environment is given by

$$z_m \in Z_i : \max \mathbf{p}(\mathbf{y}|\mathbf{x}) = p(y|x_m), \text{ for } m \in \{1, ..., M\}. \tag{3.38}$$

In Figure 3.6, we show the trajectories of a set of agents when they follow redundant and non redundant cues. Observe how in the case of Figure 3.6a, the agents tend to repel each other, contributing to the environment exploration. On the other hand, in Figure 3.6b, we can notice how the agents tend to follow the trajectories of their neighborhood,

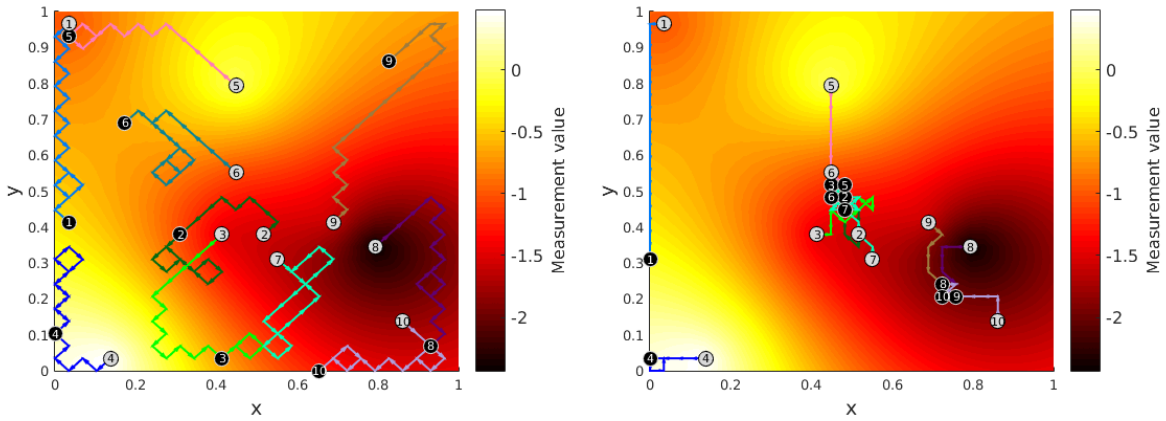


(a) Trajectories of agents when $\mathfrak{s} = -\frac{1}{2\Sigma}$.

(b) The difference between $p(x)$ and $p(y|x)$ when $\mathfrak{s} = -\frac{1}{2\Sigma}$.

Figure 3.5: Multi-agent system behavior for the lowest rationality.

promoting the cluster formation. Notice how in this case, the isolated agents do not have enough environment information, since their *training data* are reduced to their own measurements.



(a) Trajectories of agents folowwing non redundant cues.

(b) Trajectories of agents folowwing redundant cues.

Figure 3.6: Agent redundancy tracking.

The behavior exposed above, gains relevance in a system requiring that the agents develop a repulsive or an attracting behavior, depending on the number of individuals attending a specific event. This will be described in detail in Chapter 4, when the proposed model is applied in a sensor coverage problem.

So far, we have defined a multi-agent learning model in which agents have the ability to decide between two specific behaviors established by the parameter $\mathfrak{s}$ of the Blahut-Arimoto algorithm. First, they can adopt a behavior conditioned by the lowest rationality level, in which the environment exploration is promoted. Second, agents can understand their environment without the necessity of a very low distortion or high information levels as established in Theorem 3.2. Additionally, agents are able to define their actions according to the level of redundancy in their environments.

In the next chapter, we include the game theory approach in our model, in order to improve the agent performance, by means of a distributed optimization scheme that allows the system to find a *Nash equilibrium*.

CHAPTER 4

INFORMATION THEORY LEARNING MODEL AND EQUILIBRIUM CONVERGENCE

In this chapter, we include the potential games approach in order to define a scheme for the action selection used by the agents towards an equilibrium. First, we use the expected distortion, provided by the rate distortion function, as a potential function, whose minimization becomes the system objective. Second, the conditional distribution obtained through the minimization of the mutual information, which, as we stated in Chapter 3, describes the agent understanding about the environment, defines a *Logit dynamics* pattern that the agents use to choose the actions offering the best utility in an event tracking setting. Finally, the rationality levels also defined in Chapter 3, which are determined by the parameter $\mathfrak{s}$ of the Blahut-Arimoto algorithm, are used to establish a convergence time towards a *Nash equilibrium*.

We show the performance of the final model in a mobile sensing setting, in which the coverage problem is addressed. We first show the agent behavior in an invariant environment, where the results resemble the consensus based potential function described in [60]. Then, the model is implemented in a variant environment, and the distortion minimization demonstrates to be dependent on the environment measurements and the agent locations. Finally, our model exhibits a very good performance in a more realistic setting, in which the trending network coverage problem is addressed with satisfactory results in terms of convergence time, and number of agents, in comparison with the results exposed in [60] and . However, in spite of the importance of this convergence time

and the number of agents to reach the *Nash equilibrium*, the main result of this work, is the ability of the agents to detect redundancy in their environment, an in this way, the capacity to decide about the actions to assume. This, to the best of our known, has not been addressed in literature.

4.1 The Logit Dynamics Pattern

As we shown in Chapter 2, one of the outputs of the Blahut-Arimoto algorithm is the conditional distribution, which for a point $m \in \{1,...,M\}$, i.e., a point belonging to the agent action set, is given by

$$p(y|x=\mu_m) = \frac{p(y)e^{\mathfrak{s}(y-\mu_m)^2}}{\sum_y p(y)e^{\mathfrak{s}(y-\mu_m)^2}}, \tag{4.1}$$

which measures the agent understanding about the environment, under a distortion value determined by $\mathfrak{s}$. Equation (4.1), can also be shown as a Boltzmann distribution, in which $\mathfrak{s}$ resembles the inverse temperature. This lead us to think on the conditional distribution, as the expression used by the agents to choose their strategies, according to a rationality measure given by the parameter $\mathfrak{s}$, whose limits were established through Theorems 3.1 and 3.2 in Chapter 3.

Definition 4.1 (*Logit dynamics* pattern)**.** The expression in (4.1) defines the strategy updating rule for an agent i, according to the utility function

$$U_i = (y-\mu_m)^2, \tag{4.2}$$

with $y \in \mathcal{Y}$[1].

This means that an agent will choose the action having the highest similitude in relation with the environment when it is developing a tracking event task. However, there are some cases in which the number of agents attending a specific event, could produce redundant measurements and unnecessary data transmission to sink points. In this sense, the agents can revert the benefit described in (4.2), which lead them to refuse dense locations and to explore new field positions, following the *redundancy tracking behavior* exposed in Section 3.1.4. This will be demonstrated in the network coverage problem described in Section 4.4.4. So far, we have defined the dynamics to be used by the agents in order to make their action choice. In the next section, we define the distortion based potential function that allows us to find an equilibrium.

[1]The alphabet $\mathcal{Y}$ represents a set of possible measurement values taken by the agent.

4.2 Distortion Based Potential Game

As we shown in Chapter 2, the *Nash equilibrium* of a multi-agent system can be found if the utility change of the each agent can be mapped through a potential function. In our case, this potential function is based on the expected distortion established by the rate distortion function. This is described by means of the next theorem.

Theorem 4.1. *Let $Z = \prod_1^N Z_i$ be the set of strategy profiles for a potential game $\mathcal{G}$ having N agents, in which the agent $i \in \{1,...,N\}$, has the action set $Z_i = \{z_1,...,z_M\}$. The agent utility function given in (4.2), constitutes an ordinal potential game whose potential function is the expected distortion measure given by*

$$\phi(z) = \sum_{\mu(z),y} p\left(y,\mu(z)\right)\left(y-\mu(z)\right)^2, \tag{4.3}$$

where $\mu(z)$ is the set of estimated mean values at the set of locations $z = (z_1,...,z_N) \in Z$ constituting the system action profile.

Proof. According to (2.55), $\phi(z)$ is a potential function if

$$\phi(z_2,z_{-i}) - \phi(z_1,z_{-i}) > 0 \Longleftrightarrow U_i(z_2,z_{-i}) - U_i(z_1,z_{-i}) > 0, \tag{4.4}$$

where, $z_1, z_2 \in Z_i$, and z_{-i} is the set of actions assumed by the agents different to i. Then, without loss of generality, from (2.22), and Theorem 3.2 we have that

$$\phi(z_2,z_{-i}) - \phi(z_1,z_{-i}) = \frac{\Sigma_2}{2} - \frac{\Sigma_1}{2} > 0. \tag{4.5}$$

The conditional probability $p(y|x=\mu_m)$ is inversely proportional to Σ_m, then

$$\begin{aligned} p(y|\mu_2) < p(y|\mu_1) = \\ p(y)e^{\mathfrak{s}(y-\mu_2)^2} < p(y)e^{\mathfrak{s}(y-\mu_1)^2} = \\ \log(e^{\mathfrak{s}(y-\mu_2)^2}) < \log(e^{\mathfrak{s}(y-\mu_1)^2}) = \\ \mathfrak{s}(y-\mu_2)^2 < \mathfrak{s}(y-\mu_1)^2. \end{aligned} \tag{4.6}$$

Since $\mathfrak{s} \in \mathbb{R}_{<0}$, then $(y-\mu_2)^2 > (y-\mu_1)^2$, and

$$U_i(z_2,z_{-i}) - U_i(z_1,z_{-i}) > 0. \tag{4.7}$$

■

With this in mind, the system objective is the minimization of the expected distortion.

At this point, we have defined all the aspects involved in the learning model proposed in this work. In the next section, we summarize it and explain its computational implementation.

4.3 Information Theory Based Learning Model

Figure 4.1, summarizes the proposed multi-agent learning framework. Recalling the multi-agent model described in Section 3.1, first we have the sets of training points $S = \{s_1,...,s_N\}$, and training values $V = \{v_1,...,v_N\}$, coming, respectively, from the positions and the measurements of the agents belonging to the neighborhood of the agent i, which are used to infer, by means of GPR, the testing values $W = \{w_1,...,w_M\}$, in a set of previously known testing points $Z_i = \{z_1,...,z_M\}$. In this way, an agent i obtains a set of mean and covariances for each point $m \in \{1,...,M\}$ of its action set, which define the set of Gaussian distributions $\mathbf{p(x)}$ describing its environment. The rationality values, found through Theorems 3.1 and 3.2, are used to define the desired amount of information that the agents want to have about the environment, in order to promote the field exploration when it is low, or to accelerate the equilibrium convergence of the system when it is high. Once, the source distribution $\mathbf{p(x)}$ and the rationality value of $\mathfrak{s}$ are defined, the conditional probability $p(y|x = \mu_m)$ describing the similitude between the agent and a point $m \in \{1,...,M\}$ of its environment[2], is used as a *Logit dynamics* pattern, though which the agent choose their actions. The selected action, can be oriented to follow high or low redundancy, improving the event tracking in the first case, or the avoidance of repetitive lectures due to the cluster formation in the second case, as we stated in Section 3.1.4. Finally, the distortion based potential function $\phi(\mathbf{z})$, allows the system to find a *Nash equilibrium* at which the agents do not receive incentive to change their strategies unilaterally.

[2]The set of points $Z_i = \{z_1,...z_m\}$ corresponding to the *action set* of the agent i, is the same set of points corresponding to its environment, as we stated in Chapter 3.

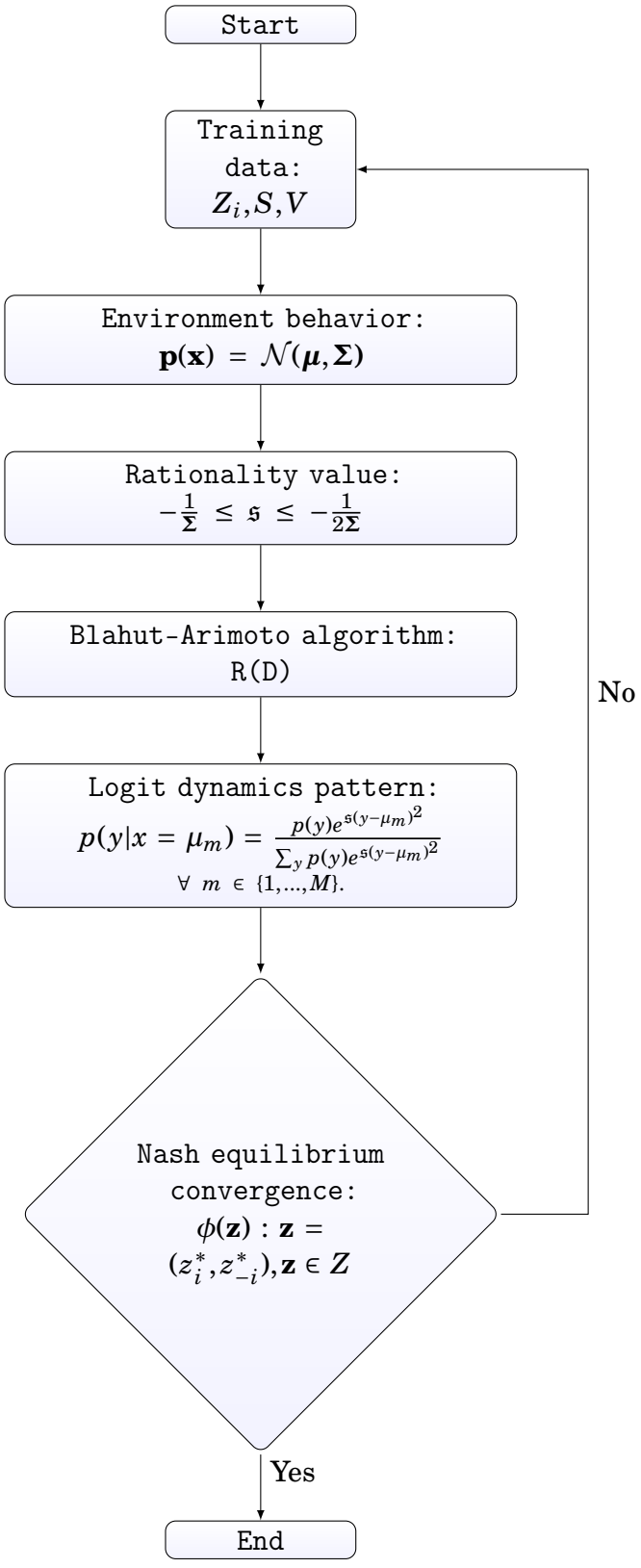


Figure 4.1: The information theory based learning model.

4.3.1 Computational Implementation

Algorithm 1 shows the computational implementation of our model. The Blahut-Arimoto algorithm, described in Section 2.2.1, is included to obtain the conditional distribution that defines the action selection rule of the agents. The set of covariance matrices, C_*, C, and C_{**}, are obtained using the kernel function RBF, as defined in the GPR approach. The selected action z_i depends on the desired behavior of the agent, which can be to follow redundancy or not. This behavior, will be applicable in the network coverage problem described in Section 4.4.4, where the redundancy avoidance is used for the agents to scape from positions covered by others. This algorithm calculates the action to be selected by each agent, until the utility of each position belonging to the action set is the same or almost the same, in such a way that there is no incentive to move,i.e., until the *Nash equilibrium* condition is satisfied. In Section 4.4, we firstly show the model

Algorithm 1 REDUNDANCY BASED LEARNING

Input: S, V, Z_i
Output: $z_i^* \in Z_i$

while $U_i(z_i, z_{-i}^*) \leq U_i(z_i^*, z_{-i}^*)$ **do**
  $M \leftarrow length(V)$
  $\mu \leftarrow C_* C^{-1} V$
  $\Sigma \leftarrow C_{**} - C_*^T C^{-1} C_*$
  for $m \leftarrow 1$ **to** M **do**
    $p(x) \leftarrow \mathcal{N}(\mu_m, \Sigma_m)$
    $-\frac{1}{\Sigma_m} \leq \mathfrak{s}_m \leq -\frac{1}{2\Sigma_m}$
    $p(y|x = \mu_m) \leftarrow$ BLAHUT-ARIMOTO$(p(x), \mathfrak{s}_m)$
  end for
  if Event tracking **then**
    $z_i \leftarrow \{z_m : \max \mathbf{p(y|x)} = p(y|x_m)\},\ m \in \{1, ..., M\}$
  else if Environment redundancy **then**
    $z_i \leftarrow \{z_m : \min \mathbf{p(y|x)} = p(y|x_m)\},\ m \in \{1, ..., M\}$
  end if
  return z_i
end while

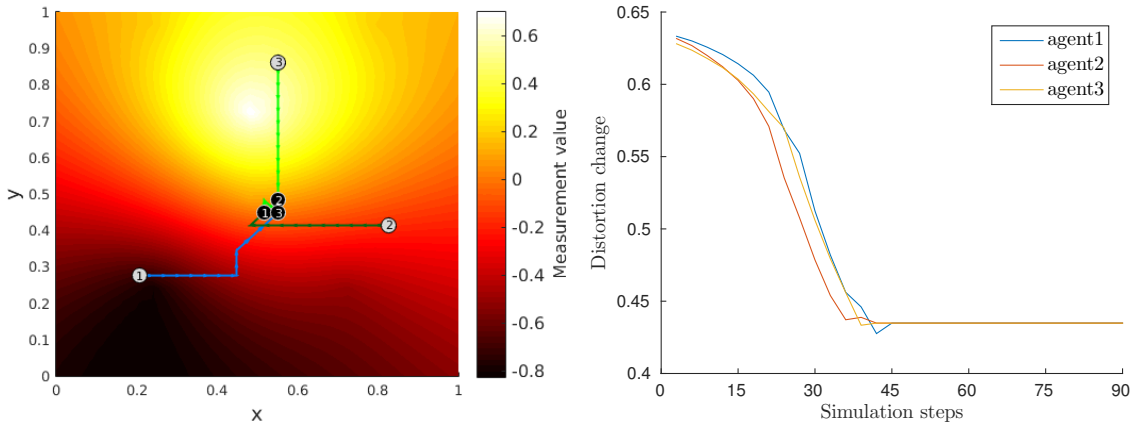
performance in a variant and an invariant environment, in order to demonstrate the effect of the measurements on the equilibrium convergence. Finally, we show how this approach, works in a realistic setting that solves the network coverage problem.

4.4 Model Implementation

In this section, we show the results of the model implementation through a set of simulations developed in a spatial 2-D field given by a 100×100 grid, in which, the agents are deployed at random locations. The initial and final positions of the agents, are represented by grey and black small circles, respectively. At each time step, an agent i, selected at random, chooses its action from the *action set* Z_i, according to (4.1), until the potential function minimization is reached. First, by means of a simple example using three agents, we demonstrate the effect of the parameter $\mathfrak{s}$ on the learning rationality. Second, we show how the agents behave in an invariable and in a variable setting to demonstrate the effect of the environment change on the agent learning process. Finally, we show a more realistic case in which a higher number of agents move in the spatial field in order to fulfill a sensor coverage problem.

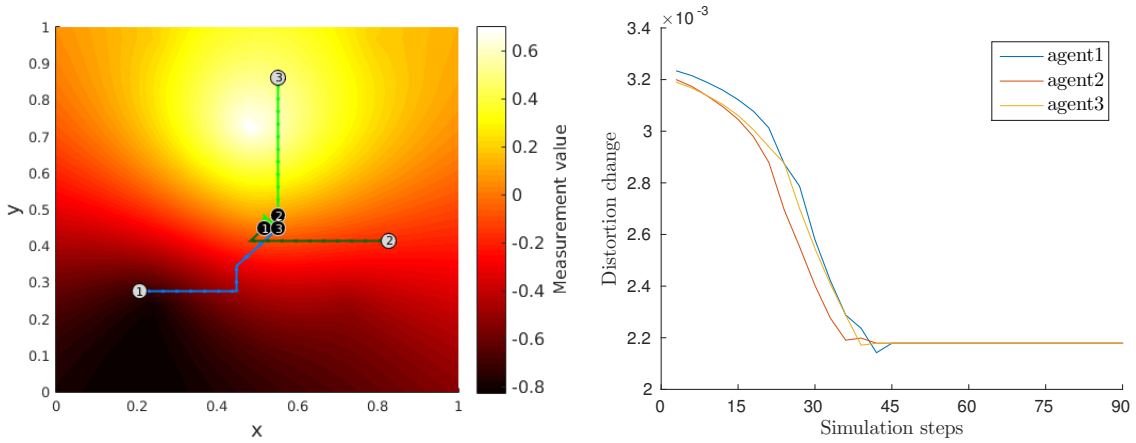
4.4.1 The Rationality Effect in Equilibrium Convergence

As we demonstrated in Theorem 3.2, there is a highest rationality value at which an agent obtains the maximum information about the environment, and at which, a distortion decrease does not improve the convergence time time to a Nash equilibrium. This is shown in Figures 4.2a and 4.3a, in which we can observe, respectively, the trajectories followed by the agents for $\mathfrak{s} = \frac{-1}{\Sigma}$, and for $\mathfrak{s} = \frac{-200}{\Sigma}$, i.e., the highest rationality value, and a value of $\mathfrak{s}$ corresponding to a very low distortion, equivalent to $D = \frac{\Sigma}{400}$. We can notice that in addition to the similitude in the trajectories in both cases, the convergence time is the same, in spite of the difference in the distortion levels, as shown in Figures 4.2b and 4.3b. This proves that a decrease of the distortion value below the one settled by the highest rationality, does not produce an improvement in the convergence time towards the *Nash equilibrium*. On the other hand, in Figure 4.4a, we show the trajectories of the agents when the parameter $\mathfrak{s}$ is equal to $\frac{-1}{2\Sigma}$, i.e., the lowest rationality. In this case, we can observe how the path of agent 1, initially deviates to locations far from its neighbors, due to the increase in the distortion perceived about the environment, as shown in Figure 4.4b. Additionally, in this last case, the convergence time to the Nash equilibrium is longer than the one obtained in the previous cases.



(a) Trajectories of agents when $\mathfrak{s} = \frac{-200}{\Sigma}$ or $D = \frac{\Sigma}{400}$. (b) The corresponding potential function minimization.

Figure 4.2: The rationality effect when $\mathfrak{s} = \frac{-200}{\Sigma}$.

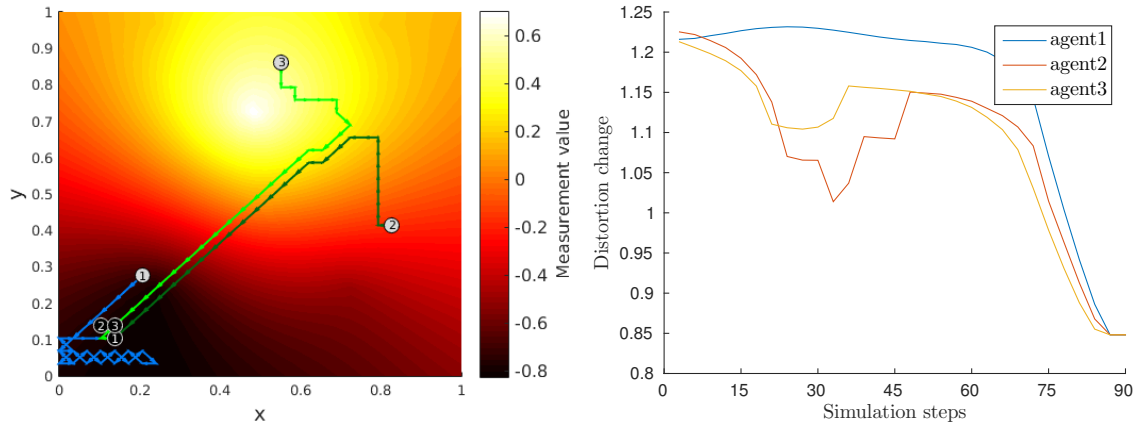


(a) Trajectories of agents when $\mathfrak{s} = \frac{-1}{\Sigma}$, i.e., the highest rationality. (b) The corresponding potential function minimization.

Figure 4.3: The rationality effect when $\mathfrak{s} = \frac{-1}{\Sigma}$.

4.4.2 Model Performance in an Invariable Environment

In Figure 4.5, we show the behavior of three agents moving in an invariant environment. In this case, we assume that all the agents are connected, which means that the *training data* set for each one is $V = [v_1, ..., v_3]^T$, where $v_1 = v_2 = v_3$, since the environment is not variable. As a result, we can notice in Figure 4.5a, how the agents reach a final arrangement in which the *distortion based potential function* is minimized in approximately 45 simulation steps, as shown in Figure 4.5b. Additionally, we can observe

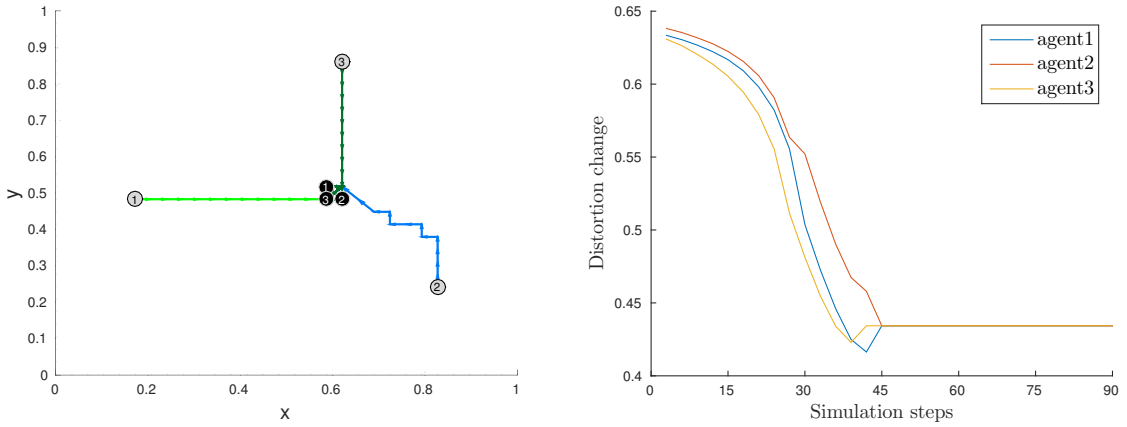


(a) Trajectories of agents when $\mathfrak{s} = \frac{-1}{2\Sigma}$, i.e., the lowest rationality.

(b) The corresponding potential function minimization.

Figure 4.4: The rationality effect when $\mathfrak{s} = \frac{-1}{2\Sigma}$.

how the distortion for the three agents is almost the same during the whole simulation, due to the uniform state of the environment.



(a) Trajectories followed by the agents until the *Nash equilibrium*.

(b) The potential function minimization.

Figure 4.5: Agent behavior in an invariable environment.

Due to the unchanging environment, the agent utility function

$$U_i = (y - C_* C^{-1} V)^2, \tag{4.8}$$

only depends on the matrices

$$C_* = \alpha e^{\frac{-||s_n - z_m||}{\gamma}}, \tag{4.9}$$

and

$$C = \alpha e^{\frac{-||s_i - s_j||}{\gamma}}, \tag{4.10}$$

with $z_m \in Z_i$, and $s_n \in S$, i.e., the utility function, mostly depends on the distances between the agents and the locations of the *action set*. This resembles the consensus based utility function described in [60] [70], which only depends on the euclidean distances between the agents, and is given by

$$U_i(z_i, z_{-i}) = -\sum_{j \in \mathcal{N}_i} ||z_i - z_j||, \tag{4.11}$$

where $\mathcal{N}_i$ is the neighborhood of the agent i. The potential function in this case, is equated to the utility. Then $\phi(\mathbf{z}) = U_i(\mathbf{z})$, where $\mathbf{z} \in Z$, i.e., $\mathbf{z}$ is the current action profile of the system. The action choice, in a spatial field with obstacles, follows a model named restrictive spatial adaptive play (RSAP), where a trial action $\hat{z_i}$, is selected from the highest of the two following probabilities, which determine if the agent has to move or to stay in the current position, respectively.

$$Pr[\hat{z_i} = z_i] = \frac{1}{k_i}, z_i \in R(z_i(t-1)) \backslash z_i(t-1) \tag{4.12}$$

$$Pr[\hat{z_i} = z_i(t-1)] = 1 - \frac{(|R_i(z_i(t-1)| - 1)}{k_i} \tag{4.13}$$

where $R_i(z_i(t-1))$ is the set of restricted actions due to the obstacles, $k_i = max_{z_i \in Z_i} |R_i(z_i)|$ denotes the maximum number of possible actions if were not obstacles, and $z_i(t-1)$ represents the current position. Once the trial action is selected, the agent action for the next time step is chosen according to the probabilities

$$Pr[z_i = \hat{z_i}] = \frac{e^{\beta U_i(\hat{z_i}, z_{-i}(t-1))}}{G}, \tag{4.14}$$

and

$$Pr[z_i = z_i(t-1)] = \frac{e^{\beta U_i(z_i(t-1))}}{G}, \tag{4.15}$$

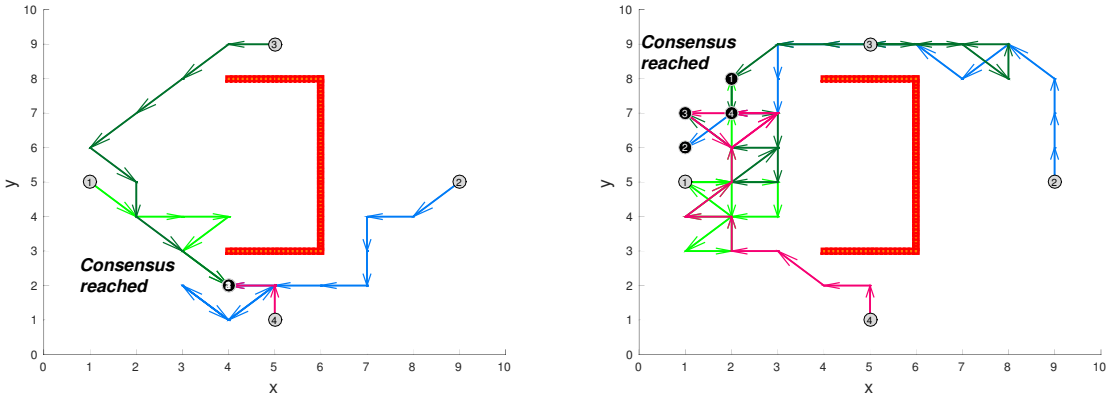
where β is the rationality measure, and

$$G = e^{\beta U_i(\hat{z_i}, z_{-i}(t-1))} + e^{\beta U_i(z_i(t-1))}. \tag{4.16}$$

Equations (4.14) and (4.15) define if the agent assumes the trial action or if it prefers to keep its current location.

Coming back to our approach, in Figures 4.6a and 4.6b, we compare its performance with the results of the model described above in an invariant environment having

obstacles. As we can notice, in terms of simulation steps, the *information theory based model* outperforms the results of the *consensus based model* shown in [60], since the action choice in our case, is directly defined through the environment information, established by the *Logit pattern* of (4.1), without the previous selection of a trial action. Additionally, in our case, the rationality measure is constant and determines a highest understanding about the environment, whereas in [60], this parameter is changed arbitrarily until a good performance is found.



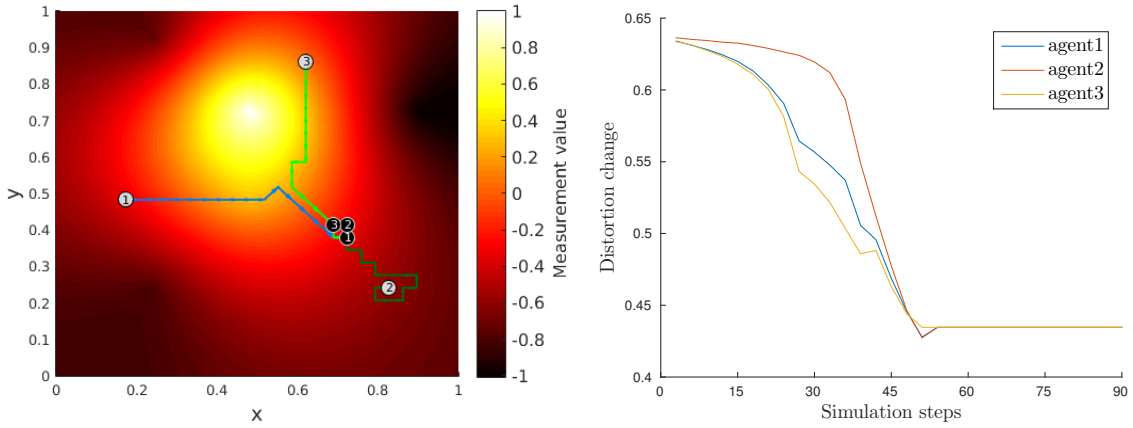
(a) Concensus based model of [60]. Convergence after 1500 simulation steps.

(b) Information theory based learning model. Convergence after 240 simulation steps.

Figure 4.6: Comparison of our approach with the results shown in [60] in an invariant field with obstacles.

4.4.3 Model Performance in a Variable Environment

The case of a variable environment is shown in Figure 4.7. Here, we can observe how the measurement variations found in the environment of an agent i, affect the conditional probability value $p(y|x = \mu_m)$ at each location $z_m \in Z_i$. In this sense, the final system arrangement, shown in Figure 4.7a, not only depends on the distances between the agents and their action sets, but also of the agents measurements and the inferred values. In this case, in Figure 4.7b, we can observe at the beginning, how the distortion value has a considerable difference for agent 2 in relation to the values of agents 1 and 3. This is a consequence of the measurements variation around them, since agents 1 and 3 have a similar setting, which is very different to the case of agent 2. Additionally, due to the environment variability, the potential function requires more time steps to be minimized in comparison with the case shown in Section 4.4.2.



(a) Trajectories followed by the agents until the *Nash equilibrium*.

(b) The potential function minimization.

Figure 4.7: Agent behavior in a variable environment.

4.4.4 Distributed Coverage Control

In Figure 4.8, we show the model performance in a field having 10 agents trying to cover a specific area at which an event has occurred. In this setting, agents have different sensing radius (rs), which are represented by colored circles surrounding them, whereas each connection radius is defined as $2rs$. As in [60], [19], and [123], in our model the event detection is associated to a threshold level, which we have named t_{hr}. In this regard, as we stated in Section 3.1.4, an agent i chooses the action having the lowest conditional probability, i.e.,

$$z_m \in Z_i : \min[p(y|x = t_{hr})] = p_m(y|x = t_{hr}), \tag{4.17}$$

when its measurement is far from t_{hr}. In other words, the agent moves towards the environment locations having the highest difference to its current measurement.

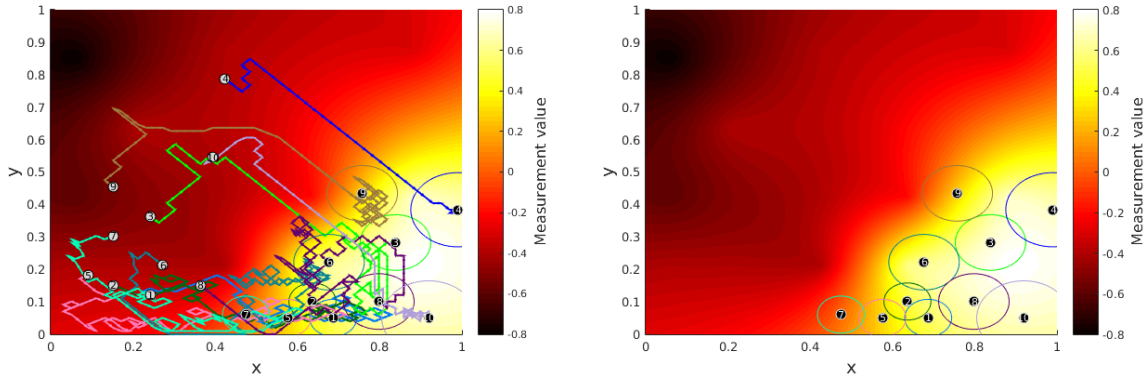
On the other hand, an agent i, chooses the action having the highest conditional probability, i.e.,

$$z_m \in Z_i : \max[p(y|x = t_{hr})] = p_m(y|x = t_{hr}), \tag{4.18}$$

when its measurement is close to t_{hr}. This means that the agent moves towards the environment locations having the lowest difference to its current measurement. In this case, after a finite number of simulation steps, all the conditional probabilities $p(y|x = t_{hr})$ of each action $z_m \in Z_i$, begin to exhibit similar values, which means that the agent i does not have an action that improves its utility, whereas the other agents remain static, in other words, they reach the *Nash equilibrium*. In Figure 4.8a, we can observe

the initial positions and the trajectories followed by each agent until they reach the locations around the environment event. Here, we can notice how some agents such as the numbers 9 and 7 try to explore far from the profitable locations, and how they decide to change their path towards more interesting positions reported by their neighbors. The final configuration after 1000 time steps is shown in Figure 4.8b, and the behavior of the *distortion based potential function* is shown in Figure 4.9. As we can observe, the convergence to a minimum distortion in the system is obtained after 500 time steps, with some slight increases after this point, which are caused by the repulsive movement of agents in the borders, since they try to avoid locations covered by others.

The exhibited results outperform the results shown in [60] in terms of the number of agents necessary to cover an event in a spatial field. This can be attributed to the fact that in our model, each agent requires less time to decide about its next action, since it does not require to make a previous calculation about a trial action, as stated in the RSAP algorithm described in [60]. Additionally, in our case we have defined the value at which the rationality measure ($\mathfrak{s}$) determines the highest understanding about the environment, whereas in [60] this parameter is changed arbitrarily until a good performance is found. Although we have shown the results in an environment having only a covered region, the model is not limited to this type of settings, which means that multiple events can be attended by the agents, leading the system to multiple local Nash equilibrium. This is demonstrated through the results shown in Fig. 4.10. In Fig. 4.10a we can observe the trajectories followed by the agents to attend two relevant locations, whereas Fig. 4.10b shows how the *distortion based potential function* is minimized in less time steps due to the change in the initial conditions and the increase of the rewarding regions. The distortion increase exhibited at the beginning is produced by the initial isolation of agents 3 and 4, which is reduced when they begin to find each other through the communication link represented by the colored circles.



(a) The initial deployment of agents in the spatial field and their trajectories towards the *Nash equilibrium*.

(b) The final network configuration.

Figure 4.8: Coverage problem.

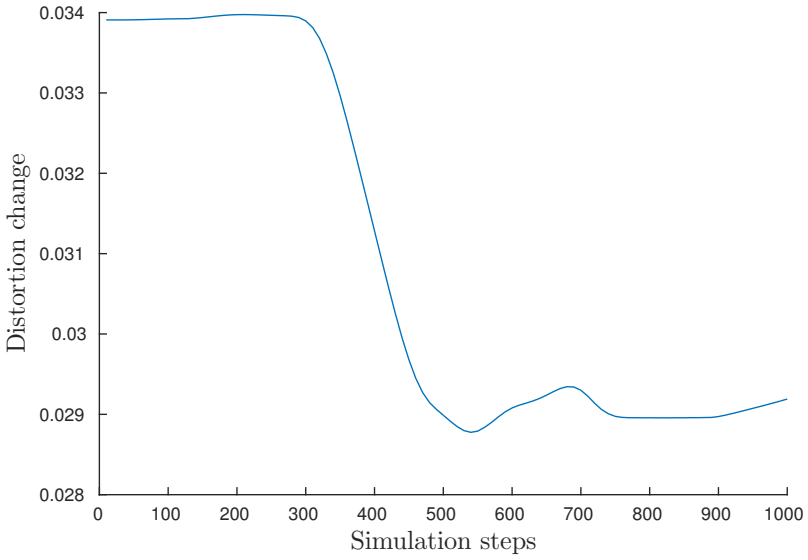
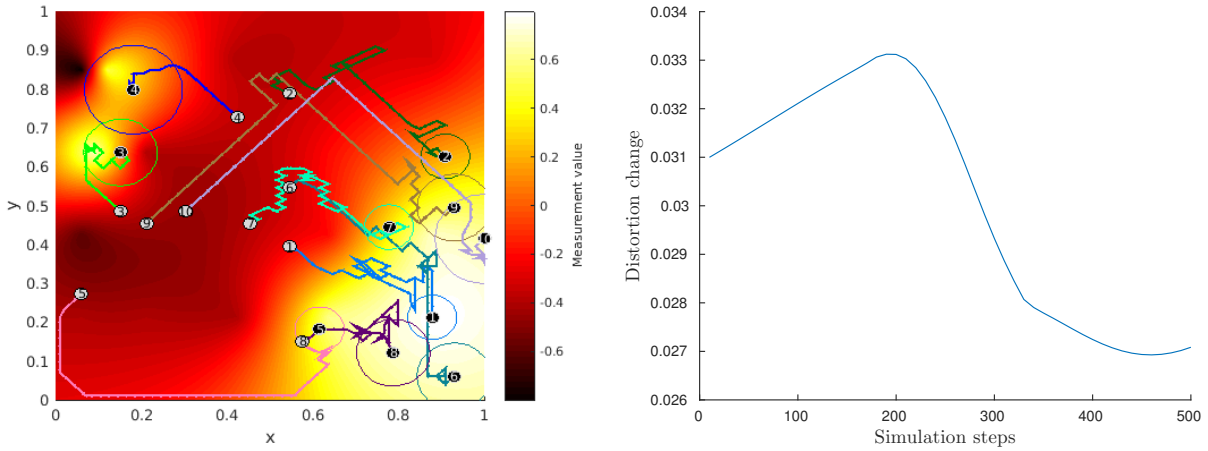


Figure 4.9: The evolution of the potential function of the system.



(a) The initial deployment of agents in the spatial field and their trajectories towards the *Nash equilibrium*.

(b) Evolution of the potential function of the system.

Figure 4.10: Network coverage problem for two rewarding regions and change in the initial conditions shown in Fig. 4.8.

CHAPTER 5

Conclusions and Future Directions

In this work, we use some of the advantages offered by the information theory to define a multi-agent learning framework in which the information acquires importance in the agent decision making process. The model is based on three approaches. The first one is the Gaussian process regression (GPR), through which agents infer their environment. The second one is the rate distortion function (RDT), which defines a redundancy border of the environment understanding for each agent. The last one is the potential games approach, which, along with the established *distortion based potential function* and the rationality levels given by the Lagrangian multiplier $\mathfrak{s}$, allow the system to find a Nash equilibrium.

By means of the Blahut-Arimoto algorithm, used to calculate the RDF, we found two relevant values that determine the lowest and the highest rationality measures, which, as we demonstrated in a mobile sensor network, improve the agent understanding about the environment and the field exploration, respectively. Additionally, the redundancy based decision making, allows the agents to avoid overpopulated locations and migrate to positions with promising welfare. This redundancy avoidance, has an important role in the current sensing network design, especially in IoT applications, where the number of monitoring devices increases continuously.

On the other hand, the Boltzmann form of the obtained conditional distribution, which in our case represents the similitude between the agent and the environment behavior, demonstrated to be an effective action selection rule for the agents in variant and invariant spatial fields, even in cases involving obstacles, in which we outperform

the consensus based potential game presented in [60] in terms of time steps.

From the distributed optimization point of view, the proposed *distortion based potential function* demonstrated a good performance in terms of the distortion minimization of the environment information, and consequently the *Nash equilibrium* convergence in an acceptable number of time steps, as we shown in the network coverage problem. Additionally, due to the reduced number of probability calculations, in comparison with the RSAP model initially proposed in [60], the action choice in our model does not require a considerable number of agents to track and cover locations exhibiting relevant events for the system.

Since this model is based on strategic games, in which agents choose their action before the others choose theirs, the GPR inference tool provides an indirect connection to the agents far from the agent neighborhood, since the actions taken by a neighbor of one of its neighbors, are finally reflected in the *training data* set. This effect is visible in the event tracking case of the coverage problem, in which some agents, in spite of being outside of the connection radius of the ones located in wealth positions, move towards these regions. In addition to this virtual neighborhood extension, the increment of the *training data* points improves the environment prediction, and in this way, the quality of the decisions. This is proven when the points inside of the sensing radius are used to decrease the environment distortion, and therefore, the convergence time.

In spite of the fact that the model performance was described for mobile sensor networks, its applicability was also demonstrated in the smart grids context, in which the reactive power sharing problem was solved in an acceptable way, without considerable affectations in the voltage regulation in a set of four DG's when one of them suffers an overload. Additionally, in this case we demonstrate the versatility of the model to involve other types of game theory approaches in order to accomplish a distributed optimization requirement, which was proven through the use of evolutionary game theory and the replicator dynamics concept.

In general terms, the proposed model offers a learning structure in which the redundant information of the environment is a determinant factor in the agent decision making process, which is a relevant factor if we consider the continuous raising of mobile sensor networks and their use in IoT applications, which generate high amounts of redundant data. Additionally, due to the permanent node mobility, these types of networks require distributed synchronization schemes, in which the nodes can decide when and which one has to transmit data to the sink. In this sense, we are working in a distributed synchronization model that combines the maximum entropy principle and

hybrid dynamical systems.

APPENDIX A

INFORMATION THEORY LEARNING MODEL FOR REACTIVE POWER SHARING IN MICROGRIDS

In this chapter, we show a variation of the model described in Chapters 3 and 4, which is used in a power system application. The implementation combines the known concepts of information theory such as the maximum entropy (MaxEnt), and the rate distortion function (RDF), to control the reactive power sharing in an islanded microgrid. In this case, the agents are representations of the distributed generators, named DG's, and the environment behavior is determined by the shared information between them through a communications network. The distortion level of the information that each agent has about the environment, determines a *distortion based fitness function*, which is used in a *replicator dynamics* setting to control, in a distributed way, the power support in the DG's when they are overloaded.

A.1 Motivation

The massive rising of technological solutions focused to energy generation in isolated locations, using renewable and environmentally friendly sources, has led to the implementation of distributed generators (DG) for electrical networks with low and medium scope, designed to have a connection with the conventional power network, providing significant benefits in the operation, such as power ancillary services for management [18].

The idea of the distributed generation, has been embraced by the concept of microgrids, which can be described as a resource to interconnect conventional voltage transmission systems with the mentioned isolated distributed generators [52]. The elements comprising a microgrid, such as the storage devices, and loads, among others, can be managed through a centralized control system, named the point of common coupling (PCC), or through a distributed system with abilities to stabilize voltage and frequency faults on each node, without the intervention of a centralized entity [58].

When a microgrid operates in this islanded mode, the main challenge is the sharing of the reactive power demand between all the nodes (DG's), in a way that has coherence with the capacity of each one. The complexity of this objective is increased due to the conflict between the voltage regulation and the reactive power sharing, caused by the operating characteristics of the DG's, since both variables have a dependency conditioned by the drop control [96]. Such dependency, avoids to have a good performance in the secondary control in terms of voltage regulation and reactive power simultaneously, which has become in an interesting issue for the research community. In this sense, we propose a multi-agent learning model, based on information theory that addresses the problem of reactive power sharing, without affecting the voltage regulation considerably.

In a first step, the model calculates the maximum entropy (see Section 2.2.1.2) of every reactive power input around an expected value that depends on the capacity in every DG. Second, a fitness function is determined according to the value of the distortion of every node with the environment, using the rate distortion function described in Section 2.2. Finally, we use this fitness function in a replicator dynamics context, in order to identify where and when some DG's require reactive power support from the system.

In the next section, we describe the main technical concepts related to a microgrid.

A.2 Microgrid Control

In order to describe the control system of a microgrid, let us begin considering the microgrid model shown in Figure A.1. This model is composed of a set of $N = 6$ buses, or agents, in which two are loads, and four are DG's. A reactance line connecting the pair of buses i and j, is denoted as X_{ij}. The active and the reactive power injections for a bus i, are denoted by P_i and Q_i, respectively, and are given by

$$P_i = \sum_{j=1}^{N} \frac{E_i E_j}{X_{ij}} \sin\left(\theta_i - \theta_j\right), \tag{A.1}$$

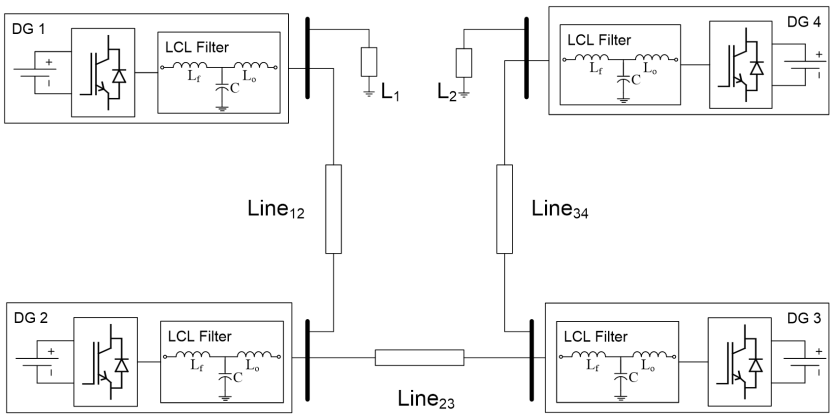


Figure A.1: Microgrid model.

and

$$Q_i = \frac{E_i^2}{X_i} - \sum_{j=1}^{N} \frac{E_i E_j}{X_{ij}} \cos\left(\theta_i - \theta_j\right), \tag{A.2}$$

where E_i is the voltage magnitude, θ_i is the phase angle, and $X_i = \frac{1}{\sum_{j=1}^{N} X_{ij}^{-1}}$.

The heterogeneity of the generation in microgrids, includes systems such as photovoltaic, wind, and micro turbine, among others, which normally produce DC or variable frequency power that require inverters to allow the DG's connect to a synchronous AC power system [97]. This connection process, demands control actions to achieve adequate reactive power sharing between DG's in order to avoid overcharges on them. In this regard, there are three levels associated to the voltage control in microgids [96], which we describe below.

A.2.1 Primary "Droop" Control

A main objective of a primary control is the microgrid stabilization by means of "droop" controllers for the inverters [20], which causes voltage deviations in the buses of the microgrid. In an islanded operation, the inverters operate as VSIs (Voltage Source Inverters) with controlled voltage magnitudes. The "droop" controllers also provide the voltage references, which are based on the decoupling between the active and reactive power. For the inductive lines, these controllers specify the inverter voltage magnitudes E_i and frequency ω_i, which are given by

$$E_i = E^* - n_i\left(Q_i - Q_{i,set}\right), \tag{A.3}$$

and

$$\omega_i = \omega^* - m_i\left(P_i - P_{i,set}\right), \tag{A.4}$$

where E^* is a nominal network voltage, Q_i is the measured reactive (non-active) power injection, ω^* is the nominal network frequency, and P_i is the measured active power injection. The constants n_i and m_i, are the "droop" coefficients. Finally, the quantities $Q_{i,set}$ and $P_{i,set}$, represent the reactive and active power set points, respectively.

A.2.2 Secondary Control

In conventional interconnected electrical power systems, the sharing of reactive power demand among generators, is not a relevant problem, due to the capacitive compensation of the loads and the transmission lines. On the other hand, in microgrids, the low ratings of DG units, the short electrical distances between nodes, and the lack of compensation, require an accurate sharing of the reactive power demand among DG's in order to avoid overloading. In this sense, due to the impedance of the transmission lines, the primary "droop" controller is unable to share reactive power among identical or different inverters [58], which creates the necessity of a secondary controller that fulfills this purpose.

However, this reactive power sharing attempt, produces a conflict with the voltage regulation, which can be observed in Figure A.2. First, without secondary control, two DG's operate at voltages E_1 and E_2 with their corresponding reactive power injections Q_1 and Q_2, as represented by the black line. Once the secondary voltage-regulating control is applied, the voltage in both DG's is restored to a common rating denoted as E^*, being the green line for the first DG, and the blue one for the second. In this case, the power injections of both inverters have changed to $Q_1^{'}$ and $Q_2^{'}$, respectively, which evidences the deterioration of the reactive power sharing, because in this case, these values in both DG's are more distant than before the secondary control application.

The accuracy of reactive power sharing depends on the upper and the lower limits of the DG voltage magnitudes, and of the homogeneity of the transmission line reactances. An ideal secondary voltage controller, should ensure a compromise between voltage regulation and the reactive power sharing. In this sense, we propose a secondary controller based on information theory approach, as we describe in Section A.3.

A.2.3 Tertiary Control

Since this kind of controller is out of the scope of this work, we only mention its main purpose. The tertiary controller, is associated with a global economic dispatch. It is possible to use different techniques to solve the economic dispatch problem in microgrids, such as dynamic population games, as exposed in [72].

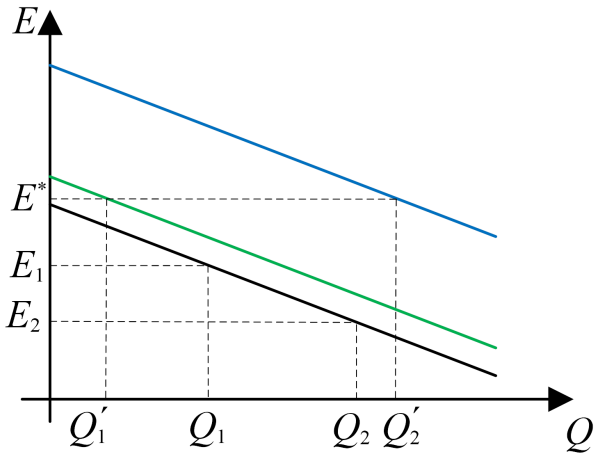


Figure A.2: E-Q "Droop" Controller.

A.3 Information Theory Based Model for Reactive Power Sharing

In Figure A.3, we summarize the proposed multi-agent learning model. As we have mentioned, it combines the rate distortion function, the maximum entropy principle, and the replicator dynamics approach. The agent decision rule, is determined by a fitness function that depends on its knowledge about the environment, which is calculated through the distortion measure $\mathcal{L}(x,y)$ associated to the rate distortion function, where the agent information is represented by y, and the environment information is represented by x. This means that a reduced distortion about the environment implies a good fitness or utility for an agent.

For the microgrid case, the scheme depicted in Figure A.3 represents the modules composing the DG controller. Let us describe it through the following steps:

1. Consider a microgird composed of N DG's, supported by a communications network to allow the information interchange. The neighborhood of a DG_i, defined by the set $\mathcal{N}_i = [DG_1, ..., DG_M]$, is composed of all the DG's having a communication channel towards DG_i. The environment information of a DG_i, which is obtained through the intrinsic communications network, contains the reactive power data of $\mathcal{N}_i$, which we denote as $Q = [Q_1, ...Q_M]$.

2. The expected value $\langle Q_{DG} \rangle$, is determined by the reactive power at which each DG normally should operate.

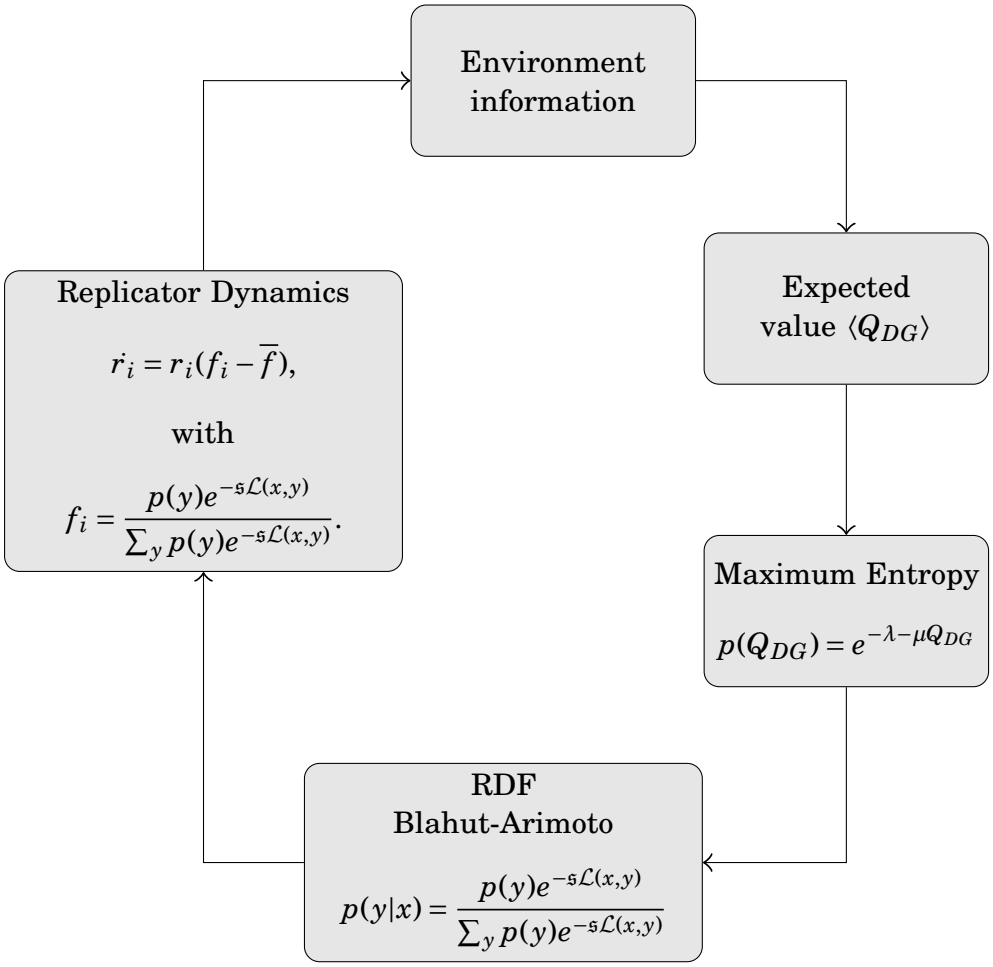


Figure A.3: Information Theory Based Learning Model.

3. By means of the maximum entropy principle, described in Section 2.2.1.2, the model calculates a probability distribution $p(Q_{DG})$ through the expression

$$\begin{aligned} \underset{p(Q_{DG})}{\text{maximize}} \quad & \sum_M p(Q_{DG}) \log p(Q_{DG}) \\ \text{subject to} \quad & \sum_M p(Q_{DG}) Q_{DG} = \langle Q_{DG} \rangle \\ & \sum_M p(Q_{DG}) = 1 \\ & p(Q_{DG}) \geq 0, \end{aligned} \tag{A.5}$$

which describes the deviation of each DG in relation to the expected value $\langle Q_{DG} \rangle$.

4. The resulting probability $p(Q_{DG})$, is used as the source distribution in the Blahut-Arimoto algorithm. At this point, we make $p(x) = p(Q_{DG})$, since we assume $p(Q_{DG})$

as the distribution describing the environment.

5. The Blahut-Arimoto algorithm, described in Section 2.2.1, is used to measure the similitude of each DG and the environment behavior, which is determined by the conditional distribution[1]

$$p(y|x) = \frac{p(y)e^{-\mathfrak{s}\mathcal{L}(x,y)}}{\sum_y p(y)e^{-\mathfrak{s}\mathcal{L}(x,y)}}. \tag{A.6}$$

6. Finally, the similitude between the DG an its environment, define the fitness function f_i, and the replicator dynamics equation

$$\dot{r_i} = r_i(f_i - \overline{f}), \tag{A.7}$$

with

$$f_i = \frac{p(y)e^{-\mathfrak{s}\mathcal{L}(x,y)}}{\sum_y p(y)e^{-\mathfrak{s}\mathcal{L}(x,y)}}, \tag{A.8}$$

where r_i is the proportion of the population assuming the strategy followed by DG_i, and $\dot{r_i}$ represents its variation in time.

A.4 Model Implementation

In this section, we describe the characteristics of the primary and the secondary controllers, as well as the microgrid used to evaluate the effects of the proposed model on the reactive power sharing.

A.4.1 Primary Control

Although the main interest of this approach is focused in the secondary controller, we give a short description of the primary controller because of its influence. This is shown in Figure A.4. In order to model the inverters, they are represented by controlled-voltage sources, since in islanded operation, each inverter acts as a VSI (voltage source inverter), to control the exported and the imported power to and from the conventional power network to stabilize the microgrid [42]. The main idea behind the *"droop" controllers* is to imitate the behavior of a synchronous machine, which in this case, reduces the frequency when the active power load increases, and reduces the voltage magnitude, when the reactive power increases.

[1] As stated in Section 2.1, x and y belong to the alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, which in the case of the microgrid, correspond to a set of possible values for the reactive power.

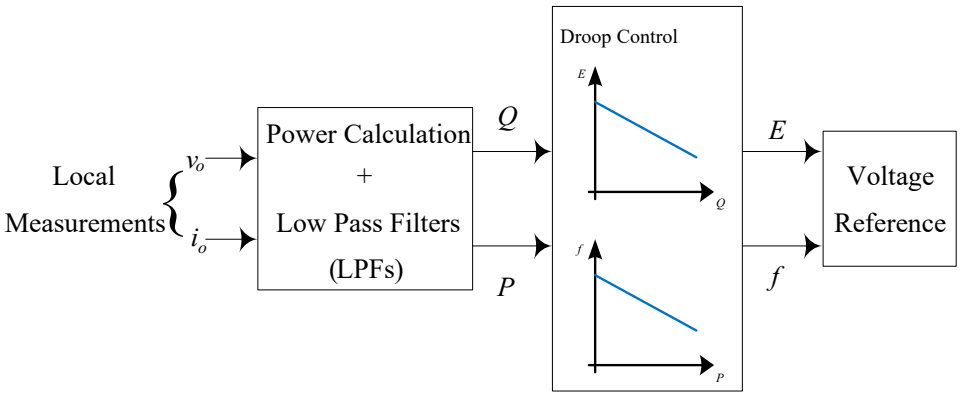


Figure A.4: Primary Control.

A.4.2 Secondary Controller

The secondary controller is used to compensate the deviations for frequency and voltage magnitude, ensuring that they tend to zero after a change in load or generation in the microgrid. A detailed diagram of the implemented secondary control for voltage, in the case of a single DG, is depicted in Figure A.5. It is important to mention that although the primary frequency control of each DG must be considered to obtain the reference voltage signal, no secondary frequency control action is performed.

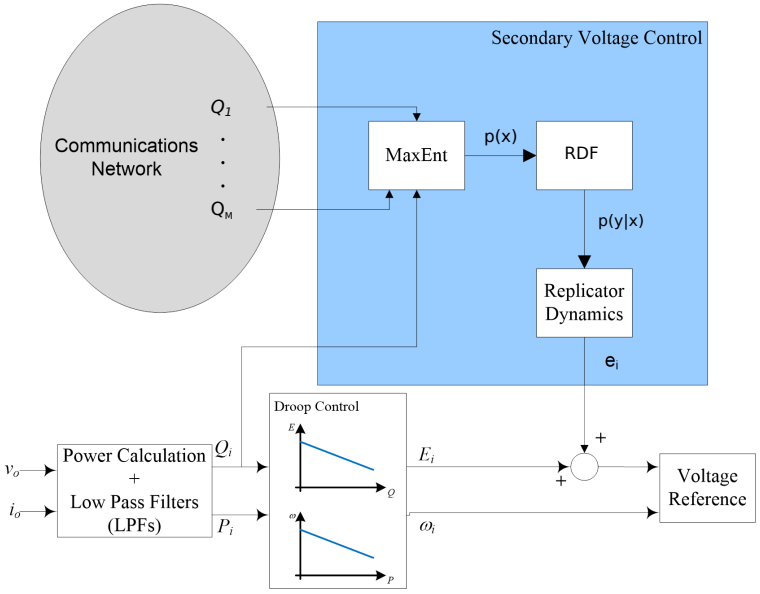


Figure A.5: Secondary Control.

A.4.3 Microgrid Model

The microgrid used to simulate the behavior of the proposed *information theory based controller*, is shown in Figure A.1, which is composed of four inverters, three interconnection lines and two loads. The lines are modeled as RL branches connected in series, the loads are connected to units 1 and 4, and are modeled as constant power devices. In Table A.1, we provide the most relevant parameters. Additional information is reported in [96].

Table A.1: Microgrid Parameters

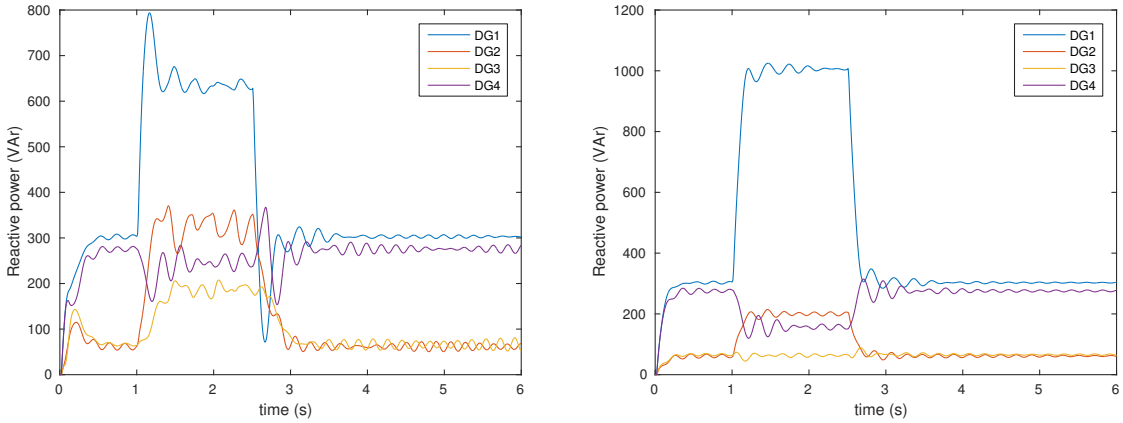
Parameter	**Value**
Nominal frequency	50 Hz
DC Voltage	650 V
AC Voltage	325.3 V
Filter capacitance	25 μF
Filter inductance	1.8 mH
Output inductance	1.8 mH
Line Impedance Z_{12}	0.8+j1.131 Ω
Line Impedance Z_{23}	0.4+j0.565 Ω
Line Impedance Z_{34}	0.7+j0.597 Ω
m_i	2.5×10^{-3} $\frac{\text{rad/s}}{\text{W}}$
n_i	1.5×10^{-3} $\frac{\text{V}}{\text{VAr}}$

A.5 Results

The simulation results show the voltage and reactive power behavior on each DG when the load is abruptly duplicated in the DG_1 during the time interval from t=1s to t=2.5s. In Figure A.6, we show the effect of the proposed controller on the reactive power sharing. In Figure A.6a, we show how the *information theory based controller* reduces the reactive power in the DG_1, which is distributed to the others. On the other hand, in Figure A.6b, we show that the reactive power is poorly shared.

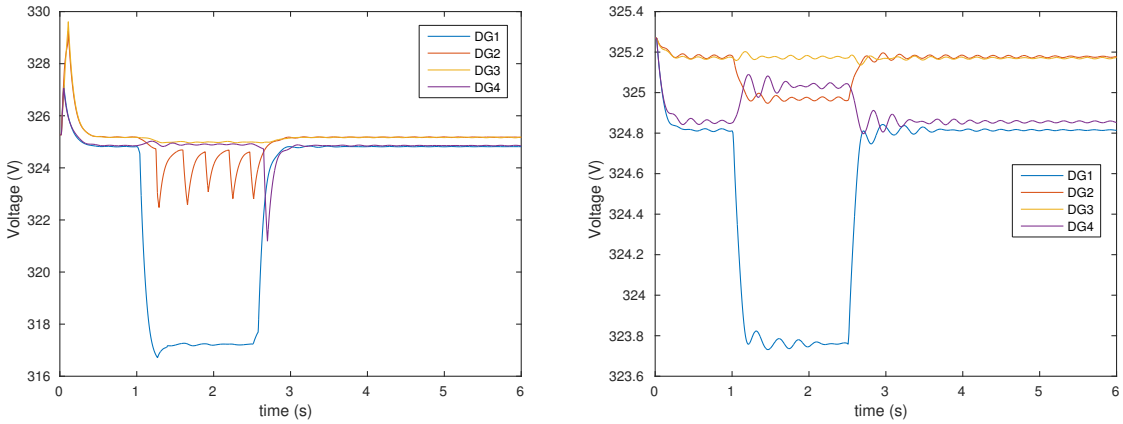
In contrast, a comparison between Figures A.7a and A.7b, illustrates a non-significant difference between the voltage magnitudes before and after applying the proposed secondary control when the load increase occurs, which means that there is not a considerable voltage variation.

Finally, in Table A.2, we show the relationship between the maximum reactive power values when the MaxEnt principle is included or not in the proposed controller. In the



(a) Reactive power sharing using secondary control. (b) Reactive power sharing without secondary control.

Figure A.6: The effect of the secondary controller on the reactive power sharing.



(a) Voltage response using secondary control. (b) Voltage response without secondary control.

Figure A.7: The effect of the secondary controller on the voltage.

first case, when the MaxEnt module is included, the maximum reactive power peak, reached when the load increases in the DG_1, is higher in relation to the value obtained when this module is ignored, i.e., the RDF stage assumes $p(x)$ as a uniform distribution. This result demonstrates that in addition to the distortion reduction between the DG behavior and its environment, provided by the RDF, there is an extra reduction in uncertainty produced by the MaxEnt principle, which is demonstrated in Figure A.8. Notice how the MaxEnt module produces and extra reduction in the distortion value when the DG_1 is overloaded, which impacts favorably in the reactive power sharing.

Table A.2: Maximum DG_1 peaks and times in overload event

	Maximum Q (VAr)	**time (s)**
Without controller	1025	1.1629
Using MaxEnt	779	1.1629
Without MaxEnt	794	1.1652

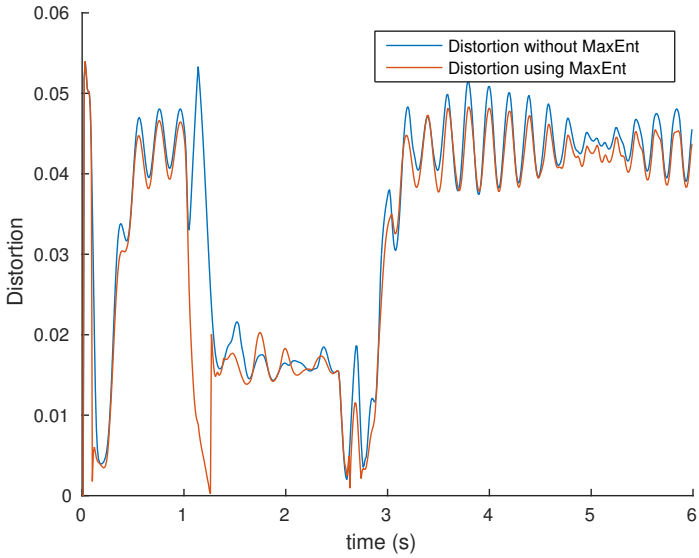


Figure A.8: Distortion when MaxEnt is included or not in the controller.

A.6 Conclusions

We implemented a secondary voltage controller based on information theory concepts such as the rate distortion function and the maximum entropy in order to define a *distortion based fitness function*. The distributed control is implemented by means of the replicator dynamics approach, which uses the previously calculated fitness function to define the strategy selection of each distributed generator (DG) of a microgrid. The implemented controller improves considerably the power sharing condition between all of the DG's operating in an islanded mode, specially when an abrupt load increase is applied to the system.

When the MaxEnt module is included in the controller, the distortion is reduced and consequently the reactive power peaks decrease when abrupt changes in load are present in any of the DG's.

On the other hand, in terms of the voltage magnitudes, in spite of the fact that they are not directly controlled, there is not a considerable degradation in their behavior

when the *information theory based controller* is used to obtain a reactive power sharing. It demonstrates that it is possible to use the proposed control technique without compromising the microgrid performance.

APPENDIX B

KULLBACK-LEIBLER DIVERGENCE OF TWO GAUSSIAN DISTRIBUTIONS

As we described in Section 2.1, the Kullback-Leibler divergence, also known as the relative entropy, is a measure of the *distance* between two probability distributions. It is commonly used to find the *gain or loss of information* obtained for describing with a distribution Q a random variable X whose original distribution is P. This difference is given by

$$KL(P||Q) = \sum P \log \frac{P}{Q} \; [bits], \tag{B.1}$$

which is always non-negative and equal to zero if and only if $P = Q$.

B.1 Kullback-Leibler Divergence of Two Gaussian Distributions

Theorem B.1. *Let* $p(x) = \mathcal{N}(\mu_1, \sigma_1)$ *and* $q(x) = \mathcal{N}(\mu_2, \sigma_2)$*, then*

$$KL(p||q) = \frac{1}{2} \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{1}{2\sigma_2{}^2} \left[\sigma_1{}^2 + (\mu_1 - \mu_2)^2 \right] \tag{B.2}$$

Proof. Using (B.1),

$$KL(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \tag{B.3}$$

$$= \int p(x) \log \frac{\frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{(x-\mu_1)}{2\sigma_1{}^2}^2}}{\frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{(x-\mu_2)}{2\sigma_2{}^2}^2}} dx \tag{B.4}$$

$$= \int p(x) \log \sqrt{\frac{\sigma_2{}^2}{\sigma_1{}^2}} dx \tag{B.5}$$

$$+ \int p(x) \left[\frac{-(x-\mu_1)^2}{2\sigma_1{}^2} + \frac{(x-\mu_2)^2}{2\sigma_2{}^2} \right] dx \tag{B.6}$$

$$= \frac{1}{2} \log \frac{\sigma_2}{\sigma_1} + \frac{1}{2\sigma_1{}^2} \left[- \int (x-\mu_1)^2 p(x) dx \right] \tag{B.7}$$

$$+ \frac{1}{2\sigma_2{}^2} \int (x-\mu_2)^2 p(x) dx \tag{B.8}$$

$$= \frac{1}{2} \log \frac{\sigma_2}{\sigma_1} - \frac{\sigma_1{}^2}{2\sigma_1{}^2} \tag{B.9}$$

$$+ \frac{1}{2\sigma_2{}^2} \int (x-\mu_1+\mu_1-\mu_2)^2 p(x) dx \tag{B.10}$$

$$= \frac{1}{2} \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{1}{2\sigma_2{}^2} \left[\int (x-\mu_1)^2 p(x) dx \right. \tag{B.11}$$

$$+ (\mu_1-\mu_2)^2 \int p(x) dx \tag{B.12}$$

$$\left. + 2(\mu_1-\mu_2) \int (x-\mu_1) p(x) dx \right] \tag{B.13}$$

$$= \frac{1}{2} \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{1}{2\sigma_2{}^2} \left[\sigma_1{}^2 + (\mu_1-\mu_2)^2 \right] \tag{B.14}$$

■

APPENDIX C

STEPS TO MINIMIZE THE MUTUAL INFORMATION

As we stated in Section 2.2, if we ignore temporarily the inequality restriction $p(y|x) \geq 0$, the rate distortion function can be described by the expression

$$\begin{aligned} R(D) = \underset{p(y|x)}{\text{minimize}} \Bigg[& \sum_{x,y} p(y|x)p(x) \log \frac{p(y|x)}{p(y)} \\ & - s\Big(\sum_{x,y} p(y|x)p(x)\mathcal{L}(x,y) - D\Big) \\ & + \sum_{x} \lambda_x \Big(\sum_{y} p(y|x) - 1\Big) \Bigg]. \end{aligned} \tag{C.1}$$

If we make

$$\begin{aligned} J[p(y|x), p(y)] = & \sum_{x,y} p(y|x)p(x) \log \frac{p(y|x)}{p(y)} \\ & - s \sum_{x,y} p(y|x)p(x)\mathcal{L}(x,y) + \sum_{x} \lambda_x \sum_{y} p(y|x), \end{aligned} \tag{C.2}$$

then

$$R(D) = sD - \lambda + \min_{p(y|x)} \min_{p(y)} J[p(y|x), p(y)]. \tag{C.3}$$

The above double minimization problem, is solved through the next two steps:

1. For fixed $p(y)$: Since $I(x,y)$ is a convex function for $p(y|x)$, we have

$$\begin{aligned} R(D) = & \frac{\partial}{\partial p(y|x)} \Big[sD - \lambda + J[p(y|x), p(y)] \Big] \\ & = p(x) \log p(y|x) + p(x) - p(x) \log p(y) \\ & - s p(x) \mathcal{L}(x,y) + \lambda = 0. \end{aligned} \tag{C.4}$$

If we make $\log u = 1 + \frac{\lambda}{p(x)}$, then

$$p(y|x) = \frac{p(y)e^{s\mathcal{L}(x,y)}}{u}. \tag{C.5}$$

Since $\sum_y p(y|x) = 1$,

$$u = \sum_y p(y)e^{s\mathcal{L}(x,y)}, \tag{C.6}$$

i.e., λ is selected to accomplish the condition $\sum_y p(y|x) = 1$. Therefore, the initially ignored condition $p(y|x) \geq 0$ is also satisfied.

Hence,

$$p^*(y|x) = \frac{p(y)e^{s\mathcal{L}(x,y)}}{\sum_y p(y)e^{s\mathcal{L}(x,y)}}. \tag{C.7}$$

2. For fixed $p(y|x)$: In this case, it is enough to prove that

$$I(X;Y) = \min_{p(y)} J[p(y)]. \tag{C.8}$$

Then

$$I(X;Y) - \sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p^*(y)} = 0 \tag{C.9}$$

$$\tag{C.10}$$

since $I(X;Y) = \sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p^(y)}$, we have

$$\begin{aligned} &\sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p^(y)} \\ &-\sum_{x,y} p(y|x)p(x)\log\frac{p(y|x)}{p^*(y)} = 0 \end{aligned} \tag{C.11}$$

so that

$$\sum_{x,y} p(x)p(y|x)\log\frac{p^*(y)}{p(y)} = 0. \tag{C.12}$$

Setting $\sum_x p(x)p(y|x) = p^*(y)$ and using the Kullback-Leibler divergence (see appendix B), we obtain

$$\sum_y p^*(y)\log\frac{p^*(y)}{p(y)} = KL(p^*||p) = 0, \tag{C.13}$$

which means that the equality is satisfied if and only if the probability distributions p and p^* are the same. Therefore,

$$p^*(y) = \sum_x p(x)p(y|x). \tag{C.14}$$

BIBLIOGRAPHY

[1] Z. ABRAMS, A. GOEL, AND S. PLOTKIN, *Set k-cover algorithms for energy efficient monitoring in wireless sensor networks*, in Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks, IPSN '04, New York, NY, USA, 2004, Association for Computing Machinery, p. 424–432.

[2] X. AI, V. SRINIVASAN, AND C.-K. THAM, *Optimality and Complexity of Pure Nash Equilibria in the Coverage Game*, IEEE Journal on Selected Areas in Communications, 26 (2008), pp. 1170–1182.

[3] R. B. ALBERT, *Topology of evolving networks: Local events and universality*, Physical Review Letters, 24 (2000), pp. 5234–5237.

[4] T. ALSKAIF, M. GUERRERO ZAPATA, AND B. BELLALTA, *Game theory for energy efficiency in Wireless Sensor Networks: Latest trends*, Journal of Network and Computer Applications, 54 (2015), pp. 33–61.

[5] K. ANAND AND G. BIANCONI, *Entropy measures for networks: Toward an information theory of complex topologies*, Physical Review E, 80 (2009), p. 045102.

[6] L.-M. ANG AND K. PHOOI SENG, *Big Sensor Data Applications in Urban Environments*, Big Data Research, 4 (2016), pp. 1–12.

[7] E. ANSHELEVICH, A. DASGUPTA, J. KLEINBERG, E. TARDOS, T. WEXLER, AND T. ROUGHGARDEN, *The Price of Stability for Network Design with Fair Cost Allocation*, in 45th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2004, pp. 295–304.

[8] V. AULETTA, D. FERRAIOLI, F. PASQUALE, P. PENNA, AND G. PERSIANO, *Convergence to Equilibrium of Logit Dynamics for Strategic Games*, Algorithmica, 76 (2016), pp. 110–142.

[9] Z. BAO, Y. CAO, L. DING, Z. HAN, AND G. WANG, *Dynamics of load entropy during cascading failure propagation in scale-free networks*, Physics Letters A, 372 (2008), pp. 5778–5782.

[10] P. BAROOAH AND J. HESPANHA, *Estimation on graphs from relative measurements*, IEEE Control Systems, 27 (2007), pp. 57–74.

[11] J. BARREIRO-GOMEZ, G. OBANDO, AND N. QUIJANO, *Distributed Population Dynamics: Optimization and Control Applications*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, (2016), pp. 1–11.

[12] M. BARTHÉLEMY, *Spatial networks*, Institut de Physique Théorique, (2014).

[13] D. P. BERTSEKAS, *Nonlinear Programming: 3rd Edition*, Athena Scientific, 2016.

[14] J. BINNEY, A. KRAUSE, AND G. S. SUKHATME, *Informative path planning for an autonomous underwater vehicle*, in 2010 IEEE International Conference on Robotics and Automation, IEEE, May 2010, pp. 4791–4796.

[15] R. BLAHUT, *Computation of channel capacity and rate-distortion functions*, IEEE Transactions on Information Theory, 18 (1972), pp. 460–473.

[16] D. BLOEMBERGEN, K. TUYLS, D. HENNES, AND M. KAISERS, *Evolutionary dynamics of multi-agent learning: A survey*, 2015.

[17] L. E. BLUME, *The Statistical Mechanics of Strategic Interaction*, Games and Economic Behavior, 5 (1993), pp. 387–424.

[18] S. BOLOGNANI, R. CARLI, G. CAVRARO, AND S. ZAMPIERI, *Distributed reactive power feedback control for voltage regulation and loss minimization*, IEEE Transactions on Automatic Control, 60 (2015), pp. 966–981.

[19] C. G. CASSANDRAS AND W. LI, *Sensor Networks and Cooperative Control*, European Journal of Control, 11 (2005), pp. 436–463.

[20] M. C. CHANDORKAR, D. M. DIVAN, AND R. ADAPA, *Control of parallel connected inverters in standalone ac supply systems*, IEEE Transactions on Industry Applications, 29 (1993), pp. 136–143.

[21] C.-Y. CHANG, G. CHEN, G.-J. YU, T.-L. WANG, AND T.-C. WANG, *TCWTP: Time-Constrained Weighted Targets Patrolling Mechanism in Wireless Mobile Sensor*

Networks, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45 (2015), pp. 901–914.

[22] L. CHEN, S. ARAKAWA, H. KOTO, N. OGINO, H. YOKOTA, AND M. MURATA, *An evolvable network design approach with topological diversity*, Computer Communications, 76 (2016), pp. 101–110.

[23] M. CHEN, S. MAO, Y. LIU, M. CHEN, S. MAO, AND Y. LIU, *Big Data: A Survey*, Mobile Netw Appl, 19 (2014), pp. 171–209.

[24] Y. CHENG, X. LI, Z. LI, S. JIANG, AND X. JIANG, *Fine-grained air quality monitoring based on gaussian process regression*, in Neural Information Processing, C. K. Loo, K. S. Yap, K. W. Wong, A. Teoh, and K. Huang, eds., Cham, 2014, Springer International Publishing, pp. 126–134.

[25] T. CLOUQUEUR, V. PHIPATANASUPHORN, P. RAMANATHAN, AND K. K. SALUJA, *Sensor Deployment Strategy for Detection of Targets Traversing a Region*, Mobile Networks and Applications, 8 (2003), pp. 453–461.

[26] J. CORTES, S. MARTINEZ, T. KARATAS, AND F. BULLO, *Coverage Control for Mobile Sensing Networks*, IEEE Transactions on Robotics and Automation, 20 (2004), pp. 243–255.

[27] T. M. COVER AND J. A. THOMAS, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, Wiley-Interscience, New York, NY, USA, 2006.

[28] F. F. D. ACEMOGLU, G. COMO AND A. OZDAGLAR, *Opinion fluctations and disagreement in social networks*, Mathematics of Operations Reserch, 38 (2013), pp. 1–27.

[29] M. H. DE BADYN, A. CHAPMAN, AND M. MESBAHI, *Network entropy: A system-theoretic perspective*, in Proceedings of the IEEE Conference on Decision and Control, vol. 2016-Febru, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 5512–5517.

[30] M. DEHMER, *A NOVEL METHOD FOR MEASURING THE STRUCTURAL INFORMATION CONTENT OF NETWORKS*, Cybernetics and Systems, 39 (2008), pp. 825–842.

[31] L. DEMETRIUS, V. M. GUNDLACH, AND G. OCHS, *Complexity and demographic stability in population models.*, Theoretical population biology, 65 (2004), pp. 211–25.

[32] L. DEMETRIUS AND T. MANKE, *Robustness and network evolution?an entropic principle*, Physica A: Statistical Mechanics and its Applications, 346 (2005), pp. 682–696.

[33] C. DESOER, *The Optimum Formula for the Gain of a Flow Graph or a Simple Derivation of Coates' Formula*, Proceedings of the IRE, 48 (1960), pp. 883–889.

[34] S. DHILLON AND K. CHAKRABARTY, *Sensor placement for effective coverage and surveillance in distributed sensor networks*, in 2003 IEEE Wireless Communications and Networking, 2003. WCNC 2003., vol. 3, IEEE, 2003, pp. 1609–1614.

[35] M. C. DONALDSON-MATASCI, C. T. BERGSTROM, AND M. LACHMANN, *The fitness value of information.*, Oikos (Copenhagen, Denmark), 119 (2010), pp. 219–230.

[36] DONGBING GU AND HUOSHENG HU, *Spatial Gaussian Process Regression With Mobile Sensor Networks*, IEEE Transactions on Neural Networks and Learning Systems, 23 (2012), pp. 1279–1290.

[37] E. ESTRADA, *The structure of complex networks: theory and applications*, Oxford University Press, 2011.

[38] S. GONZÁLEZ-VALENZUELA, M. CHEN, AND V. C. M. LEUNG, *Mobility support for health monitoring at home using wearable sensors*, IEEE Transactions on Information Technology in Biomedicine, 15 (2011), pp. 539–549.

[39] J. P. GOULD, *Risk, stochastic preference, and the value of information*, Journal of Economic Theory, 8 (1974), pp. 64–84.

[40] C. GROS, *Complex and adaptive dynamical systems*, A Primer. Springer, (2008).

[41] T. GROSS AND H. SAYAMA, *Adaptive networks*, Springer, 2009.

[42] J. M. GUERRERO, M. CHANDORKAR, T. LEE, AND P. C. LOH, *Advanced control architectures for intelligent microgrids—part i: Decentralized and hierarchical control*, IEEE Trans. Ind. Electron, 60 (2013), pp. 1254–1262.

[43] M. HATA, *Empirical formula for propagation loss in land mobile radio services*, IEEE Transactions on Vehicular Technology, 29 (1980), pp. 317–325.

[44] J. HE AND A. KOLOVOS, *Bayesian maximum entropy approach and its applications: a review*, Stochastic Environmental Research and Risk Assessment, (2017), pp. 1–19.

[45] M. O. JACKSON, *Social and economic networks*, Physical Review Letters, 24 (2000), pp. 5234–5237.

[46] E. T. JAYNES, *Information Theory and Statistical Mechanics*, Physical Review, 106 (1957), pp. 620–630.

[47] B. J. JULIAN, M. ANGERMANN, M. SCHWAGER, AND D. RUS, *Distributed robotic sensor networks: An information-theoretic approach*, The International Journal of Robotics Research, 31 (2012), pp. 1134–1154.

[48] A. KAHN, J. MARZAT, H. PIET-LAHANIER, AND M. KIEFFER, *Global extremum seeking by Kriging with a multi-agent system*, IFAC-PapersOnLine, 48 (2015), pp. 526–531.

[49] A. B. KAO, N. MILLER, C. TORNEY, A. HARTNETT, AND I. D. COUZIN, *Collective learning and optimal consensus decisions in social animal groups.*, PLoS computational biology, 10 (2014), p. e1003762.

[50] A. KRAUSE, A. SINGH, AND C. GUESTRIN, *Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies*, Journal of Machine Learning Research, 9 (2008), pp. 235–284.

[51] S. KUNZ, T. USLÄNDER, AND K. WATSON, *A testbed for sensor service networks and the fusion SOS: Towards plug & measure in sensor networks for environmental monitoring with OGC standards*, in 18th World IMACS / MODSIM Congress, 2009, pp. 973–979.

[52] R. H. LASSETER, *Microgrids*, in IEEE Power Engineering Society Winter Meeting, vol. 1, 2002, pp. 305–308.

[53] N. LI AND J. R. MARDEN, *Designing games for distributed optimization*, IEEE Journal of Selected Topics in Signal Processing, 7 (2013), pp. 230–242.

[54] S. LI, L. D. XU, AND X. WANG, *Compressed Sensing Signal and Data Acquisition in Wireless Sensor Networks and Internet of Things*, IEEE Transactions on Industrial Informatics, 9 (2013), pp. 2177–2186.

[55] W. LI, *Mutual information functions versus correlation functions*, Journal of Statistical Physics, 60 (1990), pp. 823–837.

[56] B. LIU, P. BRASS, O. DOUSSE, P. NAIN, AND D. TOWSLEY, *Mobility improves coverage of sensor networks*, in Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '05, New York, NY, USA, 2005, ACM, pp. 300–308.

[57] B. LIU, O. DOUSSE, P. NAIN, AND D. TOWSLEY, *Dynamic Coverage of Mobile Sensor Networks*, IEEE Transactions on Parallel and Distributed Systems, 24 (2013), pp. 301–311.

[58] J. A. P. LOPES, C. L. MOREIRA, AND A. G. MADUREIRA, *Defining control strategies for microgrids islanded operation*, IEEE Transactions on Power Systems, 21 (2006), pp. 916–924.

[59] X. MA, H. YU, Y. WANG, AND Y. WANG, *Large-scale transportation network congestion evolution prediction using deep learning theory*, in PloS one, 2015.

[60] J. MARDEN, G. ARSLAN, AND J. SHAMMA, *Cooperative Control and Potential Games*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39 (2009), pp. 1393–1407.

[61] J. R. MARDEN, *The role of information in multiagent coordination*, in 53rd IEEE Conference on Decision and Control, 2014, pp. 445–450.

[62] D. A. MARTÍNEZ AND E. MOJICA-NAVA, *Correlation as a measure for fitness in multi-agent learning systems*, in 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE, Nov 2016, pp. 1–6.

[63] ——, *Entropy measures in evolving networks*, in Complex Networks: from theory to interdisciplinary applications, July 2016.

[64] ——, *Graph transfer function representation to measure network robustness*, in Impact and Advances of Automatic Control in Latinamerica, Oct 2016, pp. 172–176.

[65] D. A. MARTÍNEZ, E. MOJICA-NAVA, K. WATSON, AND T. USLÄNDER, *Multi-agent learning framework for environment redundancy identification for mobile sensors in an iot context*, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W11 (2018), pp. 33–41.

[66] D. A. MARTINEZ, R. RINCON, E. MOJICA-NAVA, AND A. PAVAS, *Reactive power sharing in microgrids: An information-theoretical approach*, in 2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC), IEEE, Oct 2017, pp. 1–6.

[67] D. A. MARTÍNEZ, E. MOJICA-NAVA, A. S. AL-SUMATI, AND S. RIVERA, *A distortion-based potential game for secondary voltage control in micro-grids*, IEEE Access, (2020), pp. 1–1.

[68] J. M. MCNAMARA AND S. R. X. DALL, *Information is a fitness enhancing resource*, Oikos, 119 (2010), pp. 231–236.

[69] S. MEI, H. LI, J. FAN, X. ZHU, AND C. R. DYER, *Inferring air pollution by sniffing social media*, in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Aug 2014, pp. 534–539.

[70] M. MESBAHI AND M. EGERSTEDT, *Graph Theoretic Methods in Multiagent Networks*, Princeton University Press, stu - student edition ed., 2010.

[71] S. M. MOHAMED, H. S. HAMZA, AND I. A. SAROIT, *Coverage in mobile wireless sensor networks (M-WSN): A survey*, Computer Communications, 110 (2017), pp. 133–150.

[72] E. MOJICA-NAVA, C. BARRETO, AND N. QUIJANO, *Population games methods for distributed control of microgrids*, IEEE Transactions on Smart Grid, 6 (2015), pp. 2586 – 2595.

[73] D. MONDERER AND L. S. SHAPLEY, *Potential Games*, Games and Economic Behavior, 14 (1996), pp. 124–143.

[74] K. P. MURPHY, *Machine Learning: A Probabilistic Perspective*, The MIT Press, 2012.

[75] M. NEWMAN, *Networks: an introduction*, Oxford University Press, 2010.

[76] L. V. NGUYEN, S. KODAGODA, R. RANASINGHE, AND G. DISSANAYAKE, *Information-Driven Adaptive Sampling Strategy for Mobile Robotic Wireless Sensor Network*, IEEE Transactions on Control Systems Technology, 24 (2016), pp. 372–379.

[77] M. NOWAK, *Evolutionary Dynamics*, Harvard University Press, 2006.

[78] S. OH, Y. XU, AND J. CHOI, *Explorative navigation of mobile sensor networks using sparse Gaussian processes*, in 49th IEEE Conference on Decision and Control (CDC), IEEE, Dec 2010, pp. 3851–3856.

[79] H. OHTSUKI AND M. A. NOWAK, *The replicator equation on graphs.*, Journal of theoretical biology, 243 (2006), pp. 86–97.

[80] R. OLFATI, J. A. FAX, AND R. MURRAY, *Consensus and cooperation in networked multi-agent systems*, Proc. IEEE, 95 (2007), pp. 215–233.

[81] B. T. ONG, K. SUGIURA, AND K. ZETTSU, *Dynamic pre-training of deep recurrent neural networks for predicting environmental monitoring data*, in 2014 IEEE International Conference on Big Data (Big Data), Oct 2014, pp. 760–765.

[82] A. PANTOJA AND N. QUIJANO, *Distributed optimization using population dynamics with a local replicator equation*, in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, dec 2012, pp. 3790–3795.

[83] F. PASSERINI AND S. SEVERINI, *The von Neumann entropy of networks*, 2008.

[84] S. RAHILI, J. LU, W. REN, AND U. M. AL-SAGGAF, *Distributed Coverage Control of Mobile Sensor Networks in Unknown Environment Using Game Theory: Algorithms and Experiments*, IEEE Transactions on Mobile Computing, 17 (2018), pp. 1303–1313.

[85] N. RASHEVSKY, *Life, information theory, and topology*, The Bulletin of Mathematical Biophysics, 17 (1955), pp. 229–235.

[86] C. E. RASMUSSEN AND C. K. I. WILLIAMS, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.

[87] R.HEGSELMANN AND U.KRAUSE, *Opinion dynamics and bounded confidence*, Physical Review Letters, 5 (2000).

[88] R. RINCÓN, A. PAVAS, AND E. MOJICA-NAVA, *Long-term voltage stability analysis and network topology in power systems*, in IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2016, pp. 1–6.

[89] S. ROGERS AND M. GIROLAMI, *A First Course in Machine Learning, Second Edition*, Chapman & Hall/CRC, 2nd ed., 2016.

[90] W. H. SANDHOLM, *Population Games and Evolutionary Dynamics*, MIT Press, 2010.

[91] A. SCAGLIONE AND S. SERVETTO, *On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks*, Wireless Networks, 11 (2005), pp. 149–160.

[92] S. SENDRA, E. GRANELL, J. LLORET, AND J. J. P. C. RODRIGUES, *Smart collaborative mobile system for taking care of disabled and elderly people*, Mobile Networks and Applications, 19 (2014), pp. 287–302.

[93] C. E. SHANNON, *A Mathematical Theory of Communication*, Bell System Technical Journal, 27 (1948), pp. 379–423.

[94] J. W. SIMPSON-PORCO, F. DÖRFLER, AND F. BULLO, *Voltage stabilization in microgrids via quadratic droop control*, in IEEE CDC, Florence, Italy, 2013, pp. 7582–7589.

[95] J. W. SIMSON-PORCO, F. DÖRFLER, AND F. BULLO, *Synchronization and power sharing for droop-controlled inverters in islanded microgrid*, Automatica, 49 (2013), pp. 2603–2611.

[96] J. W. SIMSON-PORCO, F. DÖRFLER, Q. SHAFIEE, J. M. GUERRERO, AND F. BULLO, *Secondary frequency and voltage control of islanded microgrids via distributed averaging*, IEEE Transactions on Industrial Electronics, 46 (2015), pp. 1–12.

[97] J. W. SIMSON-PORCO, Q. SHAFIEE, F. DÖRFLER, J. C. VASQUEZ, J. M. GUERRERO, AND F. BULLO, *Stability, power sharing, & distributed secondary control in droop-controlled microgrids*, in IEEE Conf. Smart Grid Comm, Vancouver, BC, Canada, 2013, pp. 672–677.

[98] V. S. SOLÉ R. V., *Complex Networks*, vol. 650 of Lecture Notes in Physics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[99] X. SONG, Q. ZHANG, Y. SEKIMOTO, T. HORANONT, S. UEYAMA, AND R. SHIBASAKI, *Intelligent system for human behavior analysis and reasoning following large-scale disasters*, IEEE Intelligent Systems, 28 (2013), pp. 35–42.

[100] J. A. STANKOVIC, *Research Directions for the Internet of Things*, IEEE Internet of Things Journal, 1 (2014), pp. 3–9.

[101] T. TATARENKO, *Stochastic payoff-based learning in multi-agent systems modeled by means of potential games*, in 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE, dec 2016, pp. 5298–5303.

[102] ——, *Game-Theoretic Learning and Distributed Optimization in Memoryless Multi-Agent Systems*, Springer Publishing Company, Incorporated, 1st ed., 2017.

[103] S. F. TAYLOR, N. TISHBY, AND W. BIALEK, *Information and fitness*, 2007.

[104] N. TISHBY, F. C. PEREIRA, AND W. BIALEK, *The information bottleneck method*, 2000.

[105] E. TRUCCO, *A note on the information content of graphs*, The Bulletin of Mathematical Biophysics, 18 (1956), pp. 129–135.

[106] K. TUMER AND D. WOLPERT, *A Survey of Collectives*, in Collectives and the Design of Complex Systems, Springer New York, New York, NY, 2004, pp. 1–42.

[107] A. VISERAS, T. WIEDEMANN, C. MANSS, L. MAGEL, J. MUELLER, D. SHUTIN, AND L. MERINO, *Decentralized multi-agent exploration with online-learning of Gaussian processes*, in 2016 IEEE International Conference on Robotics and Automation (ICRA), IEEE, May 2016, pp. 4222–4229.

[108] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Science Editions, Princeton University Press, 1944.

[109] M. VOORNEVELD, *Potential games and interactive decisions with multiple criteria*, PhD thesis, Tilburg University, 1999.

[110] G. WANG, G. CAO, AND T. LA PORTA, *Movement-assisted sensor deployment*, IEEE Transactions on Mobile Computing, 5 (2006), pp. 640–652.

[111] H. WANG, H. E. ROMAN, L. YUAN, Y. HUANG, AND R. WANG, *Connectivity, coverage and power consumption in large-scale wireless sensor networks*, Computer Networks, 75 (2014), pp. 212–225.

[112] X. WANG, H. ZHANG, S. FAN, AND H. GU, *Coverage Control of Sensor Networks in IoT Based on RPSO*, IEEE Internet of Things Journal, 5 (2018), pp. 3521–3532.

[113] J. WEBB, *Game Theory: Decisions, Interaction and Evolution*, Springer Undergraduate Mathematics Series, Springer, 2007.

[114] D. WHITNEY, *Basic network metrics-notes*, Notes, (2008).

[115] D. H. WOLPERT, *Information Theory - The Bridge Connecting Bounded Rational Game Theory and Statistical Physics*, Complex Engineered Systems, 2006 (2006), pp. 262–290.

[116] D. H. WOLPERT AND K. TUMER, *An introduction to collective intelligence*, 1999.

[117] W.REN, W. BEARD, AND E. M. ATKINS, *Information consensus in multivehicle cooperative control*, IEEE Control Syst, Mag, 27 (2007), pp. 71–82.

[118] Y.-H. XIAO, W.-T. WU, H. WANG, M. XIONG, AND W. WANG, *Symmetry-based structure entropy of complex networks*, Physica A: Statistical Mechanics and its Applications, 387 (2008), pp. 2611–2619.

[119] Y. XU AND J. CHOI, *Adaptive Sampling for Learning Gaussian Processes Using Mobile Sensor Networks*, Sensors, 11 (2011), pp. 3051–3066.

[120] E. ZEYDAN, D. KIVANC, C. COMANICIU, AND U. TURELI, *Energy-efficient routing for correlated data in wireless sensor networks*, Ad Hoc Networks, 10 (2012), pp. 962–975.

[121] L. ZHANG, *Game Theoretical Algorithm for Coverage Optimization in Wireless Sensor Networks*, in World Congress on Engineering 2008 Vol I WCE 2008, 2008, pp. 764–769.

[122] Y. ZHENG, F. LIU, AND H.-P. HSIEH, *U-air: When urban air quality inference meets big data*, in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, New York, NY, USA, 2013, ACM, pp. 1436–1444.

[123] M. ZHU AND S. MARTINEZ, *Distributed coverage games for mobile visual sensors (II) : Reaching the set of global optima*, in Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference, IEEE, dec 2009, pp. 175–180.