Distributed Strategic Learning for Wireless Engineers

Hamidou Tembine

CRC Press
Taylor & Francis Group

Distributed Strategic Learning for Wireless Engineers

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Distributed Strategic Learning for Wireless Engineers

Hamidou Tembine

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2012 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
Version Date: 20120330

International Standard Book Number: 978-1-4398-7637-4 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Dedicated to Bourere Signipily

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

List of Figures	xv
List of Tables	xvii
Foreword	xix
Preface	xxi
The Author Bio	xxxiii
Contributors	xxxv
1 Introduction to Learning in Games	1
1.1 Basic Elements of Games	5
1.1.1 Basic Components of One-Shot Game	5
1.1.4 State-Dependent One-Shot Game	9
1.1.4.1 Perfectly-Known State One-Shot Games	9
1.1.4.2 One-Shot Games with Partially-Known State	10
1.1.4.3 State Component is Unknown	10
1.1.4.4 Only the State Space Is Known	11
1.1.5 Perfectly Known State Dynamic Game	11
1.1.6 Unknown State Dynamic Games	12
1.1.7 State-Dependent Equilibrium	20
1.1.8 Random Matrix Games	21
1.1.9 Dynamic Robust Game	21
1.2 Robust Games in Networks	22
1.3 Basic Robust Games	27
1.4 Basics of Robust Cooperative Games	29
1.4.0.1 Preliminaries	29
1.4.0.4 Cooperative Solution Concepts	30
1.5 Distributed Strategic Learning	33
1.5.1 Convergence Issue	39
1.5.2 Selection Issue	39
1.5.2.1 How to Select an Efficient Outcome?	40
1.5.2.2 How to Select a Stable Outcome ?	40
1.6 Distributed Strategic Learning in Wireless Networks	41

1.6.1	Physical Layer	41
1.6.2	MAC Layer	42
1.6.3	Network Layer	42
1.6.4	Transport Layer	47
1.6.5	Application Layer	48
1.6.6	Compressed Sensing	49
2	Strategy Learning	53
2.1	Introduction	53
2.2	Strategy Learning under Perfect Action Monitoring	53
2.2.1	Fictitious Play-Based Algorithms	54
2.2.2	Best Response-Based Learning Algorithms	66
2.2.5	Better Reply-Based Learning Algorithms	72
2.2.6	Fixed Point Iterations	77
2.2.7	Cost-To-Learn	80
2.2.8	Learning Bargaining Solutions	86
2.2.9	Learning and Conjectural Variations	91
2.2.10	Bayesian Learning in Games	94
2.2.11	Non-Bayesian Learning	96
2.3	Fully Distributed Strategy-Learning	96
2.3.1	Learning by Experimentation	98
2.3.2	Reinforcement Learning	101
2.3.3	Learning Correlated Equilibria	111
2.3.4	Boltzmann-Gibbs Learning Algorithms	114
2.3.5	Hybrid Learning Scheme	118
2.3.6	Fast Convergence of Evolutionary Dynamics	121
2.3.7	Convergence in Finite Number of Steps	122
2.3.8	Convergence Time of Boltzmann-Gibbs Learning	123
2.3.9	Learning Satisfactory Solutions	127
2.4	Stochastic Approximations	130
2.5	Chapter Review	131
2.6	Discussions and Open Issues	132
3	Payoff Learning and Dynamics	137
3.1	Introduction	137
3.2	Learning Equilibrium Payoffs	140
3.3	Payoff Dynamics	144
3.4	Routing Games with Parallel Links	144
3.5	Numerical Values of Payoffs Are Not Observed	146

4 Combined Learning 149

4.1	Introduction	149
4.2	Model and Notations	152
4.2.1	Description of the Dynamic Game	153
4.2.2	Combined Payoff and Strategy Learning	155
4.3	Pseudo-Trajectory	162
4.3.1	Convergence of the Payoff Reinforcement Learning . .	163
4.3.2	Folk Theorem	163
4.3.3	From Imitative Boltzmann-Gibbs CODIPAS-RL to Replicator Dynamics	165
4.4	Hybrid and Combined Dynamics	166
4.4.1	From Boltzmann-Gibbs-Based CODIPAS-RL to Composed Dynamics	166
4.4.2	From Heterogeneous Learning to Novel Game Dynamics	167
4.4.3	Aggregative Robust Games in Wireless Networks . . .	171
4.4.3.2	Power Allocation as Aggregative Robust Games	172
4.4.4	Wireless MIMO Systems	178
4.4.4.1	Learning the Outage Probability	179
4.4.4.2	Learning the Ergodic Capacity	180
4.5	Learning in Games with Continuous Action Spaces	180
4.5.1	Stable Robust Games	181
4.5.2	Stochastic-Gradient-Like CODIPAS	183
4.6	CODIPAS for Stable Games with Continuous Action Spaces	183
4.6.1	Algorithm to Solve Variational Inequality	184
4.6.2	Convergence to Variational Inequality Solution	185
4.7	CODIPAS-RL via Extremum-Seeking	186
4.8	Designer and Users in an Hierarchical System	188
4.9	From Fictitious Play with Inertia to CODIPAS-RL	191
4.10	CODIPAS-RL with Random Number of Active Players . . .	192
4.11	CODIPAS for Multi-Armed Bandit Problems	197
4.12	CODIPAS and Evolutionary Game Dynamics	200
4.12.1	Discrete-Time Evolutionary Game Dynamics	201
4.12.4	CODIPAS-Based Evolutionary Game Dynamics . . .	201
4.13	Fastest Learning Algorithms	202

5 Learning under Delayed Measurement 207

5.1	Introduction	207
5.2	Learning under Delayed Imperfect Payoffs	208
5.2.1	CODIPAS-RL under Delayed Measurement	209
5.3	Reacting to the Interference	212
5.3.1	Robust PMAC Games	214
5.3.2	Numerical Examples	216

5.3.2.1	Two Receivers	216
5.3.2.2	Three Receivers	216
5.3.3	MIMO Interference Channel	218
5.3.3.1	One-Shot MIMO Game	222
5.3.4.1	MIMO Robust Game	225
5.3.4.5	Without Perfect CSI	227
6	Learning in Constrained Robust Games	231
6.1	Introduction	231
6.2	Constrained One-Shot Games	231
6.2.1	Orthogonal Constraints	231
6.2.2	Coupled Constraints	232
6.3	Quality of Experience	233
6.4	Relevance in QoE and QoS satisfaction	234
6.5	Satisfaction Levels as Benchmarks	235
6.6	Satisfactory Solution	236
6.7	Efficient Satisfactory Solution	237
6.8	Learning a Satisfactory Solution	237
6.8.3	Minkowski-Sum of Feasible Sets	239
6.9	From Nash Equilibrium to Satisfactory Solution	239
6.10	Mixed and Near-Satisfactory Solution	240
6.11	CODIPAS with Dynamic Satisfaction Level	242
6.12	Random Matrix Games	243
6.12.1	Random Matrix Games Overview	244
6.12.2	Zero-Sum Random Matrix Games	245
6.12.4	NonZero Sum Random Matrix Games	247
6.12.5.1	Relevance in Networking and Communication	248
6.12.7	Evolutionary Random Matrix Games	250
6.12.8	Learning in Random Matrix Games	250
6.12.9	Mean-Variance Response	251
6.12.11	Satisfactory Solution	253
6.13	Mean-Variance Response and Demand Satisfaction	253
7	Learning under Random Updates	255
7.1	Introduction	255
7.2	Description of the Random Update Model	258
7.2.1	Description of the Dynamic Robust Game	260
7.3	Fully Distributed Learning	263
7.3.1	Distributed Strategy-Reinforcement Learning	263
7.3.2	Random Number of Interacting Players	269
7.3.3	CODIPAS-RL for Random Updates	273
7.3.4	Learning Schemes Leading to Multi-Type Replicator Dynamics	274

7.3.5	Heterogeneous Learning with Random Updates	276
7.3.6	Constant Step-Size Random Updates	279
7.3.7	Revision Protocols with Random Updates	279
7.4	Dynamic Routing Games with Random Traffic	280
7.5	Extensions	282
7.5.1	Learning in Stochastic Games	282
7.5.2.1	Nonconvergence of Fictitious Play	283
7.5.2.3	Q-learning in Zero-Sum Stochastic Games	284
7.5.3	Connection to Differential Dynamic Programming	286
7.5.4	Learning in Robust Population Games	286
7.5.4.1	Connection with Mean Field Game Dynamics	286
7.5.5	Simulation of Population Games	291
7.6	Mobility-Based Learning in Cognitive Radio Networks	292
7.6.1	Proposed Cognitive Network Model	295
7.6.2	Cognitive Radio Network Model	296
7.6.2.1	Mobility of Users	296
7.6.3	Power Consumption	298
7.6.4	Virtual Received Power	300
7.6.5	Scaled SINR	300
7.6.6	Asymptotics	301
7.6.8	Performance of a Generic User	303
7.6.8.1	Access Probability	303
7.6.8.3	Coverage Probability	305
7.7	Hybrid Strategic Learning	308
7.7.1	Learning in a Simple Dynamic Game	309
7.7.1.1	Learning Patterns	309
7.7.1.2	Description of CODIPAS Patterns	310
7.7.1.3	Asymptotics of Pure Learning Schemes	311
7.7.1.4	Asymptotics of Hybrid Learning Schemes	312
7.8	Quiz	312
7.8.1	What is Wrong in Learning in Games?	312
7.8.2	Learning the Action Space	314
7.9	Chapter Review	314
8	Fully Distributed Learning for Global Optima	317
8.1	Introduction	317
8.2	Resource Selection Games	317
8.3	Frequency Selection Games	318
8.3.1	Convergence to One of the Global Optima	319
8.3.2	Symmetric Configuration and Evolutionarily Stable State	322
8.3.3	Accelerating the Convergence Time	323
8.3.4	Weighted Multiplicative imitative CODIPAS-RL	324
8.3.5	Three Players and Two Frequencies	329

8.3.5.1	Global Optima	329
8.3.5.2	Noisy Observation	329
8.3.6	Similar Learning Rate	331
8.3.7	Two Time-Scales	332
8.3.8	Three Players and Three Frequencies	332
8.3.9	Arbitrary Number of Users	332
8.3.9.1	Global Optimization	333
8.3.9.2	Equilibrium Analysis	334
8.3.9.3	Fairness	334
8.4	User-Centric Network Selection	335
8.4.1	Architecture for 4G User-Centric Paradigm	337
8.4.2	OPNET Simulation Setup	342
8.4.3	Result Analysis	344
8.5	Markov Chain Adjustment	345
8.5.1	Transitions of the Markov Chains	346
8.5.2	Selection of Efficient Outcomes	347
8.6	Pareto Optimal Solutions	348
8.6.1	Regular Perturbed Markov Process	350
8.6.2	Stochastic Potential	350
9	Learning in Risk-Sensitive Games	361
9.1	Introduction	361
9.1.1	Risk-Sensitivity	363
9.1.2	Risk-Sensitive Strategic Learning	365
9.1.3	Single State Risk-Sensitive Game	366
9.1.4	Risk-Sensitive Robust Games	366
9.1.5	Risk-Sensitive Criterion in Wireless Networks	367
9.2	Risk-Sensitive in Dynamic Environment	368
9.2.1	Description of the Risk-Sensitive Dynamic Environment	368
9.2.2	Description of the Risk-Sensitive Dynamic Game . . .	369
9.2.2.8	Two-by-Two Risk-Sensitive Games	373
9.2.2.9	Type I	375
9.2.2.10	Type II	375
9.3	Risk-sensitive CODIPAS	376
9.3.1	Learning the Risk-Sensitive Payoff	376
9.3.2	Risk-Sensitive CODIPAS Patterns	378
9.3.2.1	Bush-Mosteller based RS-CODIPAS	378
9.3.2.2	Boltzmann-Gibbs-Based RS-CODIPAS . . .	378
9.3.2.3	Imitative BG CODIPAS	379
9.3.2.4	Multiplicative Weighted Imitative CODIPAS	379
9.3.2.5	Weakened Fictitious Play-Based CODIAPS .	380
9.3.2.6	Risk-Sensitive Payoff Learning	380
9.3.3	Risk-Sensitive Pure Learning Schemes	381
9.3.4	Risk-sensitive Hybrid Learning Scheme	383

9.3.5	Convergence Results	384
9.3.5.2	Convergence to Equilibria	386
9.3.5.6	Convergence Time	387
9.3.5.8	Explicit Solutions	388
9.3.5.9	Composed Dynamics	390
9.3.5.11	Non-Convergence to Unstable Rest Points . .	391
9.3.5.13	Dulac Criterion for Convergence	391
9.4	Risk-Sensitivity in Networking and Communications	393
9.5	Risk-Sensitive Mean Field Learning	405
9.6	Extensions	409
9.6.1	Risk-Sensitive Correlated Equilibria	409
9.6.2	Other Risk-Sensitive Formulations	410
9.6.3	From Risk-Sensitive to Maximin Robust Games	410
9.6.4	Mean-Variance Approach	412
9.7	Chapter Review	413
9.7.1	Summary	413
9.7.2	Open Issues	415
A	Appendix	417
A.1	Basics of Dynamical Systems	417
A.2	Basics of Stochastic Approximations	423
A.3	Differential Inclusion	438
A.4	Markovian Noise	442
	Bibliography	443
	Index	459

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

List of Figures

1.1	Strategic Learning.	33
1.2	A generic combined learning scheme.	35
2.1	Convergence of best-reply.	68
2.2	Nonconvergence of best-reply	69
2.3	Nonconvergent aggregative game.	76
2.4	Design of step size.	79
2.5	Mann iteration. Design of step size.	80
2.6	Multiple access game between two mobiles.	99
2.7	Cognitive MAC Game.	101
2.8	Reduced Cognitive MAC Game.	101
2.9	A generic RL algorithm.	103
4.1	A generic CODIPAS-RL algorithm.	156
4.2	Mixed strategy of P1 under CODIPAS-RL BG.	175
4.3	Mixed strategy of P2 under CODIPAS-RL BG.	175
4.4	Probability to play the action 1 under Boltzmann-Gibbs CODIPAS-RL.	176
4.5	Average payoffs of action 1 under Boltzmann-Gibbs CODIPAS- RL.	176
4.6	Estimated payoffs for action 1 under Boltzmann-Gibbs CODIPAS-RL.	177
4.7	Probability to play the action 1 under Imitative CODIPAS- RL.	177
4.8	Estimated payoff for action 1 under Imitative CODIPAS-RL.	178
4.9	Two jammers and one regular node.	197
5.1	A delayed CODIPAS-RL algorithm.	211
5.2	Heterogeneous CODIPAS-RL: Convergence of the ODEs of strategies.	218
5.3	CODIPAS-RL: Convergence to global optimum equilibria.	219
5.4	CODIPAS-RL: Convergence of payoff estimations.	220
5.5	CODIPAS-RL under two-step delayed payoffs.	221
7.1	Large population of users.	287
7.2	Bad RSP: Zoom around the stationary point.	292
7.3	Mean field simulation of good RSP ternary plot and zoom.	293

7.4	Typical cognitive radio scenario under consideration.	297
7.5	A generic Brownian mobility.	298
7.6	Evolution of remaining energy.	299
8.1	Convergence to global optimum under imitation dynamics. .	319
8.2	Vector field of imitation dynamics.	320
8.3	Vector field of replicator dynamics.	320
8.4	Strategies.	325
8.5	Estimations and average payoffs.	326
8.6	Three users and two choices.	330
8.7	Three users and two actions.	353
8.8	Impact of the initial condition.	354
8.9	IMS based integration of operators with trusted third party	355
8.10	OPNET simulation scenario	355
8.11	The scenario	356
8.12	Evolution of randomized actions for underloaded configura- tion	356
8.13	Evolution of randomized actions for congested configuration.	357
8.14	Convergence to equilibrium.	357
8.15	Convergence to global optimum.	358
8.16	Evolution of randomized actions.	358
8.17	Evolution of estimated payoffs.	359
9.1	Global optima: $\mu_j < 0$	398
9.2	Convergence to global optima, $\mu_j < 0$	399
9.3	Convergence to global optimum $(1, 0, 0)$. $\mu_i > 0$	400
9.4	Two risk-averse users and one risk-seeking user. $\mu_1 < 0, \mu_2 <$ $0, \mu_3 > 0$	401
9.5	Imitative CODIPAS-RL. Impact initial condition. $\mu_i = -0.01$.	402
9.6	Imitative CODIPAS-RL, $\mu_j > 0$. Impact of initial condition.	403
9.7	Imitative CODIPAS-RL: 3D plot.	404

List of Tables

1.1	2×2 expected robust game.	13
1.2	Robust game with dominant strategy.	28
1.3	Anti-coordination robust game.	29
2.1	Comparison of analytical model estimates and auditory judgments (MOS).	86
2.2	Comparative properties of the different learning schemes . .	130
3.1	Information assumptions	138
3.2	Routing versus game theoretic parameters	145
4.1	Basic assumptions for CODIPAS.	157
4.2	CODIPAS: information and computation assumptions . . .	194
4.3	CODIPAS: learnable data.	195
5.1	Assumptions for games under channel uncertainty.	214
6.1	2×2 expected robust game.	248
8.1	Strategic form representation of 2 nodes and 2 technologies. .	318
8.2	Strategic form representation for 3 users - 2 frequencies . . .	329
8.3	Frequency selection game: 3 players, 3 frequencies	333
8.4	QoS parameters and ranges from the user payoff function . .	338
9.1	Asymptotic pseudo-trajectories of pure learning	382
9.2	Frequency selection games	393
9.3	Frequency selection games: random activity	395
9.4	Risk-sensitive frequency selection games	395
9.5	Frequency selection games: random activity	396
9.6	Summary	415

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Foreword

We live today in a truly interconnected world. Viewed as a network of decision making agents, decisions taken and information generated in one part, or one node, rapidly propagate to other nodes, and have impact on the well being (as captured by utilities) of agents at those other nodes. Hence, it is not only the information flow that connects different agents (or players, in the parlance of game theory), but also the cross-impact of individual actions. Individual players therefore know that their performance will be affected by decisions taken by at least a subset of other players, just as their decisions will affect others. To expect a collaborative effort toward picking the “best” decisions is generally unreasonable, and for various reasons, among which are nonalignment of individual objectives, limits on communication, incompatibility of beliefs, and lack of a mechanism to enforce a stable cooperative solution. Sometimes a player will not even know the objective or utility functions of other players, their motivations, and the possible cross-impacts of decisions.

How can one define an equilibrium solution concept that will accommodate different elements of such an uncertain decision making environment? How can such an equilibrium be reached when players operate under incomplete information? Can players learn through an iterative process and with strategic plays the equilibrium-relevant part of the game? Would such an iterative process converge, and to the desired equilibrium, when players learn at different rates, employ heterogeneous learning schemes, receive information at different rates, and adopt different attitudes toward risk (some being risk-neutral, other being risk-sensitive)?

The questions listed above all relate to issues that sit right in the heart of multi-agent networked systems research. And this comprehensive book meets the challenge of addressing them all, in the nine chapters to follow.

Professor Tamer Başar,
Urbana-Champaign,
Illinois, 11-11-11.

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface

Preface to the book Distributed Strategic Learning for Wireless Engineers

Much of Game Theory has developed within the community of Economists, starting from the book “Theory of Games and Economic behavior” by Morgenstern and Von Neumann (1944). To a lesser extent, it has had an impact on biology (with the development of evolutionary games) and on road traffic Engineering (triggered by the concept of Wardrop equilibrium introduced already in 1952 along with the Beckmann potential approach introduced in 1956). Since 1999 game theory has had a remarkable penetration into computer science with the formation of the community of Algorithmic game theory.

I am convinced that game theory will play a much more central role in many fields in the future including telecommunication network engineering. I use the term Network Engineering Games (NEGs) to call games that arise within the latter context. NEG is the young brother of the Algorithmic game theory. NEG is concerned with competition that arises at all levels of a network. This includes aspects related to information theory, to power control and energy management, to routing, to the transport and application layers of communication networks. It also includes competition arising in spread of information over a network as well as issues related to the economy of networks. Finally, it includes security issues, service denial attacks, spread of virus in computers and measures to fight it.

This book is the first to consider a systematic analysis of games arising in all network layers and is thus an important contribution to NEGs.

The word “game” may have connotations to “toys” or of “playing” (as opposed to decision making). But in fact it stands for decision making by several decision makers, each having her (or his) own individual objectives. Is game theory a relevant tool for research in communication networks? On 20/12/2011 I searched on Scholar Google the documents containing “wireless networks” together with “power control”. 20500 documents were found. Of these, 3380 appeared in 2011, and 1680 dated from 2000 or earlier. I then repeated the experience restricting further to documents containing “game theory”. 2600 documents were found. Of these, 20 dated from prior to 2001

and 580 dated from the single year 2011. The share of documents containing “game theory” thus increased from 1.2% to 17% within 10 years.

Is game theory relevant in wireless Engineering?

A user that changes some protocols in his cellular telephone may find out that a more aggressive behavior is quite beneficial and allows to obtain better performances. Yet if all the population tried to act selfishly and use more aggressive protocols, then everyone may loose in performance. But in practice we do not bother to change the protocols in our cellular phones. Making such changes would require access to the hardware, skills and training which is too much to invest. This may suggest that game theory should be used for other networking issues, perhaps in other scales (such as auctions over bandwidth, competition between service providers etc). So is there a need in NEG? Here are two different angles that one can use to look at this problem. First, we made here the assumption that decisions are taken concerning how to use equipment. But we can instead consider the decisions as being which equipment to buy. The user’s decisions concerning which protocol to use are taken when one purchases a telephone terminal. One prefers a telephone that is known to perform better. The game is then between equipment constructors. Secondly, not all decisions require special skills and knowhow. The service providers and/or the equipment constructors can often gain considerably by delegating to the users to take decisions. For example, when you wish to connect to the Internet from your laptop, you often go to a menu that provides you with a list of available connections along with some of their properties. The equipment provider has decided to leave us, the users, this choice. It also decides what information to let us have when we take the decision.

Leaving the decisions to the users is beneficial for service providers because of scalability issues: decentralizing a network may reduce signaling, computations and costs. When designing a network that relies on decisions taken by users, one needs to predict the users behavior. Learning is part of their behavior. Much of the theory of learning in games has been developed by biologists that used mathematical tools to model learning and adaptation within competing species. In NEG one need not restrict to describing existing learning approaches, one can propose and design learning procedures.

Why learn to play an equilibrium?

Sometimes it’s better not to learn. For example, assume that there are two players, one chooses x and the other chooses y , where both x and y lye in the half closed unit interval $[0, 1[$. Assume that both have the same utility

to maximize, which is given by $r(x, y) = xy$. It is easily seen that this game has an equilibrium which is unique: $(0, 0)$. This is the worst possible choice for both players. Any values of x and y that are strictly different from the equilibrium value give both a strictly better utility!

When a service provider delegates to users some decisions, it can control what parameter to let them control and what information to let them have so as to avoid such situations. Learning to play equilibrium may then be in the interest of the players, and exploring learning algorithms enrich the tools available in designing networks.

This book is unique among the books on learning in game theory in focusing on problems relevant to games in wireless engineering. It is a masterpiece bringing the state-of-the art foundations of learning in games to wireless.

Professor Eitan Altman
INRIA Sophia Antipolis
February 3rd, 2012

Strategic learning has made substantial progress since the early 1950s and has become a central element in economics, engineering, and computer science. One of the most significant accomplishments in strategic decision making during the last decades have been the development of game dynamics. Learning and dynamics are necessary when the problem to be solved is under uncertainty, time-variant and depends on the structure of the dynamic environment. This book develops distributed strategic learning schemes in games [15, 16, 17]. It offers several examples in networking, communications, economics and evolutionary biology in which learning and dynamics play an important role in understanding the behavior of the system.

As a first example, consider a spectrum access problem where the secondary users can sense a subset of channels. If there are unused channels by primary users at a given time slot, then the secondary users which sensed can access to the free channels. The problem is that even under slotted time and frames, several secondary users can simultaneously sense the same channels at the same time. We can explicitly describe this problem depending the channel conditions, the throughput, the set of primary users, the set of malicious users, the set of altruistic users (relays), the set of secondary users, their arrivals, departure rates, their past activities, but we are unable to explain how the secondary users do it if they sensed the same channel at the same time. Thus, it is useful to find a learning mechanism that allows an access allocation in the long-run.

As a second example, consider a routing packet over a wireless ad hoc network. The wireless path maximizing the quality of service with minimal end-to-end delay from a source to a destination changes continuously as the network traffic and the topology change. A learning-based routing protocol is therefore needed to estimate the network traffic and to predict the best stochastic path.

Already there are many successful applications of learning in networked games but also in many other domains: robotics, machine learning, bio-informatics, economics, finance, cloud computing, network security and reliability, social networks, etc. A great many textbooks have been written about learning in dynamic game theory. Most of them adopt either an economic perspective or a mathematical perspective. In the past several years, though, the application of game theory to problems in networking and communication systems has become more important. Specifically, game-theoretic models have been developed to better understand flow control, congestion control, power control, admission control, access control, network security, quality of service, quality of experience management and other issues in wireline and wireless systems. By modeling interdependent decision makers such as users, transmitters, radio devices, nodes, designer, operators, etc, game theory allows us to model scenarios in which there is no centralized entity with a full picture of the system conditions. It allows also teams, collaborations, and coalitional behaviors among the participants. The challenges in applying game theory to networking systems has attracted a lot of attention in the last decade. Most

of the game-theoretic models can abstract away important assumptions and mask critical unanswered questions. In absence of observation of the actions of the other participants and under unknown dynamic environment, the prediction of the outcome are less clear. It is our hope that this book will illuminate both the promise of learning in dynamic games as a tool for analyzing network evolution and the potential pitfalls and difficulties likely to be encountered when game theory is applied by practicing engineers, undergraduate, graduate students, and researchers. We have not attempted to cover either learning in games or its applications to networking and communications. We have severely restricted our exposition to those topics that we feel are necessary to give the reader a grounding in the fundamentals of learning in games under uncertainty or robust games and their applications to networking and communications.

As most of wireless networks are dynamic and evolve in time, we are seeing a tendency toward decentralized networks, in which each node may play multiple roles at different times without relying on an access point or a base station (small base station, femto-cell BS or macro-cell BS) to make decisions such as in what frequency band to operate, how much power to use during transmission frame, when to transmit, when to go in sleep mode, when to upgrade, etc. Examples include cognitive radio networks, opportunistic mobile ad hoc networks, and sensor networks that are autonomous and self-organizing and support multihop communications. These characteristics lead to the need for distributed decision making that potentially takes into account network conditions as well as channel conditions. In such distributed systems, an individual terminal may not have access to control information regarding other terminal's actions and network congestion may occur. We address the following questions:

- Question One: How much information is enough for effective distributed decision making?
- Question Two: Is having more information always useful in terms of system performance (value/price of information)?
- Question Three: What are the individual learning performance bounds under outdated and imperfect measurement?
- Question Four: What are the possible dynamics and outcomes if the players adopt different learning patterns?
- Question Five: If convergence occurs, what is the convergence time of heterogeneous learning (at least two of the players use different learning patterns)?
- Question Six: What are the issues (solution concepts, non-convergence, convergence rate, convergence time etc) of hybrid learning (at least one player changes its learning pattern during the interaction)?

- Question Seven: How to develop very fast and efficient learning schemes in scenarios where some players have more information than the others?
- Question Eight: What is the impact of risk-sensitivity in strategic learning systems?
- Question Nine: How do we construct learning schemes in a dynamic environment in which one of the players does not observe a numerical value of its own-payoffs but only a signal of it?
- Question Ten: How to learn “unstable” equilibria and global optima in a fully distributed manner?

These questions are discussed through this book. There is an explicit description of how players attempt to learn over time about the game and about the behavior of others (e.g. through reinforcement, adaptation, imitation, belief updating, estimations or combination of these etc.). The focus is both on finite and infinite systems, where the interplay among the individual adjustments undertaken by the different players generate different learning dynamics, heterogeneous learning, risk-sensitive learning, and hybrid dynamics.

How to use this book?

This Guide is designed to assist instructors in helping students grasp the main ideas and concepts of *distributed strategic learning*. It can serve as the text for learning algorithm courses with a variety of different goals, and for courses that are organized in a variety of different manners. The Instructor's note and supporting materials is developed for use in a course using *distributed strategic learning* with the following goals for students:

Students will be better able to think about iterative process for engineering problems;

Students will be better able to make use of their algorithmic, graphing, and computational skills in real wireless networks based on data;

Students will be better able to independently read, study and understand the topics that are new to the students such as solution concepts in robust games;

Students will be better able to explain and describe the learning outcomes and notions orally and to discuss both qualitative and quantitative topics with others;

We would like to make the following remarks. The investigations of various solutions are almost independent of each other. For example, you may study the strategy dynamics by reading [Chapter 2](#) and payoff dynamics by reading [Chapter 3](#). If you are interested only in the risk-sensitive learning, you should read [Chapter 8](#). Similar possibilities exist for the random updates, heterogeneous learning, and hybrid learning (see the Table of Contents).

If you plan an introductory course on robust game theory, then you may use [Chapter 1](#) for introducing robust games in strategic-form. Remark. [Chapters 2 - 8](#) may be used for a one-semester course on distributed strategic learning.

Each chapter contains some exercises. The reader is advised to solve at least those exercises that are used in the text to complete the proofs of various results.

This book can be used for a one semester course by sampling from the chapters and possibly by discussing extra research papers; in that case, I hope that the references at the end of the book are useful. I welcome your feedback via email to [tembineh\(at\)gmail.com](mailto:tembineh(at)gmail.com). I very much enjoyed writing this course, I hope you will enjoy reading it.

Notation and Terminology

The book is comprised of nine chapters and one [appendix](#). Each chapter is divided into sections, and sections occasionally into subsections. [Section 2.3](#), for example, refers to the third section of [Chapter 2](#), while [Subsection 2.3.1](#) is the first section of [Subsection 2.3](#).

Items like theorems, propositions, lemmas, etc., are identified within each chapter according to the standard numbering; Equation (7.1) would be the first equation of [Chapter 7](#).

Organization of the book

The manuscript comprises nine chapters.

- [Chapter one](#) introduces basic strategic decision-making and robust games. State-dependent games with different level of information are formulated and the associated solution concepts are discussed. Later, distributed strategic learning approaches in different layers of the open systems inter-connection model (OSI) including physical layer (PHY), medium access control (MAC) layer, network layer, transport layer, and application layer are presented.
- In [Chapter two](#), we overview classical distributed learning schemes. We start with partially distributed strategy-learning algorithms and their possible implementation in wireless networks. Generically, partially distributed learning schemes, sometimes called semi-distributed schemes, assume that all players knew their own-payoff functions and, observe others' actions in previous stages. This is clearly not the case in many networking and communication problems of interest. Under this strong assumption, several game-theoretic formulations are possible for uncertain situations. Then, the question of how to learn the system characteristics in presence of incomplete information and imperfect measurements is addressed. Both convergence and nonconvergence results are provided. In the other chapters of this book, we develop strategic learning framework by assuming that each player is able to learn progressively its own-action space, knows his or her current action, and observes a numerical (possibly noisy) value of her (delayed) payoff (the mathematical structure of the payoff functions are unknown as well as the actions of the other players). This class of learning procedures is called fully distributed strategy-learning algorithm or model-free strategy-learning and is presented in [section 2.3](#).
- [Chapter 3](#) focuses on *payoff learning and dynamics*. The goal of Payoff Learning is to learn the payoff functions, the expected payoffs and the risk-sensitive payoffs. In many cases, the exact payoff functions may not be known by the players. The players try to learn the unknown data through the long-run interactions. This chapter complements the [Chapter two](#).
- [Chapter 4](#) studies *combined fully distributed payoff and strategy learning* (CODIPAS). The core chapter examines how can evolutionary game theory be used as a framework to analyze multi-player reinforcement learning algorithms in an heterogeneous setting. In addition, equilibrium seek-

ing algorithms, learning in multi-armed bandit problems and algorithms for solving variational inequality are presented. CODIPAS combines both strategy-learning and payoff-learning.

- **Chapter 5** examines *combined learning under delayed and unknown pay-offs*. Based on outdated and noisy measurements, combining learning schemes that incorporates the delays, as well as schemes that avoid the delays, are investigated. Relevant applications to wireless networks are presented.
- **Chapter 6** analyzes *combined learning in constrained-like games*. The core of the chapter comprises two parts. The first part introduces constrained games and the associated solution concepts. Then, we address the challenging question of how such a game can be played? How player can choose their actions in constrained games? The second part of the chapter focuses on satisfactory solutions. Instead of robust optimization framework, we propose a robust satisfaction theory which is relevant quality-of-experience (QoE, application layer) and quality-of-service (QoS, network layer) problems. The feasibility conditions as well as satisfactory solutions are investigated. The last part of the chapter is concerned about *random matrix games* (RMGs) with variance criterion.
- **Chapter 7** extends the heterogeneous learning to *hybrid learning*. Uncertainty, random updates and switching between different learning procedures are presented.
- **Chapter 8** develops *learning schemes for global optima*. The chapter provides specific class of games in which global optimum can be found in a fully distributed manner. Selection of larger sets, Pareto optimal solutions, are discussed. A detailed MATLAB code associated to the example of resource selection games is provided.
- **Chapter 9** presents *risk-sensitivity aspects in learning*. The classical game-theoretic approach to modeling multi-player interaction assumes that players in a game want to maximize their expected payoff. But in many settings, players instead often want to maximize some more complicated function of their payoff. The expected payoff framework for games is obviously very general, but it does exclude the possibility that players in the game have preferences that depend on the entire distribution of payoff, and not just on its expectation. For example, if a player is sensitive to risk, her objective might be to balance the variance to be closer to the expectation. Indeed, this is the recommendation of modern portfolio theory, and a version of the mean-variance objective is widely used by investors in financial markets as well as in network economics. The chapter also addresses the generalization of familiar notions of Nash and correlated equilibria to settings where players are sensitive to the risk. We especially examine the impact of risk-sensitivity in the outcome.

- Background materials on dynamical systems and stochastic approximations are provided in appendices.

Acknowledgments

I would like to thank everyone who made this book possible. I owe a special debt of gratitude to those colleagues who gave up their time to referee the Chapters. I would like to thank Professors Eitan Altman, Anatoli Iouditski, and Sylvain Sorin who initiated my interest in learning under uncertainty. The development of this book has spanned many years. The material as well as its presentation has benefited greatly from the inputs of many bright undergraduate and graduate students who worked on this topic. I would like to thank my colleagues from Ecole Polytechnique for their comments. It is a pleasure to thank my collaborators and coauthors of articles and papers on which part of the chapters of this book is based. They have played an important role in shaping my thinking for so many years. Their direct and indirect contributions to this work are significant. They are, of course, not responsible for the way I have assembled material, especially the parts I have added to and subtracted from our joint works to try to make the manuscript more coherent.

Special thank to Professor Tamer Başar who kindly accepted my invitation to write a foreword to the book, to Professor Eitan Altman who kindly accepted to write a preface. My thanks go to Professors Vivek Borkar, Mrouane Debbah, Samson Lasaulce, David Leslie, Galina Schwartz, Mihaela van der Schaar, Thanos Vasilakos, and Peyton H. Young for fruitful interactions or collaborations. I am grateful to the anonymous reviewers for assistance in proofreading the manuscript.

I thank seminar participants at the University of California at Los Angeles (UCLA), Ecole Polytechnique, University of California at Berkeley, University of Avignon, the National Institute for Research in Computer Science and Control (INRIA), University of Illinois at Urbana Champaign (UIUC), McGill University, Ecole Polytechnique Fdrale de Lausanne (EPFL), Ecole Suprieure d'Electricit (Supelec), University of California at Santa Cruz (UCSC), etc.

Artwork

The scientific graphs in the book are generated using the MATLAB software by Mathworks Inc. and the mean field package for simulation of large population games. The two figures in the cover of the book are examples of cycling learning processes.

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The Author Bio

Hamidou Tembine has two master degrees, one in applied mathematics and one in pure mathematics from respectively Ecole Polytechnique and University Joseph Fourier, France. He received a PhD degree from Avignon University, France. He is an assistant professor at Ecole Supérieure d'Electricité, Supelec, France. His current research interests include evolutionary games, mean field stochastic games and applications. He was the recipient of many student travel grant awards, and best paper awards (ACM Valuetools 2007, IFIP Networking 2008, IEEE/ACM WiOpt 2009, IEEE Infocom Workshop 2011).

Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contributors

Below is the list of contributors of the preface of the book and the foreword.

Eitan Altman, Research Director, Institut de Recherche en Informatique et en Automatique, INRIA Sophia Antipolis, France

Tamer Başar, Professor, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Illinois, US.

Symbol Description

$\mathbb{R}^k$	k -dimensional Euclidean space, $k \geq 2$.	$\mathbb{1}_{\{\cdot\}}$	Indicator function.
$\mathcal{N}$	Set of players (finite or infinite).	l^2	Space of sequences $\{\lambda_t\}_{t \geq 0}$ such that $\sum_t \lambda_t ^2 < +\infty$.
$\mathcal{B}(t)$	Random set of active players at time t .	l^1	Space of sequences $\{\lambda_t\}_{t \geq 0}$ such that $\sum_t \lambda_t < +\infty$.
$\mathcal{A}_j$	Set of actions of player j .	$\lambda_{j,t}$	Learning rate of player j at t .
$s_j \in \mathcal{A}_j$	An element of $\mathcal{A}_j$.	$\mathbf{e}_{s_j} \in \mathcal{X}_j$	Unit vector with 1 at the position of s_j , and zero otherwise.
$\Delta(\mathcal{A}_j)$	Set of probability distributions over $\mathcal{A}_j$.	$\ \cdot\ _2$	$\ x\ _2 = (\sum_k x_k ^2)^{\frac{1}{2}}$.
$\mathcal{X}_j$	Mixed actions $\Delta(\mathcal{A}_j)$.	$\langle \cdot, \cdot \rangle$	Inner product.
$a_{j,t}$	Action of the player j at time t . Element of $\mathcal{A}_j$.	$\mathcal{W}$	State space, environment state.
$\mathbf{x}_{j,t}$	Randomized action of the player j at t . Element of $\mathcal{X}_j$.	$\mathbf{w} \in \mathcal{W}$	A scalar, a vector or a matrix (finite dimension).
$r_{j,t}$	Perceived payoff by player j at t .	$2^{\mathcal{D}}$	The set of the all the subsets of $\mathcal{D}$.
$\hat{\mathbf{r}}_{j,t}$	Estimated payoff vector of player j at t . Element of $\mathbb{R}^{ \mathcal{A}_j }$.	$C^0(A, B)$	Space of continuous functions from A to B .
$\tilde{\beta}_{j,\epsilon}(\hat{r}_{j,t})$	Boltzmann-Gibbs strategy of player j . Element of $\mathcal{X}_j$.	$\mathbb{N}$	Set of natural numbers (non-negative integers).
$\tilde{\sigma}_{j,\epsilon}(\hat{r}_{j,t})$	Imitative Boltzmann-Gibbs strategy of player j . Element of $\mathcal{X}_j$.	$\mathbb{Z}$	Set of integers.
		M_t	Martingale.

1

Introduction to Learning in Games

One of the objectives in distributed interacting multi-player systems is to enable a collection of selfish players to achieve a desirable objective. There are two overriding challenges to achieving this objective. The first one is related to the complexity of finding the optimal solution. A centralized algorithm may be prohibitively complex when there are a large number of interacting players. This motivates the use of adaptive methods that enable players to self-organize into suitable, if not optimal, alternative solutions. The second challenge is limited information. Players may have limited knowledge about the status of other players, except perhaps for a small subset of neighboring players. The limitations in terms of information induce robust stochastic optimization, bounded rationality and inconsistent beliefs. As a consequence, there are many simple games in which the beliefs may converge but not the strategies. The outcome is sensitive to

- how much signalling is available to the players,
- the nature of the learning scheme,
- the way the information is exploited.

Common Knowledge

An event is part of the common knowledge [13, 14] if all the players know it; and all the players know that all other players know it; and all other players know that all other players know that all other players know it, and so on ad infinitum. This is much more than simply saying that something is known by all.

Rationality

If every player always maximizes his payoff, thus being able to perfectly calculate the possible probabilistic result of every action, we say that the player is *rational*.

Rationality under uncertainty

We use the notation $\mathbf{x}_1 \succeq \mathbf{x}_2$ to say that $\mathbf{x}_1$ is preferred to $\mathbf{x}_2$. A single decision maker (player) is called rational under uncertainty if its preferences satisfy the following conditions:

- (*Completeness*) Either $\mathbf{x}_1 \succeq \mathbf{x}_2$ or $\mathbf{x}_2 \succeq \mathbf{x}_1$ for any pair $(\mathbf{x}_1, \mathbf{x}_2)$.
- (*Transitivity*) If $\mathbf{x}_1 \succeq \mathbf{x}_2$ and $\mathbf{x}_2 \succeq \mathbf{x}_3$, then $\mathbf{x}_1 \succeq \mathbf{x}_3$.
- (*Monotonicity*) If $\mathbf{x}_1 \succeq \mathbf{x}_2$ and $\alpha_1 > \alpha_2, (\alpha_1, \alpha_2) \in [0, 1]^2$, then $\alpha_1 \mathbf{x}_1 + (1 - \alpha_1) \mathbf{x}_2 \succeq \alpha_2 \mathbf{x}_1 + (1 - \alpha_2) \mathbf{x}_2$.
- (*Continuity*) If $\mathbf{x}_1 \succeq \mathbf{x}_2$ and $\mathbf{x}_2 \succeq \mathbf{x}_3$, then there exists a probability α such that $\mathbf{x}_2 = \alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_3$.
- (*Independence*) If $\mathbf{x}_1 \succeq \mathbf{x}_2$, then $\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_3 \succeq \alpha \mathbf{x}_2 + (1 - \alpha) \mathbf{x}_3$ for any $\alpha \in [0, 1]$ and $\mathbf{x}_3$.

As above, the completeness condition ensures that all distributions can be compared with each other, and the transitivity condition implies that the distributions can be listed in order of preference. The monotonicity and continuity conditions assert that a lottery gets better smoothly as the probability of a preferred outcome increases. The independence condition implies that preferences only depend on the differences between lotteries; components that are the same can be ignored.

We will focus on payoffs representing the above properties. Expected payoffs and other type of payoffs such as risk-sensitive payoffs will be considered.

In the context of strategic decision-making, a strong property such as *Common Knowledge of Rationality* has been widely used.

Common Knowledge of Rationality (CKR)

CKR is informally defined as follows:

- The players are rational.
- The players all know that the other players are rational.
- The players all know that the other players know that they are rational.
- ... (in principle) ad infinitum.

However, such maximization is often quite difficult, and even if they wanted to, most players would be unable to carry them out in practice due to computational limitations, communication complexity, and anticipation errors [14].

Why should we be interested in distributed strategic learning?

Rationality does not imply that the outcome of a game must be an equilibrium, and neither does the common knowledge that players are rational, as equilibrium requires all players to coordinate on the same equilibrium. However, game theory results show that the outcome after multiple stages of moves is often much closer to equilibrium predictions than initial moves, which supports the idea that equilibrium arises as a result of players learning from experiences. The learning approach in games formalizes this idea, and examines how, which and what kind of equilibrium might arise as a consequence of a long-run non-equilibrium process of learning, adaptation, and/or imitation. Learning approaches are not necessarily procedures that are “*trying to reach Nash equilibrium*” but ask also the following questions:

- How can players optimize their own payoff while simultaneously learning about the strategies of other players?
- How can players optimize their own unknown payoff functions while simultaneously learning about their payoff functions and the behavior of the other players?

If players’ strategies are completely observed at the end of each time slot, and players interact in the long-term, fairly simple rules perform well in terms of the players worst-case payoffs, and also guarantee that steady state of the process must be closer to an equilibrium point (saddle point, Cournot/Nash equilibrium, Hannan set, correlated equilibrium, etc.) than at the starting point. Now, if players cannot observe the strategies chosen by the other players, then steady states of learning process are not necessarily Nash equilibria because players can maintain incorrect beliefs about off-path play. Beliefs can also be incorrect due to cognitive limitations, errors, and anticipation mistakes.

The classical learning procedures in games (see [59, 60, 194, 193] and the references therein) assume that:

- C1 The number of players is small and fixed,
- C2 The rules of the game are known by every player, At least the distribution of possible payoffs, is known by all the players, and the action spaces are fully known.
- C3 Every player is rational.

Even under these “strong” classical assumptions (C1-C3) and hyper-rationality of players, learning in games is difficult because it is interactive:

Each player’s learning process correlates and affects what has to be learned by every other player over time.

Note that the correlations of the learning processes in games make a clear difference with what is usually done in classical single player decision-making.

In many practical scenarios, the above assumptions on information about the other players may not hold or may need a central authority or a coordinator to feedback some signals/messages. We will discuss how these assumptions can be relaxed. We specially examine the case where players can adjust their behavior only in response to their own realized payoffs, they have no knowledge of the overall structure of the game, the number of players that interact can be random, they cannot observe the actions or payoffs of the other players, they occasionally mistakes on their choices and readapt themselves based on past experiences, regret, reinforcement, and imitation.

In many situations in engineering, economics, and finance, one would like to have a learning procedure that does not require any information about the other players actions or payoffs and less memory (small number of parameters in terms of past own-actions and past perceived own-payoffs) as possible. Such a rule is said to be *uncoupled* or *fully distributed*. However, for a large class of games, no such general algorithm causes the players' period-by-period behavior to converge to Nash equilibrium. Hence, there is no guarantee for uncoupled dynamics that behaviors come close to Nash equilibrium most of the time.

A common correlation scheme in a one-shot game is the following: each player chooses his/her action according to his/her observation of the value of the same public signal. A strategy assigns an action to every possible observation a player can make. If no player would want to deviate from the recommended strategy (assuming the others don't deviate), the distribution is called a *correlated* Cournot/Nash equilibrium. A set is said to be *convex* if it is empty or it contains all the convex combinations of its elements. Convex sets and functions are very useful in robust optimization models, and have a structure that is convenient for learning algorithms. The set of correlated equilibria is convex and includes the Nash equilibria but is frequently much larger.

The works in [58, 70, 46] show that *regret minimizing procedures* can cause the empirical frequency distribution of play to converge to the set of correlated Nash equilibria. But there is no arbitrator, and no random device in these models. So,

What is the origin of this correlation?

It is clear that the correlation comes from the interactions via the perceived payoff which is incorporated in the regret term. Thus, the perceived own-payoffs over several stages have played the role of private signals of the process.

Under these correlations during the interactions, it is important to know *when (rational or not) players can learn to play equilibrium starting from out of an equilibrium point.*

To answer to this question, we will distinguish two main classes of learning algorithms.

A learning algorithm is *fully distributed or uncoupled* if it does not need explicit information about the other players.

Each player constructs his/her strategies and updates them by using the knowledge of his set of own-actions and measurement of own-payoffs.

A learning scheme is *partially distributed* if each player implements its updating rules after explicitly receiving an explicit signal (e.g., action profile) or an information exchange about the other players.

This last class of learning procedures is more sophisticated than the class of fully distributed learning algorithms. These algorithms are still decentralized but they use explicit information from the others. In some scenarios, it may be possible to deduce useful information in terms of payoffs and/or strategies by combining the observed signal and the measured payoff if external knowledge is available to the player. These situations are referred as implicit information. Each of these learning classes is important in games and will be discussed in detail in this book.

1.1 Basic Elements of Games

Our approach in this book is based on some elements of dynamic robust game theory. It will be useful to present at the outset some of the basic notions of robust game theory and some general results on the existence and characterization of robust equilibria and maximin points. This section is dedicated to readers who are not familiar with strategic form games.

1.1.1 Basic Components of One-Shot Game

The four basic components of a one-shot game are

- The set of players, $\mathcal{N} = \{1, \dots, n\}$, where n is the total number of players,

- The action spaces $\mathcal{A}_1, \dots, \mathcal{A}_n$. These sets can be discrete or continuous, finite or infinite and assumed to be non-empty. Most of the time we will work with compact¹ action spaces (in Euclidean space).
- The preference structure of the players. In this book we focus on the case where the preference relation can be represented by a certain function: utility, payoff, cost, loss, reward, benefit, etc. The functions $\tilde{R}_1, \dots, \tilde{R}_n$ are real-valued functions defined over the product space $\prod_{j=1}^n \mathcal{A}_j$; the outcome.
- The information structure and order of play of the game.

If the game is played once, the game is called a one-shot game, and its *strategic form* (also called normal-form) is represented by a collection²:

$$\tilde{\mathcal{G}} = \left(\mathcal{N}, \{\mathcal{A}_j\}_{j \in \mathcal{N}}, \{\tilde{R}_j\}_{j \in \mathcal{N}} \right).$$

A generic element of $\mathcal{A}_j$ is denoted by s_j . If player $j \in \mathcal{N}$ chooses an action s_j in $\mathcal{A}_j$ according to a probability distribution $\mathbf{x}_j = (x_j(s_j))_{s_j \in \mathcal{A}_j}$ over $\mathcal{A}_j$, the choice of $\mathbf{x}_j$ is called a randomized action or mixed strategy of the one-shot game. We denote by $\mathcal{X}_j = \Delta(\mathcal{A}_j)$ the set of probability distributions over $\mathcal{A}_j$. We will work with the canonical probability space. When the vector $\mathbf{x}_j \in \mathcal{X}_j$ is on a vertex of the simplex $\mathcal{X}_j$, the mixed strategy boils down to a pure strategy i.e., the deterministic choice of an action. We use the canonical inclusion that maps an element of $\mathcal{A}_j$ to the simplex by considering the mapping $s_j \mapsto e_{s_j}$ where e_{s_j} is the unit vector with same size as the cardinality of $\mathcal{A}_j$, with 1 at the position of s_j and 0 otherwise.

⚠ Note that the notation e_{s_j} is to represent the vector $e_{j,s_j} \in \mathbb{R}^{|\mathcal{A}_j|}$. We do not use a specific notation for scalars, vectors, matrices or matrix of vectors because the action can be of any of these types or even a generic object.

The mixed extension payoff is $R_j : \prod_j \mathcal{X}_j \rightarrow \mathbb{R}$ defined by

$$R_j(\mathbf{x}_1, \dots, \mathbf{x}_n) := \mathbb{E}_{\mathbf{x}_1, \dots, \mathbf{x}_n} \tilde{R}_j \quad (1.1)$$

$$= \sum_{\mathbf{a} \in \prod_j \mathcal{A}_j} \tilde{R}_j(\mathbf{a}) \left(\prod_{j' \in \mathcal{N}} x_{j'}(a_{j'}) \right), \quad (1.2)$$

where $\mathbb{E}$ denotes the mathematical expectation operator. This is well defined by continuity and compactness assumption.

¹A topological space $\mathcal{X}$ is called compact if every collection of open subsets of $\mathcal{X}$ that covers $\mathcal{X}$ contains a finite subcollection covering $\mathcal{X}$. It is well-known that $\mathcal{X} \subseteq \mathbb{R}^k$ is compact if and only if it is bounded and closed.

²Note that one-shot games in extensive form can be transformed in normal form but one may lose some information if the information sets are not added.

What is the difference between action and pure strategy?

For a static game in normal form as above, there is no real distinction between pure strategies and actions. However, the distinction will become important when we consider state-dependent one-shot games, robust games with partial information and dynamic games. This distinction will be clarified in [Chapters 2](#) and [4](#). The next example explains the difference between actions and pure strategies in the context of access control protocols. The key idea is that

Information is crucial in the process of decision making.

Example 1.1.2. Consider a finite number of users over a common shared medium [98, 112, 165]. Each user has its own queue with arrival process described over a probability space. Since the medium is common, there is interference if several users transmit simultaneously. This means the probability of success of each user depends on the other users; thus there is an interaction between them.

An action in this game is a transmit power to be chosen among a list. There is also a “null” action that corresponds to the case where the user does not transmit, i.e., power zero. However, a user will not transmit if there is no packet to transmit. This implies that the knowledge of the size of own-queue or the knowledge of the state of the queue (empty or not) is necessary in order to construct a realistic model. Suppose that each user is able to observe the size of its queue or backoff. Then, a pure strategy is a mapping from the queue sizes to a set of transmit powers, which is completely different from just selecting a power from the list independently of everything else. Thus,

- An action is an element of the list of transmit powers.
- A pure strategy is a mapping from the set of queue sizes to the list of transmit powers.

Here an output of a pure strategy (function of queues) corresponds to a transmit power.

In the case of nonrandom payoffs, we will denote the instantaneous payoff function by $r_j(\mathbf{x})$ where $\mathbf{x} = (\mathbf{x}_j)_{j \in \mathcal{N}} \in \prod_{j \in \mathcal{N}} \mathcal{X}_j$.

Next, we define a Cournot/Nash equilibrium.

Definition 1.1.2.1 (Cournot/Nash equilibrium). A strategy profile $\mathbf{x}^*$ is a Cournot/Nash equilibrium of the one-shot game

$$\mathcal{G} = (\mathcal{N}, \{\mathcal{X}_j\}_{j \in \mathcal{N}}, \{r_j\}_{j \in \mathcal{N}}),$$

if it satisfies, for all $j \in \mathcal{N}$ and for all $\mathbf{x}_j \in \mathcal{X}_j$,

$$r_j(\mathbf{x}_j^*, \mathbf{x}_{-j}^*) \geq r_j(\mathbf{x}_j, \mathbf{x}_{-j}^*). \quad (1.3)$$

where $\mathbf{x}_{-j}^* = (\mathbf{x}_k^*)_{k \neq j}$.

We will often wish to focus on the value of $\mathbf{x}$ at which the maximum is achieved rather than the maximum value of the function itself, so we introduce a symbol $\arg \max$.

About the notation $\arg \max$.

Suppose $\mathbf{x}$ is an arbitrary member of the set $\mathcal{X}$. Let $r(\mathbf{x})$ be some function that is defined for all $\mathbf{x} \in \mathcal{X}$. The symbol $\arg \max$ is defined by the following equivalence:

$$\mathbf{x}^* \in \arg \max_{\mathbf{x} \in \mathcal{X}} r(\mathbf{x}) \iff r(\mathbf{x}^*) = \max_{\mathbf{x} \in \mathcal{X}} r(\mathbf{x}).$$

This suppose that the function has a maximum (the sup is attained).

⚠ Note that $\arg \max$ returns a set. For this reason we do not write $\mathbf{x}^* = \arg \max r(\mathbf{x})$ because a function may take its maximum value for more than one element in the set $\mathcal{X}$. Because the symbol $\arg \max$ may return a set of values rather than a unique value, it is called a correspondence rather than a function. Even in the case of unique maximizer, one has to write $\{\mathbf{x}^*\} = \arg \max r(\mathbf{x})$.

Example 1.1.3. Consider the function defined on $x \in [-1, 1]$, by $r(x) = x^2$. This function achieves its maximum at $x^* \in \{-1, +1\}$. So,

$$\arg \max_{x \in [-1, +1]} x^2 = \{-1, 1\}.$$

The following subproblem is a part of Definition 1.1.2.1.

For each $j \in \mathcal{N}$, and $\mathbf{x}_{-j}$, one would like to solve the optimization problem $OP_j(\mathbf{x}_{-j})$:

$$OP_j(\mathbf{x}_{-j}) : \sup_{\mathbf{x}_j \in \mathcal{X}_j} r_j(\mathbf{x}_j, \mathbf{x}_{-j}) \quad (1.4)$$

An interesting question that we will investigate is the following: Is it possible to solve the optimization problem $OP_j(\mathbf{x}_{-j})$ without knowledge of the mathematical structure function r_j ?

The solutions to the problem $OP_j(\mathbf{x}_{-j})$ are sometimes referred as **best-responses** of player j to the strategies $\mathbf{x}_{-j}$ of the others. A best response $BR_j(\cdot)$ is a correspondence (multi-valued mapping) from $\prod_{j' \in \mathcal{N}, j' \neq j} \mathcal{X}_{j'} \rightarrow 2^{\mathcal{X}_j}$, defined by the set

$$BR_j(\mathbf{x}_{-j}) = \arg \max_{\mathbf{x}_j} r_j(\cdot, \mathbf{x}_{-j}).$$

In the next section, we address the question of non-emptiness of this set for a given strategy profile of the other players.

Exercise 1.1. Show that the Definition (1.1.2.1) is equivalent to the following characterization: For any player j , any action $s_j \in \mathcal{A}_j$,

$$\mathbf{x}_j^*(s_j) > 0 \implies r_j(\mathbf{e}_{s_j}, \mathbf{x}_{-j}^*) = \max_{s'_j \in \mathcal{A}_j} r_j(\mathbf{e}_{s'_j}, \mathbf{x}_{-j}^*) = \max_{\mathbf{x}'_j \in \mathcal{X}_j} r_j(\mathbf{x}'_j, \mathbf{x}_{-j}^*) \quad (1.5)$$

which says exactly that

$$\forall j \in \mathcal{N}, \text{ support}(\mathbf{x}_j^*) := \{s_j \mid x_j^*(s_j) > 0\} \subseteq \arg \max_{\mathcal{X}_j} r_j(\cdot, \mathbf{x}_{-j}^*). \quad (1.6)$$

Hint: To prove the last equality in (1.5), use the concavity of the mixed extension with the respect to x_j , continuity and compactness [37].

These relations have been established in Nash (1951, [119]) in his paper entitled Non-Cooperative Games. Note that this property is related to the Wardrop first principle (Wardrop 1952, [187], Beckmann et al. 1956, [21]). Typically, if $\mathcal{P} = \{1, 2, \dots, K\}$, $K \geq 1$ denotes a set of subpopulations and $\mathcal{A}^p$ a finite set of actions in subpopulation, the population profile $\mathbf{z} = (\mathbf{z}^p)_{p \in \mathcal{P}}$, $\mathbf{z}^p \in \Delta(\mathcal{A}^p)$, is an equilibrium state if for any subpopulation p , and any action $a^p \in \mathcal{A}^p$ such that

$$z^p(a^p) > 0 \implies r_{a^p}^p(\mathbf{z}) = \max_{b^p \in \mathcal{A}^p} r_{b^p}^p(\mathbf{z}) = \max_{\mathbf{y}^p \in \Delta(\mathcal{A}^p)} \int_{b^p} \mathbf{y}^p(db^p) r_{b^p}^p(\mathbf{z}) \quad (1.7)$$

where $r_{b^p}^p(z)$ is the payoff of a generic player in subpopulation p with the action b^p . As we can see this definition of *equilibrium state* in the context of multiple subpopulations is more general than the definition in (1.5). Note that, in the multiple subpopulation game, for a fixed action a^p , the corresponding payoff $r_{a^p}^p(\cdot)$ may depend on $\mathbf{z}^{-p} = (\mathbf{z}^{p'})_{p' \in \mathcal{P}, p' \neq p}$ but also on z^p . Also, considering macro-players, one gets the first definition (1.1.2.1). However, the function $r_{a^p}^p(\mathbf{z})$ do not need to be concave/convex in z^p .

1.1.4 State-Dependent One-Shot Game

If there are independent random variables that determine the one-shot game, we add a state space $\mathcal{W}$ and the payoff function will be defined on product space: $\mathcal{W} \times \prod_{j \in \mathcal{N}} \mathcal{X}_j$. A strategy in this game depends on the information of the realized state.

A pure strategy of player j in the one-shot game is a mapping from the set of information available to player j to its action space.

1.1.4.1 Perfectly-Known State One-Shot Games

Given a state $\mathbf{w} \in \mathcal{W}$ that is known by all the players, we define a normal-form game

$$\tilde{\mathcal{G}}^1(\mathbf{w}) = (\{\mathbf{w}\}, \mathcal{N}, (\mathcal{A}_j^{\mathbf{w}}, \tilde{R}_j(\mathbf{w}, \mathbf{a}))_{j \in \mathcal{N}}),$$

where $\mathcal{A}_j^{\mathbf{w}}$ is the set of mapping from $\{\mathbf{w}\}$ to the set $\mathcal{A}_j$ and the payoff takes an element of $\prod_{j \in \mathcal{N}} \mathcal{A}_j^{\mathbf{w}}$ to a real value which we simply denote by $\tilde{R}_j(\mathbf{w}, \mathbf{a})$. This is a parameterized game with a given parameter $\mathbf{w}$ that is common knowledge and which influences both strategies and payoffs.

1.1.4.2 One-Shot Games with Partially-Known State

Given a state $\mathbf{w} \in \mathcal{W}$, each player knows only some components of the state. For example, player j observes/measures w_j but does not know the current value of w_{-j} . We assume that the distribution μ_{-j} over state w_{-j} is known by player j . This defines a one-shot game with incomplete information. Each player j is able to compute the payoff function given by

$$\tilde{R}_j^2(w_j, \mu_{-j}; a) = \mathbb{E}_{w_{-j} \sim \mu_{-j}} \left[\tilde{R}_j(w, a) \mid w_j, \mu_{-j} \right].$$

Let $\Delta(W_{-j})$ be the set of probability measures over W_{-j} , equipped with the canonical sigma-algebra. Then, $\mu_{-j} \in \Delta(W_{-j})$. Note that $\tilde{R}_j^2$ is defined over $W_j \times \left(\prod_{j \in \mathcal{N}} \mathcal{A}_j \right) \times \Delta(W_{-j})$, where W_j denotes the j -th component state space observed by player j . A pure strategy of player j is this game with partial state observation is a mapping from the given own-state w_j , and the distribution μ_{-j} to $\mathcal{A}_j$. We denote the game by

$$G^2 = (\mathcal{N}, \mathcal{A}, (w_j, \mu_{-j}), (\tilde{R}_j^2)_{j \in \mathcal{N}}).$$

A pure equilibrium for the game G^2 can be seen as a Bayesian (Nash) equilibrium, and it is characterized by

$$\begin{aligned} \sigma_j : W_j \times \Delta(W_{-j}) &\longrightarrow \mathcal{A}_j, \\ \sigma_j(w_j, \mu_{-j}) &\in \arg \max_{\tilde{a}_j \in \mathcal{A}_j} \tilde{R}_j^2(w_j, \mu_{-j}; \tilde{a}_j, \sigma_{-j}(w_{-j})), \end{aligned}$$

$\forall w_j$ and $\forall j \in \mathcal{N}$.

Note that, a pure strategy for player j in this one-shot game is a mapping from the set $\{(w_j, \mu_{-j})\}$ to the set $\mathcal{A}_j$

1.1.4.3 State Component is Unknown

The players do not know any component of the current state. The distributions over the full states are known in addition to the mathematical structure of the payoffs. Now each player j is able to compute the following payoff function

$$\tilde{R}_j^3(\mu; a) = \mathbb{E}_{w \sim \mu} \tilde{R}_j(w, a),$$

where $\mu \in \Delta(\mathcal{W})$. This leads to an expected robust game that we denote by

$$G^3 = (\mathcal{N}, \mathcal{A}, \mathcal{W}, \mu, (\tilde{R}_j^3)_{j \in \mathcal{N}}).$$

A pure strategy for player j in the game G^3 corresponds to a function of μ that gives an element of $\mathcal{A}_j$.

1.1.4.4 Only the State Space Is Known

Now, we assume that the distribution over the state is also unknown but the state space $\mathcal{W}$ is known. Then, each player can adopt different behaviors depending on how he/she sees the state space. The well-known approaches in that case are the maximin robust and the maximax approaches (pessimistic or optimistic approaches) and their variants. The payoff of player j in the maximin robust game is given by

$$\tilde{R}_j^4(a) = \inf_{w \in \mathcal{W}} \tilde{R}_j(w, a).$$

We denote the game by

$$G^4 = (\mathcal{N}, (\mathcal{A}_j)_{j \in \mathcal{N}}, \mathcal{W}, (\tilde{R}_j^4)_{j \in \mathcal{N}}).$$

The equilibria of G^4 are called *maximin robust equilibria*.

Definition 1.1.4.5. *Worse-case Cournot/Nash equilibrium if (i) there exists $w^*(x)$ such that*

$$r_j(w^*(x), x) = \inf_{\mathbf{w}} r_j(w, x),$$

and (ii):

$$r_j(w^*(x), x) = \max_{x'_j} r_j(w^*(x'_j, x_{-j}), x'_j, x_{-j})$$

⚠ The choice of the set $\mathcal{W}$ plays an important role in the Worse-case Cournot/Nash equilibrium. For example, if $\tilde{R}_j(H, A) = \log \det \left(I + \frac{1}{N_0} \sum_{j' \in \mathcal{N}} H_{j'}^* P_{j'} H_{j'} \right)$, and $H_j \in \mathbb{C}^{n \times n}$, then

$$\inf_H \tilde{R}_j(H, A) = 0.$$

Therefore everything is optimal, which does not reflect the solutions to the mutual information optimization.

Similarly, the payoff of player j in the maximax robust game is given by

$$\bar{R}_j^5(a) = \sup_{w \in \mathcal{W}} \tilde{R}_j(w, a).$$

1.1.5 Perfectly Known State Dynamic Game

At each time-step t , a state is randomly picked according to some distribution $\nu(\cdot)$ defined in a probability space. The selected state $\mathbf{w}_t$ is perfectly and publicly known. Then, the perfectly-known state simultaneous-act game $\tilde{\mathcal{G}}(\mathbf{w}_t)$ is played. Note that, in this context, the actions at time t are dependent on the state $\mathbf{w}_t$.

1.1.6 Unknown State Dynamic Games

We consider dynamic games where the players have an opportunity to act several times. Additionally, the dynamic game can be under uncertainty; that is, the game can have a state which changes randomly from time slot to another.

In network scenarios, the state will be typically a *backoff state*, a *battery state*, a *vector of channel states*, the *availability of the resources* etc.

The games under state uncertainty are sometimes called *robust games*, that is in the class of games with incomplete information and imperfect payoffs. If $R_j(\mathbf{w}, \mathbf{x})$ denotes the non-random payoff functions of player j , one way to analyze the equilibrium issue is to consider the expected game defined by

$$(\mathcal{N}, (\mathcal{X}_j)_{j \in \mathcal{N}}, \mathbb{E}_{\mathbf{w}} R_j(\mathbf{w}, \cdot)).$$

We denote the set of states by $\mathcal{W}$. The set $\mathcal{W}$ is equipped with a probability measure ν .

Problem formulation: For each $j \in \mathcal{N}$, and $\mathbf{x}_{-j}$, one would like to solve the problem $ROP_j(\mathbf{x}_{-j})$:

$$ROP_j(\mathbf{x}_{-j}) : \sup_{\mathbf{x}_j \in \mathcal{X}_j} \mathbb{E}_{\mathbf{w}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}) \quad (1.8)$$

The key question that we will address in this manuscript is: how to solve the robust optimization problem $ROP_j(\mathbf{x}_{-j})$ without knowledge of the function $R_j(\cdot)$?

The Nash equilibrium of the expected robust game is therefore defined as follows:

Definition 1.1.6.1 (Equilibrium of the expected robust game). *A strategy profile $(\mathbf{x}_j)_{j \in \mathcal{N}} \in \prod_{j=1}^n \mathcal{X}_j$ is a (mixed) state-independent equilibrium for the expected robust game if and only if $\forall j \in \mathcal{N}, \forall \mathbf{y}_j \in \mathcal{X}_j$,*

$$\int_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{y}_j, \mathbf{x}_{-j}) \nu(d\mathbf{w}) \leq \int_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}) \nu(d\mathbf{w}). \quad (1.9)$$

which is exactly the definition 1.1.2.1 with the payoff function $r_j : \prod_{j \in \mathcal{N}} \mathcal{X}_j \rightarrow \mathbb{R}$, defined by

$$r_j(\mathbf{x}_j, \mathbf{x}_{-j}) = \int_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}) \nu(d\mathbf{w}) = \mathbb{E}_{\mathbf{w}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}).$$

Note that the strategy profile is state-free, i.e., independent of the environment state $\mathbf{w}$.

Two-by-two expected robust games

Consider two players. The action spaces of each player are independent on the state. Player I has two actions $\{Up, Down\}$ in any state, and Player II's action set is $\{Left, Right\}$ in any state. We transform the robust game into expected game by taking the expectation of the payoff. Table 6.1 represents the payoff entries. where

Decisions of player I \ Decisions of player II	<i>Left</i>	<i>Right</i>
<i>Up</i>	$\bar{a}_1, \bar{a}_2$	b_1, b_2
<i>Down</i>	$\bar{c}_1, \bar{c}_2$	d_1, d_2

TABLE 1.1

2×2 expected robust game.

- $\bar{a}_1 = \mathbb{E}_{\mathbf{w}} \tilde{R}_1(\mathbf{w}, Up, Left)$, $\bar{a}_2 = \mathbb{E}_{\mathbf{w}} \tilde{R}_2(\mathbf{w}, Up, Left)$
- $\bar{b}_1 = \mathbb{E}_{\mathbf{w}} \tilde{R}_1(\mathbf{w}, Up, Right)$, $\bar{b}_2 = \mathbb{E}_{\mathbf{w}} \tilde{R}_2(\mathbf{w}, Up, Right)$
- $\bar{c}_1 = \mathbb{E}_{\mathbf{w}} \tilde{R}_1(\mathbf{w}, Down, Left)$, $\bar{c}_2 = \mathbb{E}_{\mathbf{w}} \tilde{R}_2(\mathbf{w}, Down, Left)$
- $\bar{d}_1 = \mathbb{E}_{\mathbf{w}} \tilde{R}_1(\mathbf{w}, Down, Right)$, $\bar{d}_2 = \mathbb{E}_{\mathbf{w}} \tilde{R}_2(\mathbf{w}, Down, Right)$.

It is easy to show that the optimal completely mixed state-independent strategy (if it exists) is given by

$$x^* = \frac{\bar{d}_1 - \bar{b}_1}{\bar{d}_1 - \bar{b}_1 + \bar{a}_1 - \bar{c}_1},$$

$$y^* = \frac{\bar{d}_2 - \bar{b}_2}{\bar{d}_2 - \bar{b}_2 + \bar{a}_2 - \bar{c}_2}.$$

Convexity and Quasi-Convexity

We briefly introduce the notion of convexity. Let $\mathcal{E}$ be a linear space. A subset $\mathcal{C}$ of $\mathcal{E}$ is *convex* if for all

$$\alpha c_1 + (1 - \alpha)c_2 \in \mathcal{C}, \quad \forall (c_1, c_2) \in \mathcal{C}, \alpha \in [0, 1].$$

The following properties hold:

- The closure and the interior of a convex set are convex.
- The intersection of two convex sets is a convex set.

A function $f : \mathcal{C} \rightarrow \mathbb{R}$ is convex if

$$f(\alpha c_1 + (1 - \alpha)c_2) \leq \alpha f(c_1) + (1 - \alpha)f(c_2), \quad \forall (c_1, c_2) \in \mathcal{C}^2, \alpha \in [0, 1]$$

Note that the convexity of the domain $\mathcal{C}$ is a prerequisite for saying that a function is convex.

A function $f : \mathcal{C} \rightarrow \mathbb{R}$ where the set $\mathcal{C}$ is a convex set is called *concave* if the function $(-f)$ is convex.

A function $f : \mathcal{C} \rightarrow \mathbb{R}$ where the set $\mathcal{C}$ is a convex set is called *strictly convex* if

$$\begin{aligned} f(\alpha c_1 + (1 - \alpha)c_2) &< \alpha f(c_1) + (1 - \alpha)f(c_2), \\ \forall (c_1, c_2) \in \mathcal{C}^2, c_1 &\neq c_2, \alpha \in (0, 1) = [0, 1] \setminus \{0, 1\}. \end{aligned}$$

If $f : \mathcal{C} \rightarrow \mathbb{R}$ is a function and η is a scalar, the set

$$\{c \in \mathcal{C} \mid f(c) \leq \eta\}$$

is called level set of f . One nice property is that if f is a convex function, then all its level sets are convex.

⚠ Note, however, that convexity of the level sets does not imply convexity of the function. A counterexample is obtained with the function $f : c \mapsto \sqrt{|c|}$ is not convex. The level sets are the empty set or the set $\{c \mid c^2 \leq \eta^4\}$.

What is a topological space?

A topological space is a set $\mathcal{E}$ together with $\mathcal{T}$ (a collection of subsets of $\mathcal{E}$) satisfying the following axioms:

- The empty set $\emptyset$ and the whole space $\mathcal{E}$ are in $\mathcal{T}$.
- $\mathcal{T}$ is closed under arbitrary union.
- $\mathcal{T}$ is closed under finite intersection.

The collection $\mathcal{T}$ is called a topology on $\mathcal{E}$. The sets in $\mathcal{T}$ are called the open sets, and their complements in $\mathcal{E}$ are called closed sets. A subset of $\mathcal{E}$ may be neither closed nor open, either closed or open, or both.

⚠ A topological set is not like a DOOR!

Compactness

A topological space $\mathcal{E}$ is compact if, whenever a collection of open sets covers the space, some subcollection consisting only of finitely many open sets

also covers the space. For Euclidean space, a subset is compact if it is closed and bounded subset.

A function between topological spaces is called *continuous* if the inverse image of every open set is open. A *Hausdorff space* or separated space is a topological space in which distinct points can have disjoint neighborhoods.

Semicontinuity

A real-valued function f is lower semicontinuous if $\{e \in \mathcal{E} \mid f(e) > \alpha\}$ is an open set for every $\alpha \in \mathbb{R}$. For real-valued functions, this is equivalent to say that a mapping $f : A \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous at a_0 if $\liminf_{a \rightarrow a_0} f(a) \geq f(a_0)$. f is lower-semicontinuous if it is lower semicontinuous at every point of its domain.

Upper semicontinuity is defined in a similar way by considering $\limsup$.

An important property that we will use in robust games for the emptiness of the best-response set is the following:

If $\mathcal{C}$ is a compact space and $f : \mathcal{C} \rightarrow \mathbb{R}$ is upper semicontinuous, then f has a maximum on $\mathcal{C}$.

A function $f : \mathcal{C} \rightarrow \mathbb{R}$ is quasiconvex if

$$f(\alpha c_1 + (1 - \alpha)c_2) \leq \max(f(c_1), f(c_2)), \quad \forall (c_1, c_2) \in \mathcal{C}^2, \alpha \in [0, 1]$$

⚠ The floor function $x \mapsto \lfloor x \rfloor$ is a quasiconvex function that is neither convex nor continuous.

A quasiconcave function is a function whose negative is quasiconvex. The probability density function of the normal distribution is quasiconcave but not concave.

Next, we present some well-known results in fixed point theory.

Fixed-point theorem

Theorem 1.1.6.2 (Kakutani, 1941). *Let $\mathcal{D} \subseteq \mathbb{R}^d$, $d \geq 1$ be a nonempty, compact and convex set. Let the correspondence $\phi : \mathcal{D} \rightarrow 2^{\mathcal{D}}$ be upper hemicontinuous (its graph is closed) with nonempty convex values. Then ϕ has a fixed point, i.e.,*

$$\exists x^* \in \mathcal{D}, x^* \in \phi(x^*) \in 2^{\mathcal{D}}.$$

Theorem 1.1.6.2 provides sufficiency conditions for existence of equilibria. It is a generalization of Brouwer's fixed-point theorem, Schauder's fixed point theorem, etc.

Theorem 1.1.6.3 (Existence of equilibrium, Kakutani-Glicksberg-Debreu-Fan). *Suppose that*

- *for each $j \in \mathcal{N}$, $\mathcal{X}_j$ is a nonempty, convex, and compact subset of a Hausdorff topological vector space,*
- *for each j , r_j is upper semicontinuous on $\prod_{j \in \mathcal{N}} \mathcal{X}_j$, for each j and*
- *each $x'_j \in \mathcal{X}_j$ the functions $x_{-j} \mapsto r_j(x'_j, x_{-j})$ is lower semicontinuous and,*
- *for each j and $x_{-j} \in \prod_{i \neq j} \mathcal{X}_i$, $x'_j \in \mathcal{X}_j \mapsto r_j(x'_j, x_{-j})$ is quasi-concave.*

Then, the one-shot game

$$(\mathcal{N}, (\mathcal{X}_j, r_j(\cdot))_{j \in \mathcal{N}}),$$

has at least one (Cournot/Nash) equilibrium.

⚠ The above result provides a sufficiency condition for existence of an equilibrium. It does not address the question of how an equilibrium can be reached or can be played, which will be formulated in terms of partial information and bounded reasoning. Remember that you don't have to know you are playing a game to be in a game!

Theorem 1.1.6.4. *The following results hold.*

- *If the action spaces are nonempty and finite and $\mathcal{W}$ is compact and nonempty, then the expected robust game has at least one equilibrium in mixed (state-independent) strategies.*
- *If the action spaces and $\mathcal{W}$ are compact and non-empty, and the functions $\bar{R}_j(\mathbf{w}, a_j, \mathbf{a}_{-j})$ are continuous then the expected robust game has at least one equilibrium in mixed strategies (measures over a compact set).*
- *If the action spaces are convex, compact and non-empty, $\mathcal{W}$ compact and non-empty and $\bar{R}_j(\mathbf{w}, a_j, \mathbf{a}_{-j})$ continuous in $\mathbf{a}$, quasi-concave in a_j , then the existence of pure equilibria holds in the expected robust game.*

Proof. • To prove the first statement, we show that for every player j , the payoff function

$$\mathbf{x} \mapsto r_j(\mathbf{x}) = \mathbb{E}_{\mathbf{w}}[R_j(\mathbf{w}, \mathbf{x})] = \int_{\mathbf{w}} R_j(\mathbf{w}, \mathbf{x}) \nu(d\mathbf{w})$$

is continuous. Moreover, $\mathbf{x}_j \mapsto r_j(\mathbf{x})$ is quasi-concave for any vector

$\mathbf{x}_{-j}$. The strategy spaces $\mathcal{X}_j$ are convex, compact, and non-empty. We can therefore use one of the fixed point theorems (Brouwer, Schauder, Kakutani, etc.) which guarantee the existence of least one equilibrium in mixed strategies. To use the Brouwer theorem, we may need to construct Brown-von Neumann excess payoff ratio or Nash map.

- The proofs of the second statement and the last statement follow the same line using the Ky Fan fixed point theorem (which is a generalization of the Kakutani fixed point theorem).
- The proof of the last statement follows from Kakutani-Glicksberg-Debreu-Fan's theorem applied to the payoff function $\mathbb{E}\tilde{R}_j(\mathbf{w}, a_j, \mathbf{a}_{-j})$.

□

As we can observe, fixed-point theorems play important roles in establishing existence of equilibria. But there are many classes of games (namely, weakly acyclic game, potential games, supermodular games, robust potential games, population games with monotone payoffs, etc, see [99] and the references therein) where the existence of Nash equilibrium can be established without recourse to fixed-point theorems.

Another way to analyze the equilibrium of robust games is to consider the worst-state case game or Maximin robust game defined by

$$\left(\mathcal{N}, (\mathcal{X}_j)_{j \in \mathcal{N}}, \inf_{\mathbf{w}} R_j(\mathbf{w}, \cdot) \right).$$

The Nash equilibrium of the maximin robust game is therefore defined as follows:

Definition 1.1.6.5 (Maximin Robust Equilibrium). *A strategy profile $(\mathbf{x}_j)_{j \in \mathcal{N}} \in \prod_{j \in \mathcal{N}} \mathcal{X}_j$ is a (mixed) state-independent maximin robust equilibrium if and only if*

$$\forall j \in \mathcal{N}, \quad \inf_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{y}_j, \mathbf{x}_{-j}) \leq \inf_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}), \quad \forall \mathbf{y}_j \in \mathcal{X}_j. \quad (1.10)$$

which is exactly the Definition 1.1.2.1 with the payoff function $r_j : \prod_{j \in \mathcal{N}} \mathcal{X}_j \rightarrow \mathbb{R}$, defined by

$$r_j(\mathbf{x}_j, \mathbf{x}_{-j}) = \inf_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \mathbf{x}_j, \mathbf{x}_{-j}).$$

Again, the strategy profile is distribution-free. The players do not need to know the state $\mathbf{w}$ since they focus on the worst-state case. Note that Theorem 1.1.6.4 extends to this maximin robust game under similar sufficient conditions.

Other Notions of Robust Equilibria

We describe the robust game $\mathcal{G}$ the collection $\mathcal{G} = (\mathcal{G}(\mathbf{w}))_{\mathbf{w} \in \mathcal{W}}$ where $\mathcal{G}(\mathbf{w})$ is a normal form game. All the games " $\mathcal{G}(\mathbf{w})$ " have the same set of action spaces (independent of the state " $\mathbf{w}$ "). Denote by $E(\mathcal{G}(\mathbf{w}))$ the set of Nash equilibria of the normal-form game $\mathcal{G}(\mathbf{w})$ for a fixed $\mathbf{w}$.

Definition 1.1.6.6. An action profile a^* is state-robust equilibrium of $\mathcal{G}$ if

$$\forall \mathbf{w} \in \mathcal{W}, a^* \in E(\mathcal{G}(\mathbf{w})).$$

Denote by $E(\mathcal{G})$ the set of equilibria of $\mathcal{G}$. The interpretation of this state-robust equilibrium is the following: a state-robust equilibrium is independent of $\mathbf{w}$ and it is an equilibrium for "all" the games $\mathcal{G}(\mathbf{w})$, $\mathbf{w}$ describing $\mathcal{W}$. This is a distribution-free equilibrium. This notion of equilibrium may be too demanding since one has an existence problem even for simple cases. Below, we provide examples of state-robust equilibrium as well as examples where there is no such equilibrium.

An alternative to this definition of quantifying the set of states for which a given profile a is equilibrium. It consists of finding the probability for a given action profile $\mathbf{a} \in \prod_{j \in \mathcal{N}} \mathcal{A}_j$ to be an equilibrium of $\mathcal{G}$.

Similarly, the analogue of the uniqueness problem becomes: What is the probability that $(a, a') \in E(\mathcal{G})^2$, $a \neq a'$?

This last problem consists to quantify the measure of the set: $\mathbb{P}(\{\mathbf{w} \mid (a, a') \in E(\mathcal{G}(\mathbf{w}))^2\})$ for a given pair $(a, a'), a \neq a'$. Here, we have a random variable over $\mathcal{W}$, associated with the collection $(\mathcal{W}, \mathcal{P}(\mathcal{W}), \mathbb{P})$ where $\mathcal{P}(\mathcal{W})$ is a σ -algebra.

We say that the probability of having more than one pure equilibrium in $\mathcal{G}$ is zero if $\forall a \neq a'$,

$$\mathbb{P}(\{\mathbf{w} \mid (a, a') \in E(\mathcal{G}(\mathbf{w}))^2\}) = 0.$$

Example 1.1.6.7 (Many equilibria versus domination). Consider the following two-player, two-state, two-action per state game with symmetric payoff.

$P1 \backslash P2$	s'_1 of $P2$	s'_2 of $P2$	$P1 \backslash P2$	s'_1 of $P2$	s'_2 of $P2$
s_1 of $P1$	1	1	s_1 of $P1$	1	1/4
s_2 of $P1$	1	1	s_2 of $P1$	1/4	0
state " $w = 0$ "			state " $w = 1$ "		

There are two states represented by $\{0, 1\}$ indexed by w . The set of actions of player 1 $\mathcal{A}_1(w = 0) = \{s_1, s_2\} = \mathcal{A}_1(w = 1)$, and the actions space of player 2 is $\mathcal{A}_2(w = 1) = \{s'_1, s'_2\} = \mathcal{A}_2(w = 0)$. Player 1 chooses a row and player 2 chooses a column. Lets examine the equilibria in each state. In state $w = 0$, any action is an equilibrium, i.e, the set of pure equilibria is $\{1, 2\}^2$ and the set of mixed equilibria is

$$\{((x, 1 - x), (y, 1 - y)) \mid x \in [0, 1], y \in [0, 1]\}.$$

In state $w = 1$, the first action is strictly dominant in the sense that the payoff of the action s_1 is strictly greater than the payoff of the action s_2 , independently of the choice of the other player. Thus, the equilibrium is $(1, 1)$. We deduce that the action profile which consists of playing the first action for both players leads to a state-robust equilibrium.

Example 1.1.6.8 (No pure distribution-independent equilibrium). Consider the following state-dependent one-shot game. There are two states. In state 0, the first action dominates the second one, and in state 1 the second action is dominant. This means that (s_1, s_1) is a pure equilibrium of the matrix game $\mathcal{G}(0)$, and (s_2, s_2) is an equilibrium of the game $\mathcal{G}(1)$. We verify that there is no pure distribution-independent equilibrium in $\mathcal{G}$.

$P1 \backslash P2$	s'_1 of $P2$	s'_2 of $P2$	$P1 \backslash P2$	s'_1 of $P2$	s'_2 of $P2$
s_1 of $P1$	1	1/4	s_1 of $P1$	0	1/4
s_2 of $P1$	1/4	0	s_2 of $P1$	1/4	1
state " $w = 0$ "			state " $w = 1$ "		

Example 1.1.6.9 (Three states). We consider a three-player game. There are three states. In each state, each player has two actions. In the table below, Player 1 chooses a row, player 2 chooses a column and player 3 chooses one of the two blocks.

<table><tr><td>1</td><td>1</td></tr><tr><td>1</td><td>1</td></tr></table>	1	1	1	1	<table><tr><td>1</td><td>1/4</td></tr><tr><td>1/4</td><td>0</td></tr></table>	1	1/4	1/4	0	<table><tr><td>0</td><td>1/4</td></tr><tr><td>1/4</td><td>1</td></tr></table>	0	1/4	1/4	1
1	1													
1	1													
1	1/4													
1/4	0													
0	1/4													
1/4	1													
$w = 0$	$w = 1$	$w = 2$												
<i>continuum of MEs</i>	$1 - E$	$1 - E$												

Assume that the states are drawn according the probability distribution $\tilde{w} \mapsto (\mu_1, \mu_2, 1 - \mu_1 - \mu_2)$. We compute $\mathbb{P}(\{(1, 1) \in E(\mathcal{G}(\mathbf{w}))\})$ in order to quantify the probability for $(1, 1)$ to be an equilibrium. It is easy to see that this probability is at least $\mu_1 + \mu_2$, which can be arbitrary close to 1. Furthermore, there is no pure (distribution-independent) equilibrium in $\mathcal{G}$ but each of the games $\mathcal{G}(\mathbf{w})$ has at least one pure equilibrium.

This leads naturally to the question of existence in distribution-dependent pure equilibrium.

Example 1.1.6.10. To illustrate an example of distribution-dependent equilibrium, we consider the above state-dependent scenario where the states are drawn according to a distribution $(\mu, 1 - \mu)$ for some $1 > \mu > 0$. In addition to the state-dependent game, we represent also the expected robust game:

<table><tr><td>1</td><td>1/4</td></tr><tr><td>1/4</td><td>0</td></tr></table>	1	1/4	1/4	0	<table><tr><td>0</td><td>1/4</td></tr><tr><td>1/4</td><td>1</td></tr></table>	0	1/4	1/4	1	<table><tr><td>μ</td><td>1/4</td></tr><tr><td>1/4</td><td>$1 - \mu$</td></tr></table>	μ	1/4	1/4	$1 - \mu$
1	1/4													
1/4	0													
0	1/4													
1/4	1													
μ	1/4													
1/4	$1 - \mu$													
$w = 0$	$w = 1$	<i>expected</i>												

We can take a strategy in the form:

$$x_1^*(\mu) = \begin{cases} 1 & \text{if } \mu > \mu^* \\ 0 & \text{otherwise} \end{cases}$$

independent of w but depends on the law μ .

Note that $\mathcal{G}$ has a distribution-dependent pure equilibrium.

Example 1.1.6.11 (Continuum of mixed equilibria in each game). Let $\Delta(\{1, 2, 3, 4\}) = \{x = (x_1, x_2, x_3, x_4) \in \mathbb{R}_+^4 \mid x_1 + x_2 + x_3 + x_4 = 1\}$ be the three-dimensional simplex of $\mathbb{R}^4$ and consider the two family of games

0	0	-1	-1
0	0	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1

-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	0	0
-1	-1	0	0

Each game has a continuum of mixed equilibria: $\Delta(\{1, 2\})^2$ for the first game and $\Delta(\{3, 4\})^2$ for the second game. It is not difficult to see that the robust game $\mathcal{G}$ has no pure equilibrium in distribution-independent strategies. This means that the existence of pure equilibria in each state-dependent game does not necessarily implies the existence of pure equilibria in the expected robust game.

1.1.7 State-Dependent Equilibrium

Let $\tilde{S}_j$ be the set of maps from the state space $\mathcal{W}$ to the action set $\mathcal{A}_j$ of player j . A state-dependent pure strategy of player j is an element of $\tilde{S}_j$. We define a state-dependent Cournot/Nash pure equilibrium as follows.

Definition 1.1.7.1 (Pure state-dependent equilibrium). A strategy profile σ^* is a pure state-dependent equilibrium of the robust game

$$\mathcal{G} = \left(\mathcal{W}, \mathcal{N}, \{\mathcal{A}_j\}_{j \in \mathcal{N}}, \{\tilde{R}_j\}_{j \in \mathcal{N}} \right),$$

if it satisfies, for all $j \in \mathcal{N}$ and for all $\sigma_j \in \tilde{S}_j$ such that $\sigma_j(\mathbf{w}) \in \mathcal{A}_j$,

$$\int_{\mathbf{w} \in \mathcal{W}} \tilde{R}_j(\mathbf{w}, \tilde{\sigma}_j^*(\mathbf{w}), \tilde{\sigma}_{-j}^*(\mathbf{w})) \nu(d\mathbf{w}) \geq \int_{\mathbf{w} \in \mathcal{W}} \tilde{R}_j(\mathbf{w}, \tilde{\sigma}_j(\mathbf{w}), \tilde{\sigma}_{-j}^*(\mathbf{w})) \nu(d\mathbf{w}) \quad (1.11)$$

where $\tilde{\sigma}_{-j}^* = (\tilde{\sigma}_k^*)_{k \neq j} \in \prod_{j' \neq j} \tilde{S}_{j'}$.

Let S_j be the set of maps from the state space $\mathcal{W}$ to the randomized action set $\mathcal{X}_j$.

We define a mixed state-dependent Cournot/Nash equilibrium as follows.

Definition 1.1.7.2 (Mixed state-dependent equilibrium). A strategy profile σ^* is a state-dependent equilibrium of the robust game

$$\mathcal{G} = \left(\mathcal{W}, \mathcal{N}, \{\mathcal{X}_j\}_{j \in \mathcal{N}}, \{\tilde{R}_j\}_{j \in \mathcal{N}} \right),$$

if it satisfies, for all $j \in \mathcal{N}$ and for all $\sigma_j \in S_j$ such that $\sigma_j(\mathbf{w}) \in \mathcal{X}_j$,

$$\int_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \sigma_j^*(\mathbf{w}), \sigma_{-j}^*(\mathbf{w})) \nu(d\mathbf{w}) \geq \int_{\mathbf{w} \in \mathcal{W}} R_j(\mathbf{w}, \sigma_j(\mathbf{w}), \sigma_{-j}^*(\mathbf{w})) \nu(d\mathbf{w}) \quad (1.12)$$

where $\sigma_{-j}^* = (\sigma_k^*)_{k \neq j} \in \prod_{j' \neq j} S_{j'}$.

Note that we cannot talk about uniqueness in robust games without specification. The reason is that the uniqueness is in the probabilistic sense, but it is also sensitive to the class of strategies that is under consideration. For example, a game that has a unique state-independent strategy equilibrium, it does not necessarily have a unique *state-dependent strategy* equilibrium. To see this statement, consider a modified strategy

$$\sigma_j(\mathbf{w}) = \begin{cases} \sigma_j^* & \text{if } \mathbf{w} \in \tilde{\mathcal{W}} \\ \sigma_j \in S_j & \text{arbitrary otherwise} \end{cases}$$

where $\mathcal{W} \setminus \tilde{\mathcal{W}}$ is a ν -null set.

1.1.8 Random Matrix Games

A random matrix game is a particular case of robust game where the interaction payoffs are given by matrices with random entries. Below we describe the case of two players. Player 1 has $l_1 \geq 2$ choices and Player 2 has $l_2 \geq 2$ choices. The rows are labeled with the pure strategies for Player 1, the columns with the pure strategies of Player 2. The payoff matrix of player 1 is a random matrix $M_1 \in \mathbb{R}^{l_1 \times l_2}$ and the payoff of player 2 is a random matrix $M_2 \in \mathbb{R}^{l_1 \times l_2}$. The matrix entries $M_{1,ij}, M_{2,ij}$ are the pairs of payoffs for the pure strategy pair (s_i, s'_j) . Note that the labels of the actions or pure strategies (rows or columns) do not matter. We write these matrices as follows

$$M_1 = \bar{M}_1 + (M_1 - \bar{M}_1) = \bar{M}_1 + N_1 \quad (1.13)$$

$$M_2 = \bar{M}_2 + (M_2 - \bar{M}_2) = \bar{M}_2 + N_2 \quad (1.14)$$

where $\bar{M}_j = \mathbb{E}M_j$ is the expected matrix for $j \in \{1, 2\}$ and $N_j = M_j - \bar{M}_j$, is a zero-mean random variable (a noise). More details on random matrix games (RMGs) can be found in [Chapter 6](#).

1.1.9 Dynamic Robust Game

We describe a generic model of robust difference games, i.e., games under uncertainty and in discrete time. The dynamic game is described by

$$\Gamma = (\mathcal{W}, \mathcal{N}, \mathcal{T}, (\mathcal{A}_j, \Theta_{j,t}, h_{j,t}, N_{j,t}, \Sigma_{j,t})_{j \in \mathcal{N}, t \in \mathcal{T}}, (F_j)_{j \in \mathcal{N}})$$

where

- $\mathcal{T}$ is the time space. This set is assumed to be nonempty. For difference game, the set $\mathcal{T} = \{1, 2, 3, 4, \dots\}$ represents the game stages.
- A finite or infinite set $\mathcal{W}$, w_t is the state of the environment at stage t . A state may represent a joint channel state, collection of individual locations, collection of queue sizes, weather conditions or activity/inactivity of the players, etc.

- A set of players $\mathcal{N} = \{1, 2, \dots, n\}$. There is at least one potential player in the game, $1 \leq n < +\infty$. Sometimes “Nature” can be considered as a player.
- $\mathcal{A}_j$ is an action set of player $j \in \mathcal{N}$ at time $t \in \mathcal{T}$. Its elements are allowed action $a_{j,t}$ of player j at stage t .
- A set $\Theta_{j,t}$ is the observation set of the player j at stage t , $\theta_{j,t} \in \Theta_{j,t}$.
- The function $h_{j,t} : \mathcal{W} \times \prod_{j' \in \mathcal{N}} \mathcal{A}_{j'} \longrightarrow \Theta_{j,t}$ defined by

$$\theta_{j,t} = h_{j,t}(w_t, a_t),$$

is the observation of player j concerning the current state of the game.

- A set $\eta_{j,t}$ defined as a subset of

$$\left(\prod_{1 \leq t' \leq t-1} \prod_{j \in \mathcal{N}} \theta_{j,t'} \times \mathcal{A}_{j,t'} \right) \times \prod_{j \in \mathcal{N}} \theta_{j,t},$$

which determines the information set of player j at stage t . The specification of $\eta_{j,t}$ for the stage of the game characterizes the information structure of player k and the set $\{\eta_{j,t}\}_{t \in \mathcal{T}, j \in \mathcal{N}}$ is the information structure of the game.

- $N_{j,t}$ is the information space at stage t induced by $\eta_{j,t}$. This set contains the action observation structure of the player j .
- A class $\Sigma_{j,t}$ of mappings $\sigma_{j,t} : N_{j,t} \longrightarrow \mathcal{A}_j$, are the admissible decisions at stage t . The collection $\sigma_j = (\sigma_{j,t})_{t \in \mathcal{T}}$ is the strategy of player j and Σ_j is the strategy set of player j .
- A function

$$F_j : \prod_{t \in \mathcal{T}} \left(\mathcal{W} \times \prod_{j' \in \mathcal{N}} \mathcal{A}_{j'} \right) \longrightarrow \mathbb{R},$$

called payoff function of player j in the dynamic game. An intermediary payoff is not specified. In the additive payoff function the long-term payoff is the cumulative stage payoffs. An example of nonadditive payoff F_j is the norm-sup of all the stage payoffs obtained along the trajectory.

1.2 Robust Games in Networks

Emerging wireless networks such as cognitive radio, ad hoc, sensor networks, disruption/delay tolerant networks, and distributed opportunistic networks

have been found to be more and more important in the network community. One common point between the corresponding scenarios is that nodes/mobile devices/base stations/users have to take some decisions in a quasi-autonomous manner and interact with each other under the form of interdependency such as multiuser interference, congestion cost, or capacity region in wireless networks.

Game theory, which is a branch of mathematics analyzing interactions between (inter-dependent) decision makers, appears as a natural paradigm to understand and design these types of wireless networks.

Below we mention some pioneer works on networking games. Starting from

- (i) The paper entitled *Datagram Network as a Game* by Nagle (RFC 970, 1985),
- (ii) *Multi-Objective Routing In Integrated Services Networks: A Game Theory Approach* by Economides, A.A. and Silvester J.A. in 1991,
- (iii) *A Game Theoretic Perspective to Flow Control in Telecommunication Networks* by Christos Douligeris & Ravi Mazumdar (1992),
- (iv) *Pricing in Computer Networks: Motivation, Formulation, and Example* by Ron Cocchi, Scott Shenker, Deborah Estrin, Lixia Zhang, and
- (v) *Competitive Routing Game* by Orda & Shimkin in 1993 in the first volume of IEEE Transaction in Networking,

the networking community is much interested in game theoretic modeling to understand strategic behaviors in wired and wireless networks (stochastic or deterministic).

In the last two decades, game theory has been applied more and more intensively to wireless networking and communications [5, 147, 78]. For instance, many distributed congestion control, routing algorithms, power control, and resource allocation schemes have been proposed by exploiting concepts of game theory. There is a quite general consensus to say that the corresponding analysis are useful but rely on strong information and behavior assumptions on the part of users (e.g., games with complete information and rational players, where rationality is common knowledge and predetermined beliefs). One of the consequences of this is that the predicted game outcomes (most often Cournot/Nash equilibria, Hannan set, correlated equilibria, communication equilibria, bargaining solutions, hierarchical solutions, etc.) will rarely be observed. One of the many purposes of learning procedures is to design fast and efficient algorithms exploiting partial information on the game. This chapter presents a unified view of learning algorithms with different levels of information. The goal is also to contribute to the understanding of learning in strategic

decision making that will become more acute with the increased deployment of wireless devices and limited resources.

The uncertainty/variability of the environment and the restrictive natures of wireless communications, for example, limited energy, battery, limited spectrum, scarce and shared communication medium, together with the resulting economical issues, make the distributed strategic learning approach a reasonable fit of approach to the wireless problems.

The framework covers scenarios where players are users/mobile devices/transmitters/base stations/network designers who are able to choose their actions by themselves. The actions can be typically

- to upgrade or not a technology, turn ON/OFF at device,
- to sense or not a channel,
- vector of powers in the case of a distributed power allocation problem or just a power level in the case of distributed power control in one dimension.
- to choose a subset of channels in the case of spectrum sharing and resource selection problems
- rate vectors in the case of rate allocation,
- choice of routes (concatenation of different nodes) from sources to destinations in the case routing and flow control, etc.

The payoff function can be, for example,

- the transmission rate or energy-efficiency, the cost of energy consumption per capacity, for power control/allocation problems [6].
- the individual/total throughput or mean delays, for the spectrum sharing problem,
- the cost, the queueing length, or the end-to-end delay for the routing problem.

Example 1.2.1. *As an example of interaction between nodes we consider multiple access problem in wireless networks [169]. Suppose that there are two mobile nodes $tx1$ and $tx2$ who want to access to a shared wireless channel to send information to the corresponding receivers $rx1$ and $rx2$. Both receivers are within the transmission range of both the transmitters. Each transmitter has one packet to transmit in each time step, and it can either choose to transmit during a time step or wait. If $tx1$ transmits, the packet is successfully*

transmitted if tx2 chooses not to transmit during that time step (and hence there is no collision). For successful packet transmission, tx1 obtains a benefit at the cost of transmit power. It is of interest to analyze the interactions between the transmitters under different network settings and performance objectives.

Example 1.2.2 (Energy management). *In wireless networks, most devices work on their own batteries. The limitation of battery power gets worse when devices communicate because much more power is consumed to transmit than to compute. As a result, optimal power consuming networks have been studied for various scenarios such as broadcasting, multicasting, routing, and target coverage (for wireless sensor networks). The power control problem for wireless data is to assign each device's transmission power to promote the quality and efficiency of the wireless networking system. In this manuscript, one of the main goals is the development of combined learning schemes such that the iterative power control algorithm converges both synchronously and asynchronously when outdated or incorrect interferences are used.*

Example 1.2.3. *Wireless sensor network is a network consisting of a large number of sensor devices with processing and wireless transmission capability. Wireless Sensor Networks (WSN) is considered as a class of ad hoc networks with the salient feature that it could have a centralized entity for the entire network called base station or sink. The sensors in the network sense different types of occurrences (e.g., temperature, magnetism, sound, light) from the surrounding environments and send the data to the base station. However, acquiring data from the deployment area of sensors is not always easy, and multiple issues are to be considered in dealing with a wireless sensor network. The major issue is the limited resources of the tiny sensor devices, which are run with one-time batteries. Even if rechargeable batteries are used, it is not always easy to replace them once the sensors are deployed over a geographical area. Most of the time, the application requires the network to be unattended, which means it is free from human intervention. Other problems might come from the external environment; heat, rain, wind condition, etc. Routing, data aggregation, ensuring Quality of Experience (QoE), Quality of Service (QoS), security, all these are various facets of wireless sensor networks. The interaction between the sensors can be modeled as an evolving with variable and nonreciprocal set of players. A typical objective function could be an end-to-end delay under long-term energy constraint (with renewable energy).*

The above payoff functions may depend on

- queue lengths,
- channel gains,
- location of the players,
- environment state, etc.

which are uncertain quantities. Thus, the resulting games are games under uncertainty, also called *robust games*. This class of games is frequently met in networks.

Below we provide examples where the class of robust games in wireless networking and communications is relevant.

- Outage probability

$$\mathbb{P}(\tilde{R}_j(\mathbf{w}, \mathbf{a}) < \beta_j) = \mathbb{E}_w \left(\mathbb{1}_{\{\tilde{R}_j(\mathbf{w}, \mathbf{a}) < \beta_j\}} \right)$$

- Success probability (P_{succ}) $\mathbb{P}(\tilde{R}_j(\mathbf{w}, \mathbf{a}) \geq \beta_j) = \mathbb{E}_w \left(\mathbb{1}_{\{\tilde{R}_j(\mathbf{w}, \mathbf{a}) \geq \beta_j\}} \right) = \mathbb{E}_w \left(\mathbb{1}_{\{ACK\}} \right)$

Energy-consumption based variants: $\frac{P_{succ}}{c_j(a_j)}, P_{succ} - c_j(a_j)$,

- Coverage probability

$$\mathbb{P} \left(\bigcup_{SBS} \tilde{R}_{j,SBS}(\mathbf{w}, \mathbf{a}) \geq \beta_j \right) = \mathbb{E}_w \left(\mathbb{1}_{\{\bigcup_{SBS} \tilde{R}_{j,SBS}(\mathbf{w}, \mathbf{a}) \geq \beta_j\}} \right)$$

- Ergodic Shannon-like payoff:

- $\mathbb{E}_w (\log(1 + SINR(\mathbf{w}, \mathbf{a})))$
- $\mathbb{E}_w (\log(1 + SINR(\mathbf{w}, \mathbf{a})) \mathbb{1}_{SINR_j \geq \beta_j})$
- Energy-based payoff:

$$\mathbb{E}_w \left(\sum_{j \in \mathcal{N}} \log(1 + SINR_j(\mathbf{w}, \mathbf{a})) - c_j(a_j) \right)$$

- Cognitive radio networks

- Primary users: $\mathbb{P}(SINR_{j'}(\mathbf{w}, \mathbf{a}) \geq \beta_{j'})$
- Secondary users: $\mathbb{P} \left(\frac{a_j |w_{j,rxPU}|^2}{(\epsilon^2 + d_{j,RxPU}^2)^{\alpha/2}} \leq \eta, SINR_j(\mathbf{w}, \mathbf{a}) \geq \beta_j \right)$ where $rxPU$ denotes the receiver of primary user and $d_{j,RxPU}$ is the current distance between user j and the receiver of the primary user $RxPU$.

- Energy market as a multi-level robust game

Consider a generic expected robust games in the energy market with three classes of players: producers of energy, distributors and users/customers with the following specifications

- Producers/generators (of different types of renewable energy)
 - * Each producer knows only her own production cost
 - * does not know the costs of the other producers

- * does not know the (exact) weather condition in advance (before making the decision on how much to produce)
- Distributors
 - * Each distributor has a contract with a subset of producers
 - * Contract with the users
- Users/Consumers
 - * Each user sees its realized prices per period/slot
 - * Users interact also through a network, application in the Cloud

We analyze the system by considering the expectation over type distribution, over environment distribution (weather), leading to expected robust games.

⚠ In the above example, there are *robust games* within the robust game.

1.3 Basic Robust Games

What is the difference between the classical one-shot games and robust games?

We have defined in [Subsection 1.1.1](#) the basic components of a one-shot game. If the description involves in addition a state parameter that can be deterministic or stochastic (uncertain), and which can be correlated with the action profile or independent, then one gets a robust game. In particular, if the uncertainty set is reduced to a singleton or an empty set, then we get a standard one-shot game.

Below we provide basic examples of robust games.

Example 1.3.1 (Robust games with dominated strategies). *An action is dominated if, regardless of what any other players do, the action earns a player a smaller payoff than some other action. Hence, an action is dominated if it is always better to play some other strategy, regardless of what opponents may do.*

In the context of robust games, a strategy is dominated if it is always better to play some other strategy, regardless of the states and what opponents may do.

Consider a two-player robust game. Let n_{aa}^i be an independent and identically distributed (i.i.d) random variable with zero mean and finite variance. [Table 1.2](#), the action profile (s_1, s'_1) is dominant.

$1 \backslash 2$	s'_1	s'_2		$n_{1,1}^1, n_{1,1}^2$	$n_{1,2}^1, n_{1,2}^2$
s_1	(100, 100)	(3, 10)	+	$n_{1,1}^1, n_{1,1}^2$	$n_{1,2}^1, n_{1,2}^2$
s_2	(3, 10)	(0, 0)		$n_{2,1}^1, n_{2,1}^2$	$n_{2,2}^1, n_{2,2}^2$

TABLE 1.2

Robust game with dominant strategy.

Robust games with dominant strategies can be solved using the so-called Iterated elimination of dominated strategies.

An example of such a game is a power control game in a multiple-input-single-output (MISO) system, where the maximum power is a dominant action.

Note that for some states of the robust game, the action may not be dominant but it can be a dominant of the expected payoff. This allow ones to extend the definition of robust game to expected robust game with dominant strategies.

Example 1.3.2 (Robust team game). The payoff function has the form

$$R_j(w, a) = \bar{r}(w, a).$$

Theorem 3.1.0.3 of [Chapter 3](#) characterizes the equilibrium payoffs of such team problems. This class of games can be generalized to potential robust games. See [Chapter 4](#).

Example 1.3.3 (Anti-coordination robust games). Consider the following two-player game. Each player can choose one of the two options. When they use the same action at the same time and same area, there is congestion or a collision (e.g, in same frequency band and same medium) and the data is lost. [Table 1.3](#) represents the two-by-two game. The actions are denoted by s_1 and s_2 . Player 1 chooses a row, and player 2 chooses a column. A successful configuration gives a payoff 1 to the corresponding player. In addition, we look at it in an independent and identically distributed (i.i.d) noisy environment. Let $0 < \alpha \leq 1$. In [table 1.3](#), n_{ab}^i : is a real-valued i.i.d noise. This class of game can be extended to robust crowd game and to robust congestion game, etc.

1 \ 2	s_1	s_2	+	$n_{1,1}^1, n_{1,1}^2$	$n_{1,2}^1, n_{1,2}^2$
s_1	$(0, 0)$	(α, α)		$n_{2,1}^1, n_{2,1}^2$	$n_{2,2}^1, n_{2,2}^2$
s_2	(α, α)	$(0, 0)$			

TABLE 1.3

Anti-coordination robust game.

1.4 Basics of Robust Cooperative Games

Cooperative game theory is concerned primarily with coalitions, that is, groups of players, who coordinate their actions and pool their payoffs. Consequently, one of the problems here is how to divide the payoff among the members of the formed coalition. The basis of this theory was laid by John von Neumann & Oskar Morgenstern [120] with coalitional games in characteristic function form, known also as transferable utility games (TU-games). The theory has been extended to non-transferable utility games (NTU-games). Robust cooperative games are cooperative games under uncertainties.

1.4.0.1 Preliminaries

Let $\mathcal{N}$ be a non-empty finite set of players who consider different cooperation possibilities. Each subset $\mathcal{N}' \subseteq \mathcal{N}$ is referred to as a crisp coalition. The set $\mathcal{N}$ is called the grand coalition and $\emptyset$ is called the empty coalition.

Definition 1.4.0.2. *A cooperative game in characteristic function form is an ordered pair $(\mathcal{N}, v)$ consisting of the player set $\mathcal{N}$ and the characteristic function $v : 2^{\mathcal{N}} \rightarrow \mathbb{R}$, with $v(\emptyset) = 0$ and $2^{\mathcal{N}}$ is the set of subsets of $\mathcal{N}$.*

A cooperative game in characteristic function form is usually referred to as a transferable utility game (TU-game). The real number $v(\mathcal{N}')$ can be interpreted as the maximal payoff that the members of $\mathcal{N}'$ can obtain when they cooperate.

We now address one of the basic questions in the theory of cooperative TU-games:

If the grand coalition forms, how to divide the utility $v(\mathcal{N})$?

This question is approached using different solution concepts in cooperative game theory like cores, stable sets, bargaining sets, the Shapley value, the nucleolus. A solution concept gives an answer to the question of how the payoff obtained when all players in $\mathcal{N}$ cooperate should be distributed among

the individual players while taking account of the potential utility of all different coalitions of players. Hence, a solution concept assigns to a coalitional game at least one payoff vector $r = (r_j)_{j \in \mathcal{N}}$, where r_j is the payoff allocated to player $j \in \mathcal{N}$.

Denote by $\mathcal{G}_{\mathcal{N}}$ the set of characteristic functions of coalitional games with player set $\mathcal{N}$ forms with the usual operations of addition and scalar multiplication of functions a $(2^{|\mathcal{N}|} - 1)$ -dimensional linear space. A basis of this linear space is obtained by value function $v_{\mathcal{N}'}^*(\mathcal{N}'') = \mathbf{1}_{\{\mathcal{N}'' \subseteq \mathcal{N}'\}}$ for $\mathcal{N}' \in 2^{\mathcal{N}} \setminus \{\emptyset\}$. This class cooperative games with value v^* are called *unanimity games*. An interpretation is unanimity game is the following: In the game $v_{\mathcal{N}'}^*$, the payoff one can be obtained if and only if all players in coalition $\mathcal{N}'$ are involved in cooperation.

Definition 1.4.0.3. Let $r : \mathcal{G}_{\mathcal{N}} \rightarrow \mathbb{R}^n$. Then, r satisfies

- *Individual rationality* if $r_j(v) \geq v(\{j\})$, $\forall (v, j) \in \mathcal{G}_{\mathcal{N}} \times \mathcal{N}$
- *Efficiency* if $\sum_{j=1}^n r_j(v) = v(\mathcal{N})$, $\forall v \in \mathcal{G}_{\mathcal{N}}$.
- *Additivity* if $r(v + v') = r(v) + r(v')$, $\forall (v, v')$
- *the dummy player property* if $r_j(v) = v(\{j\})$ for all v and for all dummy players k such that $v(\mathcal{N}' \cup \{j\}) = v(\mathcal{N}') + v(\{j\})$ for all $\mathcal{N}' \in 2^{\mathcal{N}} \setminus \{j\}$.
- *the anonymity property* if $r(v_{\pi}) = \pi^*(r(v))$ for all $\pi \in S_n$. Here v_{π} is the game with $v_{\pi}(\pi(\mathcal{N}')) := v(\mathcal{N}')$ for all $\mathcal{N}' \in 2^{\mathcal{N}}$ and $\pi^* : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined by $(\pi^*(r))_{\pi(j)} := r_j$ for all $r \in \mathbb{R}^n$ and $j \in \mathcal{N}$.

1.4.0.4 Cooperative Solution Concepts

Definition 1.4.0.5. A payoff vector $r \in \mathbb{R}^n$ is an *imputation* for the game $v \in \mathcal{G}_{\mathcal{N}}$ if it is efficient and individually rational, i.e.

- $\sum_{j \in \mathcal{N}} r_j = v(\mathcal{N})$,
- $r_j \geq v(\{j\})$, $\forall j \in \mathcal{N}$

We denote by $I(v)$ the set of imputations of $v \in \mathcal{G}_{\mathcal{N}}$. Clearly, $I(v)$ is empty if and only if $v(\mathcal{N}) < \sum_{j \in \mathcal{N}} v(\{j\})$.

We now define set-valued solution concepts called **core** [64].

Definition 1.4.0.6 (Core). The *core* $C(v)$ of a TU-game $v \in \mathcal{G}_{\mathcal{N}}$ is the set

$$C(v) = \left\{ r \in I(v), \sum_{j' \in \mathcal{N}'} r_{j'} \geq v(\mathcal{N}'), \forall \mathcal{N}' \in 2^{\mathcal{N}} \setminus \{\emptyset\} \right\}$$

If $r \in C(v)$, then no coalition $\mathcal{N}'$ has an incentive to split off if r is the proposed payoff allocation in $\mathcal{N}$, because the total amount $\sum_{j' \in \mathcal{N}'} r_{j'}$ allocated to $\mathcal{N}'$ is not smaller than the amount $v(\mathcal{N}')$ which the players can obtain by forming the sub-coalition. If the core is non-empty, then the core is a polytope because bounded and it is a finite system of linear inequalities.

Definition 1.4.0.7 (Balanced game). Denote by $e_{\mathcal{N}'}$ the characteristic vector of $\mathcal{N}'$ i.e $e_{\mathcal{N}',i} = \mathbf{1}_{\{i \in \mathcal{N}'\}}$.

- *Balanced map:* $\lambda : 2^{\mathcal{N}} \setminus \{\emptyset\} \longrightarrow \mathbb{R}_+$ is a balanced map if

$$\sum_{\mathcal{N}' \in 2^{\mathcal{N}} \setminus \{\emptyset\}} \lambda_{\mathcal{N}'} e_{\mathcal{N}'} = e_{\mathcal{N}}$$

- A collection $\mathcal{N}'$ of coalitions is called balanced if there is a balanced map λ such that

$$\mathcal{N}' = \{\mathcal{N}'' \in 2^{\mathcal{N}} \mid \lambda_{\mathcal{N}''} > 0\}$$

(the support of λ).

- A game v is balanced if for each balanced map $\lambda : 2^{\mathcal{N}} \setminus \{\emptyset\} \longrightarrow \mathbb{R}_+$ one gets

$$\sum_{\mathcal{N}' \in 2^{\mathcal{N}} \setminus \{\emptyset\}} \lambda_{\mathcal{N}'} v(\mathcal{N}') = v(\mathcal{N}).$$

The next theorem gives a characterization of games with a nonempty core.

Theorem 1.4.0.8. Let $v \in \mathcal{G}_{\mathcal{N}}$. Then the following two assertions are equivalent:

- $C(v) \neq \emptyset$.
- The game v is balanced.

We now focus on the notion of stable set [120].

We say that an imputation r dominates another imputation r' via coalition $\mathcal{N}'$, if

- $r_j > r'_j$ for all $j \in \mathcal{N}'$, this says that the payoff r is better than the payoff r' for all members of $\mathcal{N}'$.

- $\sum_{j \in \mathcal{N}'} r_j \leq v(\mathcal{N}')$. This guarantees that the payoff r is reachable for the coalition $\mathcal{N}'$.

For $v \in \mathcal{G}_{\mathcal{N}}$, we denote by $\text{dom}(V)$ the set consisting of all imputations that are dominated by some element in V . The dominance core $DC(v)$ is defined by $DC(v) = I(v) \setminus \text{dom}(I(v))$.

Definition 1.4.0.9. For $v \in \mathcal{G}_{\mathcal{N}}$ a subset $\mathcal{N}'$ of $I(v)$ is called a stable set if the following conditions hold:

- (Internal stability) $\mathcal{N}' \cap \text{dom}(\mathcal{N}') = \emptyset$.
- (External stability) $I(v) \setminus \mathcal{N}' \subset \text{dom}(\mathcal{N}')$.

Note that for a game v the set $\mathcal{N}'$ is a stable set if and only if $\mathcal{N}'$ and $\text{dom}(\mathcal{N}')$ form a partition of $I(v)$.

Definition 1.4.0.10. • A game v is superadditive if $v(\mathcal{N}' \cap \mathcal{N}'') \geq v(\mathcal{N}') + v(\mathcal{N}'')$ for all $(\mathcal{N}', \mathcal{N}'') \in 2^{\mathcal{N}}$ with $\mathcal{N}' \cap \mathcal{N}'' = \emptyset$

- A game v is subadditive if $-v$ is superadditive.

Theorem 1.4.0.11. Let $v \in \mathcal{G}_{\mathcal{N}}$ and $\mathcal{N}'$ be a stable set of v . Then

- $C(v) \subseteq DC(v) \subseteq \mathcal{N}'$;
- If v is superadditive, then $DC(v) = C(v)$;
- If $DC(v)$ is a stable set, then there is no other stable set.

We now focus on the Shapley value. The Shapley value uses the marginal vectors of a cooperative TU-game. First we define a marginal vector.

Definition 1.4.0.12 (Shapley value). • Let S_n is the set of permutation $\mathcal{N}$. Let $\pi \in S_n$. The set $p_{\pi,j} = \{j' \in \mathcal{N}, \mid \pi^{-1}(j') < \pi^{-1}(j)\}$ consists of all predecessors of j with respect to the permutation. The marginal contribution vector $m_{\pi,j} \in \mathbb{R}^n$ with respect to π and v has the j -th coordinate

$$m_{\pi,j}(v) = v(p_{\pi,j} \cup \{j\}) - v(p_{\pi,j}), \quad j \in \mathcal{N}$$

- The Shapley value Φ of a game v is the average of the marginal vectors of the game: $\Phi(v) = \frac{1}{n!} \sum_{\pi \in S_n} m_{\pi}(v)$

Theorem 1.4.0.13. A solution $r : \mathcal{G}_{\mathcal{N}} \rightarrow \mathbb{R}^n$ satisfies additivity, anonymity, the dummy player property, and efficiency if and only if it is the Shapley value.

For cooperative robust games, we define the value as $v : \mathcal{W} \times 2^{\mathcal{N}}$, defined by $v(w, J)$ where w is the uncertainty parameter and J is a coalition. Then, the robust cooperative game is given by $(\mathcal{N}, v, \mathcal{W})$, where $\mathcal{W}$ is the uncertainty set. For a cooperative expected robust game, the allocated expected payoff will be defined as $\bar{r}_j = \mathbb{E}_w r_j(w)$. The notions of expected robust core and expected robust Shapley value are defined in a similarly way.

FIGURE 1.1

Strategic Learning.

1.5 Distributed Strategic Learning

What is distributed strategic learning?

Distributed strategic learning is learning in the presence of other players (i.e., from, with, or about others) or (from, with or about) Nature/environment. The term *strategic learning* stands for how the players can learn from/with/about the others (players, environment) under their complex strategies/beliefs. A strategic learning consists of acting (eventually optimally or for satisfactory criterion) when outcomes are interdependent and interactive (depend on others).

In this book, distributed strategic learning is a learning-based process for creating and executing breakthrough payoff functions, strategies, and distribution laws over time. The strategic learning aspect includes learning how to communicate better but also communicating to learn better. When the players can act locally with joint decisions during the interactions, we refer to it as *coalitional strategic learning*. In [figure 1.1](#), we represent some of the learning topics related to distributed strategic learning.

Strategy learning and dynamics

We use the term *strategy learning* or *strategy dynamics* to emphasize to learning process for strategies, optimal strategies, better response strategies, satisfactory strategies, equilibrium strategies, etc.

Belief learning and dynamics

We use the term *belief learning* to refer to a process of learning the other players beliefs or one's own beliefs.

Payoff learning and dynamics

We use the term payoff learning or payoff dynamics to refer to a process of learning the payoff functions, the optimal payoffs, the equilibrium payoffs, etc. See [Chapter 3](#).

Combined learning

The term *combined learning* is used to emphasis a joint learning process. For example, learning simultaneously the payoff functions and the optimal (or satisfactory) strategies is a combined learning.

[Figure 1.2](#) represents a generic combined learning scheme. There are n entities (players) in a dynamic interactive system. Each player can make a decision and is able to measure a reaction from the system. Each player updates her or his decision based on their own-measurements without information exchange among the players. The environment parameters can be stochastic or deterministic and outcomes depend on the players decisions.

Heterogeneous learning

The term *heterogeneous learning* is used to emphasize that it refers to differences in the ways players update and learn. It does not necessarily refer to the structural heterogeneity of the model. For example, heterogeneous learning can be considered in symmetric robust games with simultaneous moves.

We will examine five types of heterogeneity: players that

FIGURE 1.2

A generic combined learning scheme.

- have different perceptions
- have different temperature-like parameters and degrees of inertia in updating and
- use different learning patterns
- use different learning rates (slow learners, medium learners, fast learners, etc.)
- different updating times, synchronicity, and activities.

In the next chapters, if at least two of the players have distinct learning patterns, we will use the term heterogeneous learning.

The asymptotic results of convergence of heterogeneous learning are examined in [Chapter 4](#). Moreover, we do not neglect the issue of how heterogeneity affects the speed of convergence of the interacting system towards equilibrium.

Hybrid learning

Different from heterogeneous learning, the term *hybrid learning* will be used in the case where at least one of the players changes the learning pattern during the interactions. At each time, each active player will be able to select from a list of learning patterns. The resulting dynamics are related to *hybrid game dynamics* in which each player can switch between different rules called *revision protocols*.

CODIPAS: COmbined fully DIstributed PAYoff and Strategy

CODIPAS is a combined learning. It consists of learning jointly and in an interdependent manner payoffs and strategies in long-run games.

Players	Information	CODIPAS: Summary
Actions	Own-action	Make a decision based on own-measurement and strategies.
Observation	Numerical measurement	New numerical payoff values are perceived.
Response	Learning pattern	Update the strategies and the estimates based on the recent reaction.

CODIPAS-RL : COmbined fully DIstributed PAYoff and Strategy Reinforcement Learning

CODIPAS-RL is a reinforcement-based CODIPAS. This class of learning will be developed in [Chapter 4](#). The basic idea of CODIPAS can be summarized as follows:

- The consequences influence behavior.
- The behaviors influence the outcomes.

The idea of CODIPAS-RL goes back at least to Bellman (1957, [23]) for joint computation of optimal value and optimal strategies. The idea has been followed in [19] with less information requirement. Since then, different versions of CODIPAS-RLs has been studied. These are:

- Bush-Mosteller-based CODIPAS-RLs: combine the scheme of Bush-Mosteller with the perceived payoff averaging (per action).
- Boltzman-Gibbs-based CODIPAS-RLs: construct Gibbs distribution

scheme based on estimated payoffs (per action). The main difference between Boltzman-Gibbs CODIPAS-RLs and the classical Boltzman-Gibbs dynamics or logit dynamics is in terms of information. In logit dynamics the vector payoff corresponding to all the actions of the player are required in order to update. In the Boltzman-Gibbs-based CODIPAS-RLs, the payoff corresponding to the others actions than the current action are not observed. Each player needs to estimate these alternatives.

- Imitative Boltzman-Gibbs-based CODIPAS-RLs: Combine imitative rules that are proportional to the Gibbs distribution.
- Multiplicative weighted Imitative CODIPAS-RLs: Instead of the exponential function, an imitation rule based on multiplicative weights is considered.
- Weakened fictitious play based CODIPAS-RLs: Construct a stochastic fictitious play scheme based only on the payoff estimations. The actions of the other players are not observed. The frequency of others play is also unknown.
- Logit-based payoff learning: The payoff dynamics is based on logit map.
- Imitative Boltzmann-Gibbs-based payoff learning.
- Multiplicative weighted payoff learning.
- No-regret based CODIPAS-RLs.
- Pairwise comparison based CODIPAS-RLs: Construct a strategy learning based on a rate that is proportional to the gap between two estimates.
- Projection-based CODIPAS-RLs.
- Excess-payoff-based CODIPAS-RL.
- Risk-sensitive CODIPAS.
- Cost-to-Learn CODIPAS, etc

Mean field learning

The term *mean field learning* refers to a learning framework for large populations of players. It is different from *learning the mean field* which is related to transport equations and Fokker-Planck-Kolmogorov (FPK) forward equations.

Risk-sensitive learning

The term *risk-sensitive learning* refers to learning in the presence of risk-sensitive players. It consists of learning the risk-sensitive payoffs, risk-sensitive equilibrium, and or other risk-sensitive solutions. See [Chapter 9](#).

Risk-sensitive payoff dynamics

We use the term *risk-sensitive payoff dynamics* to refer to a process of learning the risk-sensitive payoff functions, the optimal risk-sensitive payoffs, the risk-sensitive equilibrium payoffs, etc. The risk-neutral payoff dynamics is simply called payoff dynamics. See [Chapter 3](#).

Risk-sensitive CODIPAS

Joint payoff and strategy learning in the risk-sensitive context.

Slow, Medium, or Fast Learners

We refer to players with negligible learning rates compared to the others as *slow learners*. The fast learners have high learning rates. The medium learners are the players with learning rates between the slow learners' rates and the fast learners' rates. The other configurations are considered as *proportional learning rates*.

Evolutionary game dynamics

Evolutionary games in a large population provide a simple framework for describing strategic interactions among large numbers of players. Traditionally, predictions of behavior in game theory are based on some notion of equilibrium, typically Cournot equilibrium, Bertrand equilibrium, Nash equilibrium, Stackelberg solution, Wardrop equilibrium, mean field equilibrium, resilient by invasion, or some refinement/extension thereof. These notions require the assumption of equilibrium knowledge, which posits that each user correctly anticipates how the other players will act or react. The equilibrium

knowledge assumption is too strong and is difficult to justify, in particular, in contexts with large numbers of players.

As an alternative to the equilibrium approach, an explicitly dynamic updating choice is proposed, a model in which players myopically update their behavior in response to their current strategic environment. This dynamic procedure does not assume the automatic coordination of players actions and beliefs, and it can derive many specifications of players' choice procedures. These procedures are specified formally by defining a revision of actions called *revision protocol*. A revision protocol takes current payoffs (also called *fitness* in behavioral ecology) and aggregate behavior as inputs; its outputs are conditional switch rates, which describe how frequently players in some class playing action a who are considering switching action switch to another action b , given the current payoff vector and subpopulation state. Revision protocols are flexible enough to incorporate a wide variety of paradigms, including ones based on imitation, adaptation, learning, optimization, etc. The revision protocols describe the procedures players follow in adapting their behavior to the dynamic evolving environment such as evolving networks. The methodology is extended to delayed evolutionary game dynamics as well as stochastic evolutionary game dynamics. For recent developments of evolutionary dynamics (deterministic or stochastic), we refer the reader to [140, 158, 160] and the references therein.

1.5.1 Convergence Issue

Evolutionary game dynamics provides a powerful tool for prediction of the outcome of a game. Convergence to equilibria has been established in many classes of games. These include generic two-players-two-action games, common interest games, potential games, S-modular games, many classes of aggregative games, games with unique evolutionarily stable strategy, stable games, games with monotone payoffs, etc. As a consequence, convergence issue of networking problems including parallel routing, routing with M/M/1 cost (Poisson arrival process, exponentially distributed service time and single server queue), network congestion games, network selection games, power allocation games, rate control, spectrum access games, resource sharing games, and many others, can be investigated through evolutionary game dynamics for both linear and non-linear payoff functions.

1.5.2 Selection Issue

A fundamental question we address is the selection problem (selection of equilibrium, Pareto optimal solutions, global optimum, etc.) in a *fully distributed way* (minimal signaling to the users, no message exchange, no recommendation, etc.).

1.5.2.1 How to Select an Efficient Outcome?

The problem of selection of a global optimum in a fully decentralized way is a very challenging problem. Very little is known about this. In [Chapter 8](#) we provide examples of games where fully distributed reinforcement learning algorithms lead to evolutionary game dynamics that converge to global optima. However, in general, evolutionary game dynamics may not converge.

1.5.2.2 How to Select a Stable Outcome ?

Another important question is the stability/instability of the system. How to design fast algorithms such that a network will behave well after some iterations? This question will be translated by the fact that when time horizon is large, the behavior of the system looks like at a stationary point or set, and any small perturbation around this point or set will quickly come back to that point or set due to losses. It should be noticed that the two properties, *Stability* and *Efficiency*, may not be compatible in some situations.

Simulated Annealing

There are certain optimization problems that become unmanageable using combinatorial methods as the number of objects (players, resources, etc.) becomes large. This is the case for NP-complete problem, a problem which is both NP (verifiable in nondeterministic polynomial time) and NP-hard. The optimization community has been using *simulated annealing* for many years. Annealing is a metallurgical process in which a material is heated and then slowly brought to a lower temperature. The crystalline structure of the material can reach a global minimum in this way. The high temperature excites all atoms or molecules, but later, during the cooling phase, they have enough time to assume their optimal position in the crystalline lattice and the result is fewer fractures and fewer irregularities in the crystal.

Example 1.5.3. *A typical example is the traveling salesman problem, which belongs to the NP-complete class of problems. For these problems, there is a very effective practical algorithm called simulated annealing. While this technique is unlikely to find the optimum solution, it can often find a very good solution, even in the presence of noisy data.*

The traveling salesman problem can be used as an example application of simulated annealing. In this problem, a salesman must visit some large number of cities while minimizing the total mileage traveled. If the salesman starts with a random itinerary, he can then pairwise trade the order of visits to cities, hoping to reduce the mileage with each exchange. The difficulty with this approach is that while it rapidly finds a local minimum, it cannot get from there to the global minimum.

1.6 Distributed Strategic Learning in Wireless Networks

“It can be very costly to learn quickly, and learning can take some time”.

1.6.1 Physical Layer

Distributed strategic learning has recently been applied to determine the transmission power of mobile nodes. In [84], the authors model the distributed power control problem in an infrastructure wireless network as a nonzero-sum game between the mobile nodes. Distributed learning algorithms for distributed solution of the power control game are proposed. Each node operates a learning algorithm, which determines the probability of choosing a certain transmission power based on the feedback received from the Base Station (BS). This feedback is essentially the environmental response for choosing a certain power level and expresses the amount of information that a mobile node can transfer during the lifetime of its battery. However, the proposed algorithms can also operate for other definitions of the environmental response as well.

In [191], the authors considered Bush-Mosteller-based reinforcement learning to analyze power control in wireless network. They conducted a detailed stability analysis for the two-player-two-action case. However, the stability analysis for arbitrary number of players and or the associated performance metrics remains a challenging open problem.

Rate control

Rate control is one of the most important problems in networks. If input rates to a network are left uncontrolled, then the offered load may exceed the network capacity causing congestion, i.e., buffers fill up and packets have to be dropped. It is well known that network congestion has a negative impact on performance, potentially leading to a throughput collapse. Rate control schemes should therefore

- produce a predictable and stable throughput,
- allow a distributed implementation, and
- allocate network bandwidth among applications with a certain fairness criterion (proportional fairness, maxmin fairness, α -fairness, Bargaining solutions, etc.).

In [84, 68] the authors proposed a Stochastic Automata Rate Adaptation

References

- M. Aghassi and D. Bertsimas . Robust game theory. *Mathematical Programming*, 107(1):231–273, 2006.
- C. Alós-Ferrer and N. Netzer . The logit-response dynamics. *Games and Economic Behavior*, 68(2):413–427, March 2010.
- T. Alpcan and T. Başar . *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2010.
- E. Altman , T. Basar , I. Menache , and H. Tembine . A dynamic random access game with energy constraints. In *Proceedings 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, (WiOpt'09)*, 2009.
- E. Altman , T. Boulogne , R. El-Azouzi , T. Jimenez , and L. Wynter . A survey on networking games in telecommunications. *Computers and Operations Research*, 2006.
- E. Altman , R. El-Azouzi , Y. Hayel , and H. Tembine . Evolutionary power control games in wireless networks. In *Proceedings of the 7th international IFIP-TC6 networking conference on AdHoc and sensor networks, wireless networks, next generation internet, NETWORKING'08*, pages 930–942, Berlin, Heidelberg, 2008. Springer-Verlag.
- E. Altman , R. ElAzouzi , Y. Hayel , and H. Tembine . An evolutionary game approach for the design of congestion control protocols in wireless networks. *6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops, WiOPT*, 2008.
- Eitan Altman , Rachid El Azouzi , Yezekael Hayel , and H. Tembine . The evolution of transport protocols: An evolutionary game perspective. *Computer Networks*, 53(10):1751–1759, 2009.
- K. B. Ariyur and M. Krstic . *Real Time Optimization by Extremum Seeking*. Hoboken, NJ: Wiley, 2003.
- Kenneth J. Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22:265–290, July 1954.
- W.B. Arthur . On designing economic agents that behave like human agents. *Journal Evolutionary Econ.* 3, pages 1–22, 1993.
- P. Auer , N. Cesa-Bianchi , and P. Fischer . Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.
- R. J. Aumann . Agreeing to disagree. *The Annals of Statistics*, 4:1236–1239, 1976.
- R. J. Aumann . Rationality and bounded rationality. *Games and economic behavior*, 21:2–14, 1997.
- Robert Aumann and Sergiu Hart . *Handbook of Game Theory with Economic Applications*, volume 1. *Handbooks in Economics*, North-Holland, 1992.
- Robert Aumann and Sergiu Hart . *Handbook of Game Theory with Economic Applications*, volume 2. *Handbooks in Economics*, North-Holland, 1994.
- Robert Aumann and Sergiu Hart . *Handbook of Game Theory with Economic Applications*, volume 3. *Handbooks in Economics*, North-Holland, 2002.
- A. Azad and H. Tembine . Dynamic routing games: An evolutionary game theoretic approach. *CDC-ECC, 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, Florida, December 12–15 2011.
- A.G. Barto , R.S. Sutton , and C. Anderson . Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*, 13:834–846, 1983.
- T. Basar and G. J. Olsder . *Dynamic noncooperative game theory*. *SIAM Series in Classics in Applied Mathematics*, Philadelphia, January 1999.
- M. Beckmann , C. B. McGuire , and C. B. Winston . *Studies in the economics of transportation*. Yale University Press, 1956.
- R. Bellman . On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the U.S.A.*, 38:716–719, 1952.
- R.E. Bellman . A markov decision process. *Journal of Mathematical Mech.*, 6:679–684, 1957.
- E.V. Belmega , H. Tembine , and S. Lasaulce . Learning to precode in outage minimization games over mimo interference channels. *The IEEE Asilomar Conference on Signals, Systems, and Computers*, California, US, November 2010.
- E. Ben-Porath and M. Kahneman . Communication in repeated games with costly monitoring. *Games and Economic Behavior*, 44:227–250, 2003.
- M. Benaïm . Dynamics of stochastic approximations. *Le Seminaire de Probabilites. Lectures Notes in Mathematics*, 1709:1–68, 1999.
- M. Benaïm and M. Faure . Stochastic approximations, cooperative dynamics and supermodular games. Preprint, 2010.
- M. Benaïm , J. Hofbauer , and S. Sorin . Stochastic approximations and differential inclusions. Part II: Applications, *Mathematics of Operations Research*, 2006.
- A. Benveniste , P. Priouret , and M. Metivier . *Adaptive algorithms and stochastic approximations*. Springer-Verlag New York, Inc. New York, NY.
- Ulrich Berger . Fictitious play in $2 \times n$ games. *Journal Econo. Theory*, pages 139–154, 2005.
- D.T. Bishop and C. Cannings . The war of attrition with random rewards. *Journal of Theoretical Biology*, 74:377–388, 1978.
- T. Borgers and R. Sarin . Learning through reinforcement and replicator dynamics. Mimeo, University College London., 1993.
- V. S. Borkar . *Stochastic approximation: a dynamical systems viewpoint*. New Delhi: Hindustan Publishing Agency, and Cambridge, UK.
- V. S. Borkar . Stochastic approximation with two timescales. *Systems Control Lett.*, 29:291–294, 1997.
- Diaconis P. Xiao L. Boyd, S. Fastest mixing markov chain on a graph. *SIAM Rev.*, 46:667–689, 2004.
- O. Brandière and M. Duflo . Les algorithmes stochastiques contournent-ils les pièges?. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 32:395–427, 1996.
- Haim Brezis . *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, ISBN 978-0-387-70913-0, 2011.
- Skyrms Brian and Robert Pemantle . A dynamic model of social network formation. In *Proceedings of the National Academy of Sciences*, 97(16):9340–9346, 2000.
- W. Brian Arthur . Inductive reasoning and bounded rationality. *American Economic Review*, 84:406–411, 1994.
- G. W. Brown . Some notes on computation of games solutions. Report P-78, The Rand Corporation, 1949.
- G.W. Brown . Iterative solutions of games by fictitious play,. In *Activity Analysis of Production and Allocation*, T.C. Koopmans (Ed.), New York: Wiley., 1951.
- R. Bush and F. Mosteller . *Stochastic models of learning*. Wiley Sons, New York., 1955.
- A. Cabrales . Stochastic replicator dynamics. *International Economic Review*, 41:451–481, 2000.
- U. O. Candogan , I. Menache , A. Ozdaglar , and P. A. Parrilo . Near-optimal power control in wireless networks: A potential game approach. *Annual IEEE International Conference on Computer Communications*, 2010.

R.G. Cassidy , C.A. Field , and M.J.L. Kirby . Solution of a satisficing model for random payoff games. *Management Science*, 19:266–271, 1972.

N. Cesaro-Bianchi and G. Lugosi . *Prediction, Learning and Games*. Cambridge University Press, 2006.

R. Cominetti , E. Meloy , and S. Sorin . A simple model of learning and adaptive dynamics in traffic network games. 2007.

P. Coucheney , C. Touati , and B. Gaujal . Selection of efficient pure strategies in allocation games. In *Proceedings of the International Conference on Game Theory for Networks*, 2009.

Augustin Cournot . *Recherches sur les principes mathematiques de la theorie des richesses*. Paris, France: L. Hachette, 1838; Italian translation in *Biblioteca DelliEcon.*, 1875; English translation by N. T. Bacon in *Economic Classics*, New York, NY: Macmillan, 1897; reprinted by Augustus M. Kelly , 1960., 1838.

T. M. Cover and J. A. Thomas . *Elements of Information Theory*. New York, Wiley, 1991.

G. Debreu . A social equilibrium existence theorem. *Proceedings of the National Academy of Science*, 38:886–893, 1952.

P. Diaconis and D. Stroock . Geometric bounds for eigenvalues of markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

M. Dindos and C. Mezzetti . Better-reply dynamics and global convergence to nash equilibrium in aggregative games. *Games and Economic Behavior*, 54:261–292, 2006.

S. Ellner . Ess germination strategies in randomly varying environments. i. logistic-type models. *Theoretical Population Biology*, 28:50–79, 1985.

Behdis Eslamnour , S. Jagannathan , and Maciej Zawodniok . Dynamic channel allocation in wireless networks using adaptive learning automata. *International Journal of Wireless Information Networks*, pages 1–14, May 2011.

S. Fischer , H. Räcke , and B. Vöcking . Fast convergence to wardrop equilibria by adaptive sampling methods. *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 653–662, 2006.

J. Flesch , G. Schoenmakers , and K. Vrieze . Stochastic games on a product state space. *Math. Oper. Res.*, 33:403–420, 2008.

D. Foster and R. V. Vohra . Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

D. Fudenberg and D. Kreps . A theory of learning, experimentation, and equilibrium in games. 1988.

D. Fudenberg and D. Levine . *Learning in Games*. MIT Press, Cambridge, MA, 1998.

D. Fudenberg and S. Takahashi . Heterogeneous beliefs and local information in stochastic fictitious play. 2008.

J. Willard Gibbs . On the equilibrium of heterogeneous substances. *Connecticut Acad. Sci.*, 1875-1878.

I. Gilboa and A. Matsui . Social stability and equilibrium. *Econometrica*, 59:859–867, 1991.

D.B. Gillies . Some theorems on n-person games. Ph.D. Thesis, Princeton University Press, Princeton, 1953.

AJ Goldman . The probability of a saddle point, 64:729–730, 1957.

Sanjeev Goyal . *Learning in networks: a survey*. Group Formation in Economics: Networks, Clubs, and Coalitions edited by G. Demange and M. Wooders . Cambridge University Press., 2003.

J. Guckenheimer and P. Holmes . *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.

M.A. Haleem and R. Chandramoulli . Adaptive downlink scheduling and rate selection: a cross-layer design. *IEEE Journal On Selected Areas in Communications*, 23(6), June 2005.

S. Hart and Y. Mansour . How long to equilibrium? the communication complexity of uncoupled equilibrium procedures. *Games and Economic Behavior*, 69:107–126, 2010.

S. Hart and A. Mas-Colell . A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

Y. Hayel , H. Tembine , E. Altman , and R. ElAzouzi . A markov decision evolutionary game for individual energy management. *Annals vol. XI of ISDG International Society of Dynamic Games, Advances in Dynamic Games: Theory, Applications, and Numerical Methods*, Editors: Michele Breton , Krzysztof Szajowski , pages 313–335, 2010.

G. He , H. Tembine , and M. Debbah . Deployment analysis of cooperative ofdm base stations. *IEEE International Conference on Game Theory for Networks (GameNets)*, Istanbul, Turkey, 2009.

J. Hofbauer and L. Imhof . Time averages, recurrence and transience in the stochastic replicator dynamics. *Annals of Applied Probability*, 19:1347–1368, August 2009.

J. Hofbauer and K. Sigmund . *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

J. Hofbauer , S. Sorin , and Y. Viossat . Time average replicator and best-reply dynamics. *Mathematics of Operations Research*, 34(2):263–269, May 2009.

Ed Hopkins and M. Posch . Attainability of boundary points under reinforcement learning. *Games and Economic Behavior*, 53:110–125, 2005.

A.I. Houston and J. McNamara . Fighting for food: a dynamic version of the hawk-dove game. *Evolutionary Ecology*, 2:51–64, 1988.

J. W. Huang and V. Krishnamurthy . Transmission control in cognitive radio as a markovian dynamic game: Structural result on randomized threshold policies. *IEEE Transactions on Communications*, 58(1), Jan. 2010.

Kaplansky I. A contribution to von neumann's theory of games. *Ann. of Math.*, 2(46):474–479, 1945.

S. Ishikawa . Fixed points by a new iteration method. *Proceedings Amer. Math. Soc.* 44, (1):147–150, 1974.

Luis R. Izquierdo , Segismundo S. Izquierdo , Nicholas M. Gotts , and J. Gary Polhill . Transient and asymptotic dynamics of reinforcement learning in games. *Games and Economic Behavior*, 61:259–276, 2007.

Matthew Jackson and Watts Allison . On the formation of interaction networks in social coordination games. *Games and Economic Behavior*, 41(2):265–291, 2002.

D. H. Jacobson and D. Q. Mayne . *Differential dynamic programming*. American Elsevier Pub. Co., New York, NY, 1970.

T. Joshi , D. Ahuja , D. Singh , and D. Pagrawal . Sara: Stochastic automata rate adaptation for ieee 802.11 networks. *IEEE Transactions on Parallel and Distributed Systems*, 19(10):1579–1590, October 2008.

Shizuo Kakutani . A generalization of brouwer's fixed point theorem. 8(3):457–459, 1941.

W. Karush . Minima of functions of several variables with inequalities as side constraints. M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois., 1939.

G. Kasbekar and A. Proutiere . Opportunistic medium access in multichannel wireless systems: A learning approach. in *Proceedings of Allerton Conference on Communications, Control, and Computing*, September 2010.

M. Khan and H. Tembine . *Energy-efficiency networking games*. Green It: Technologies and Applications, Springer Verlag, 2011.

M. Khan , H. Tembine , and A. Vasilakos . Game dynamics and cost of learning in heterogeneous 4g networks. *IEEE Journal on Selected Areas in Communications*, 30(1):198–213, January 2012.

M. Khan , H. Tembine , and T. Vasilakos . Evolutionary coalitional games: Design and challenges in wireless networks. *IEEE Wireless Communications Magazine*, 2012.

Manzoor Ahmed Khan and H Tembine . Evolutionary coalitional games in network selection. *Wireless Advanced 20th-22nd*, June 2011.

Manzoor Ahmed Khan , H. Tembine , and Stefen Marx . Learning in user-centric iptv services selection in heterogeneous wireless network. *IEEE INFOCOM 2011 International Workshop on Future Media Networks and IP-based TV*, 2011.

J. Kiefer and J. Wolfowitz . Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23:462–466, 1952.

Alan Kirman . The economy as an evolving network. *Journal of Evolutionary Economics*, 7:339–353, 1997.

H. W. Kuhn and A. W. Tucker . Nonlinear programming. In the *Proceedings of 2nd Berkeley Symposium*. Berkeley: University of California Press, pages 481–492, 1951.

H. J. Kushner and D. S. Clark . *Stochastic approximation methods for constrained and unconstrained systems*. Springer, New York, 1978.

H. J. Kushner and G. Yin . *Stochastic Approximation and Recursive Algorithms and Applications*. 2nd Edition, Springer-Verlag, New York, 2003, [Applications of Mathematics, Volume 35], xxii+474 pp. *Stochastic Approximation Algorithms and Applications*, 1st Edition, 1997. xxi+417 pp.

S. H. Low L. Chen , T. Cui and J. C. Doyle . A game-theoretic model for medium access control. *IEEE Journal on Selected Areas in Communications*, 26(7), September 2008.

S. Lasaulce and H. Tembine . *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*. Academic Press, Elsevier, ISBN: 978-0-12-384698-3, 336 pages, 2011.

M. Le Treust , H. Tembine , S. Lasaulce , and M. Debbah . Coverage games in small cells networks. *IEEE Proceedings of the Future Network and Mobile Summit (FUNEMS'10)*, 2010.

D. Leslie and E. Collins . Generalised weakened fictitious play. *Games and Economic Behavior*, 56:285–298, 2005.

D. Leslie and E. Collins . Individual q-learning in normal form games. *SIAM Journal Control Optim.*, 44:495–514, 2005.

D. S. Leslie and E. J. Collins . Convergent multiple timescales reinforcement learning algorithms in normal form games. *The Annals of Applied Probability*, 13(4):1231–1251, 2003.

Shu-Jun Liu and Miroslav Krstic . Stochastic nash equilibrium seeking for games with general non-linear payoffs. *SIAM Journal CONTROL OPTIMIZATION*, 49(4):1659–1679, 2011.

L. Ljung . Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22(4):551–575, 1977.

Chengnian Long , Z. Qian , Li Bo , Y. Huilong , and G. Xinping . Noncooperative power control for wireless ad hoc networks with repeated games. *IEEE JSAC*, 25(6):1101–1112, August 2007.

J. R. Marden , G. Arslan , and J. S. Shamma . Joint strategy fictitious play with inertia for potential games. In *Proceedings 44th IEEE Conf. Decision Control*, pages 6692–6697, Dec. 2005.

J. R. Marden , G. Arslan , and J. S. Shamma . Joint strategy fictitious play with inertia for potential games. *IEEE Transactions on Automatic Control*, 54(2), February 2009.

J. R. Marden , H. Peyton Young , G. Arslan , and J. S. Shamma . Payoff-based dynamics for multi-player weakly acyclic games. *SIAM Journal on Control and Optimization*, 2009.

L. W. McKenzie . On equilibrium in graham's model of world trade and other competitive systems. *Econometrica*, 22:147–161, 1954.

P. Mehta and S. Meyn . Q-learning and pontryagin's minimum principle. in *IEEE Proceedings CDC*, 2009.

I. Menache and N. Shimkin . Efficient rate-constrained nash equilibrium in collision channels with state information. *IEEE Journal on Selected Areas in Communications*, special issue on Game Theory, 26(7):1070–1077.

P. Mertikopoulos and A. L. Moustakas . Correlated anarchy in overlapping wireless networks. *IEEE Journal on Selected Areas in Communications*, 26(7):1160–1169, 2008.

Mihail Mihaylov , Karl Tuyls , and Ann Nowé . Decentralized learning in wireless sensor networks. In the *Proceedings of the Adaptive and Learning Agents Workshop (ALA 2009)*, Taylor and Tuyls (eds.), Budapest, Hungary.

K. Miyasawa . On the convergence of learning processes in a 2x2 non-zero-person game. *Princeton University Research Memo* 33, 1961.

D. Monderer and L. S. Shapley . Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

Anthony P. Morse . The behaviour of a function on its critical set. *Annals of Mathematics*, 40:62–70, January 1939.

J. Nachbar . Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, 19:59–89, 1990.

John Nash . Non-cooperative games. *The Annals of Mathematics*, 54:286–295, 1951.

John Von Neumann and Oskar Morgenstern . *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

P. Nicopolitidis , G. I. Papadimitriou , P. Sarigiannidis , Mohammad S. Obaidat , and A. S. Pomportsis . Adaptive wireless networks using learning automata. *IEEE Wireless Communications*, pages 75–81, 2011.

P. Nicopolitidis , G.I. Papadimitriou , and A.S. Pomportsis . Learning-automata-based polling protocols for wireless lans. 51(3):453–463, March 2003.

H. Nikaido and K. Isoda . Note on non-cooperative convex games. *Pacific Journal of Mathematics*, 5:807–815, 1955.

R. Pemantle . Nonconvergence to unstable points in urn models and stochastic approximations. *Annals of Probability*, 18:698–712, 1990.

S. M. Perlaza , V. Q. Florez , H. Tembine , and S. Lasaulce . On the convergence of fictitious play in channel selection games. *IEEE Latin America Transactions*, 9, April 2011.

S. M. Perlaza , H. Tembine , S. Lasaulce , and M. Debbah . Satisfaction equilibrium for quality-of-service provisioning in decentralized networks. *IEEE Journal on Selected Topics in Signal Processing*. Special Issue in Game Theory for Signal Processing, 2012.

S. M. Perlaza , H. Tembine , and S. Lasaulce . How can ignorant but patient cognitive terminals learn their strategy and utility? In *IEEE Proceedings of the 11th International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, 2010.

S. M. Perlaza , H. Tembine , S. Lasaulce , and M. Debbah . Satisfaction equilibrium: A general framework for qos provisioning in self-configuring networks. *IEEE Globecom*, 2010.

Samir Medina Perlaza , H. Tembine , Samson Lasaulce , and V. Quintero-Florez . On the fictitious play and channel selection games. In *Proceedings of the Latin-American Conference on Communications (LATIN-COM)*, Bogota, Colombia, September 2010.

J. Ponstein . Existence of equilibrium points in non-product spaces. *SIAM Journal on Applied Mathematics*, 14(1):181–190, Jan. 1966.

A. Poznyak and K. Najim . Learning through reinforcement for n-person repeated constraint games. *IEEE Trans. on Systems, Man, and Cybernetics: Part B Cybernetics*, 32:759–771, 2002.

Bary S. R. Pradelski and H. Peyton Young . Efficiency and equilibrium in trial and error learning. 2010.

H. Robbins . Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 55:527–535, 1952.

H. Robbins and S. Monro . A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

David P. Roberts . Kernel sizes for random matrix games. Working paper, 2006.

David P. Roberts . Nash equilibria in cauchy-random zero-sum and coordination matrix games. *International Journal of Game Theory*, 34:167184, 2006.

J. Robinson . An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951.

J. B. Rosen . Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33(3):520–534, 1965.

A. Roth and I. Erev . Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212, 1995.

W. H. Sandholm . *Population games and evolutionary dynamics*. MIT Press, 2010.

Arthur Sard . The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society*, 48:883–890, 1942.

K. P. R. Sastry and G. V. R. Babu . Approximation of fixed points of strictly pseudo-contractive mappings on arbitrary closed convex sets in a banach space. *Proceedings of Amer. Math. Soc.*, 128, (10):2907–2909, 2000.

L. Shapley . Some topics in two-person games. In *Advances in Game Theory* M. Drescher , L.S. Shapley , and A.W. Tucker (Eds.), Princeton: Princeton University Press, 1964.

L. S. Shapley . Stochastic games. *Proceedings Nat. Acad. Sciences*, 39:1095–1100, 1953.

Alonso Silva , Eitan Altman , Mérouane Debbah , H. Tembine , and Chloé Jimenez . Optimal mobile association on hybrid networks: centralized and decentralized case. *CoRR*, abs/0911.0257, 2009.

Alonso Silva , H. Tembine , Eitan Altman , and Mérouane Debbah . Spatial games and global optimization for the mobile association problem: the downlink case. 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, December 15–17 2010.

V. Srivastava , J. Neel , A. B. MacKenzie , R. Menon , L. A. DaSilva , J. E. Hicks , J. H. Reed , and R. P. Gilles . Using game theory to analyze wireless ad hoc networks. *IEEE Communications Surveys and Tutorials*, 7(4):46–56, 2005.

W. Stanford . On the number of pure strategy nash equilibria in finite common payoffs games. *Economics Letters*, 62:29–34, 1999.

G. Stoltz and G. Lugosi . Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59:187–208, April 2007.

D.W. Stroock and SRS Varadhan . *Multidimensional Diffusion Processes*. Springer, 1979.

P. D. Taylor and L. Jonker . Evolutionarily stable strategies and game dynamics. *Mathematical Bioscience*, 40:145–156, 1978.

Emre Telatar . Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–596, 1999.

H. Tembine . Hybrid mean field game dynamics in large populations. *American Control Conference, ACC*, San Francisco, California, US.

H. Tembine . Stochastic optimization: the non-differentiable case. Master thesis dissertation, INRIA, 2004.

H. Tembine . Folk theorem for stochastic games with public signals. Master dissertation, 2006.

H. Tembine . Repeated games with imperfect monitoring. Master Thesis Dissertation, Ecole Polytechnique, France, 2006.

H. Tembine . Evolutionary network formation games and fuzzy coalition in heterogeneous networks. 2nd IFIP Wireless Days (WD), 2009.

H. Tembine . Population games with networking applications. Ph.D Dissertation, University of Avignon, September 2009.

H. Tembine . Distributed strategic learning for wireless engineers. Notes, 2010.

H. Tembine . Population games in large-scale networks. LAP, 250 pages, ISBN 978-3-8383-6392-9, 2010.

H. Tembine . Codipas: Combined fully distributed payoff and strategy learning. Notes, January 2011.

H. Tembine . Dynamic robust games for mimo systems. *IEEE Transactions Systems, Man, Cybernetics*, 99(41):990–1002, August 2011.

H. Tembine . Mean field stochastic games. Notes, 2011.

H. Tembine . Mean field stochastic games: convergence, q/h learning, optimality. *American Control Conference, ACC*, San Francisco, California, US, 2011.

H. Tembine , E. Altman , R. El-Azouzi , and H. Hayel . Evolutionary games with random number of interacting players with application to access control. *IEEE/ACM Proceedings 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)* March 31 – April 4, Berlin, Germany, 2008.

H. Tembine , E. Altman , and R. ElAzouzi . Delayed evolutionary game dynamics applied to medium access control. In *Proceedings 4th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS)*, Pisa, Italy., 2007.

H Tembine , E. Altman , R. ElAzouzi , and Y. Hayel . Stable networking games. In *Proceedings of Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, Allerton Retreat Center, Monticello, Illinois, September 2008.

H. Tembine , E. Altman , R. ElAzouzi , and Y. Hayel . Stochastic population games with individual independent states and coupled constraints. *ACM Proceedings of the Second International Workshop on Game Theory in Communication Networks (GameComm)*, Athens, Greece, October 2008.

H. Tembine , E. Altman , R. ElAzouzi , and Y. Hayel . Correlated evolutionarily stable strategies in random medium access control. *International IEEE Conference on Game Theory for Networks, Gamenets*, 2009.

H. Tembine , E. Altman , R. ElAzouzi , and Y. Hayel . Evolutionary games in wireless networks. *IEEE Trans. on Systems, Man, and Cybernetics, Part B, Special Issue on Game Theory*, June 2010.

H. Tembine , E. Altman , R. ElAzouzi , and Y. Hayel . Bio-inspired delayed evolutionary game dynamics with networking applications. *Telecommunication Systems Journal*, DOI: 10.1007/s11235-010-9307-1., 47:137–152(16), 2011.

H. Tembine , E. Altman , R. ElAzouzi , and W. H. Sandholm . Evolutionary game dynamics with migration for hybrid power control in wireless communications. 47th SIAM/IEEE CDC, December 2008.

H. Tembine and M. Assaad . Hybrid mean field learning in large-scale dynamic robust games. *International Conference on Control and Optimization with Industrial Applications, COIA*, 2011.

H. Tembine and A. Kobbane . Robust power allocation games under channel uncertainty and time delays. *IFIP Wireless Days*, 2010.

H. Tembine , A. Kobbane , and M. El Koutbi . Dynamic robust power allocation games under channel uncertainty and time delays. *Elsevier Computer Communications*, 12, 34:1529–1537, 2011.

H. Tembine , J. Y. Le Boudec , R. El-Azouzi , and E. Altman . Mean field asymptotics of markov decision evolutionary games and teams. in Proceedings of IEEE Games, May 2009.

H. Tembine , P. Vilanova , and M. Debbah . Noisy mean field game model for malware propagation in opportunistic networks. 2nd International Conference on Game Theory for Networks, Games, 2011.

M. A. L. Thathachar and P. S. Sastry . Learning automata in stochastic games with incomplete information. Systems and Signal Processing, (R. N. Madan , N. Viswanadham and R. L. Kashyap , eds.), (New Delhi), pages 417–434, 1991.

M.A.L. Thathachar , P.S. Sastry , and V.V. Phansalkar . Decentralized learning of nash equilibria in multiperson stochastic games with incomplete information. IEEE transactions on system, man, and cybernetics, 24(5), 1994.

E.L. Thorndike . Animal intelligence: An experimental study of the associative processes in animals. Psychological Review, Monograph Supplements, vol. 8. MacMillan, New York, 1898.

J. Akbari Torkestani and M.R. Meybodi . Mobility-based multicast routing algorithm for wireless mobile ad-hoc networks: A learning automata approach. Computer Communications, Elsevier, 33(6):721–735, April 2010.

D. N. C. Tse and S. V. Hanly . Multiaccess fading channels. i. polymatroid structure, optimal resource allocation and throughput capacities. IEEE Trans. on Info. Theory, 44(7):2796–2815, 1998.

M. van der Schaar and F. Fu . Spectrum access games and strategic learning in cognitive radio networks for delay-critical applications. Proceedings IEEE, 2009.

S. Verdú and V. Poor . On minimax robustness: A general approach and applications. IEEE Transactions Information Theory, 30:328–340, Mar. 1984.

Y. Viossat . The replicator dynamics does not lead to correlated equilibria. Games and Economic Behavior, 59:397–407, 2007.

P. Viswanath , D.N.C. Tse , and V. Anantharam . Asymptotically optimal water-filling in vector multiple-access channels. IEEE Transactions on Information Theory, 47(1):241–267, Jan 2001.

J. Wardrop . Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers, Part II, 1(36):352–362, 1952.

C.I.C.H. Watkins . Learning from delayed rewards. PhD thesis, University of Cambridge, Cambridge, UK, 1989.

C.I.C.H. Watkins and P. Dayan . Q-learning. Machine Learning, 8:279–292, 1992.

J. Weibull . Evolutionary game theory. MIT Press, 1995.

Y. Xing and R. Chandramouli . Stochastic learning solution for distributed discrete power control game in wireless data networks. IEEE/ACM Transactions on Networking, 16:932–944, august 2008.

H. P. Young . The evolution of conventions. Econometrica, 61:57–84, 1993.

H. P. Young . Strategic Learning and Its Limits. Oxford University Press, 2004.

H. P. Young . Learning by trial and error. Games and Economic Behavior, Elsevier, 65(2):626–643, March 2009.

Q. Zhu , H. Tembine , and T. Başar . A constrained evolutionary gaussian multiple access channel game. in the IEEE proceedings of GameNets, 2009.

Q. Zhu , H. Tembine , and T. Başar . Evolutionary game for hybrid additive white gaussian noise multiple access control. in IEEE Proceedings of Globecom, 2009.

Q. Zhu , H. Tembine , and T. Başar . Heterogeneous learning in zero-sum stochastic games with incomplete information. in 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, 2010.

Q. Zhu , H. Tembine , and T. Başar . Network security configuration: A nonzero-sum stochastic game approach. in IEEE Proceedings of American Control Conference, June 2010.

Q. Zhu , H. Tembine , and T. Başar . Distributed strategic learning with application to network security. American Control Conference, ACC 2011, San Francisco, California, US, 2011.