CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

DEPARTAMENTO DE CONTROL AUTOMÁTICO

Soluciones para juegos cooperativos y no cooperativos con cadenas de Markov

TESIS

Que presenta

M. en C. Kristal Karina Trejo Palacios

Para obener el grado de

DOCTORA EN CIENCIAS

En la especialidad de

Control Automático

Directores de tesis:

Dr. Alexander Pozniak Gorbatch

Dr. Julio Bernardo Clempner Kerik

Ciudad de México Diciembre, 2018

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

DEPARTMENT OF AUTOMATIC CONTROL

Solving Cooperative and Non-cooperative Markov Games

M.Sc. Kristal Karina Trejo Palacios

Dissertation supervisors:

Dr. Alexander Pozniak Gorbatch

Dr. Julio Bernardo Clempner Kerik

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Science in Automatic Control

Mexico City December, 2018

A mis padres, Ofelia y Juan Felipe, por siempre
brindarme su amor y apoyo incondicional,
por ser el pilar fundamental en todo lo que soy.

Agradecimientos

A través de estas líneas deseo expresar mi más sincero agradecimiento a las personas e instituciones que directa o indirectamente participaron en el desarrollo de esta tesis doctoral.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Centro de Investigación y de Estudios Avanzados (CINVESTAV-IPN) por el apoyo y patrocinio recibido durante todo el proceso de formación doctoral y realización de este proyecto de tesis. También agradezco al proyecto CONACYT No. 251552 (Ciencias Básicas) por el soporte financiero brindado para mi desarrollo académico y científico.

Mi más profundo agradecimiento al Dr. Alexander Poznyak y al Dr. Julio Clempner, por la confianza que depositaron en mí para realizar este proyecto, por su excepcional dirección, por su tiempo dedicado, paciencia y todas sus valiosas recomendaciones, enseñanzas y experiencias a lo largo del desarrollo de esta investigación.

Al Dr. Ruben Juarez, un especial agradecimiento por haberme recibido en su grupo de investigación en la Universidad de Hawaii. Por sus consejos, apoyo y ánimo que me brindó durante mi estancia, donde tuve la oportunidad de aprender y crecer de manera acedémica y personal.

A todos mis profesores, por todas sus lecciones y orientación, y en especial a mis revisores de tesis y miembros del jurado de examen de grado, Dra. Sabine Mondié Cuzange, Dra. Petra Wiederhold Grauert, Dr. Jorge León Vázquez y Dr. Wen Yu, les agradezco su tiempo y apoyo.

A mis compañeros y amigos, personas importantes en mi vida que siempre estuvieron dispuestos a brindarme su ayuda cuando fue necesario. Gracias por los buenos y malos momentos, por hacer de todo esto una mejor experiencia. Un especial agradecimiento a mis grandes amigos Marco y Chairez, por estar siempre a mi lado, por los incontables buenos momentos y excelentes pláticas, por el respaldo y la amistad.

Agradezco a mis padres y hermanos, por su comprensión, cariño, apoyo, instrucción y aliento durante el doctorado y toda mi vida. Ellos me han dado todo lo que soy como persona. Gracias a toda mi familia, que a pesar de la distancia siempre estuvieron a mi lado en el proceso de alzanzar esta meta en mi vida.

De manera muy especial a Daniel, por todo, por el impulso y la confianza, por apoyarme y hacer valer los pequeños momentos, por brindarme las palabras correctas cuando las necesité y compartir mis alegrías y angustias. Las palabras nunca serán suficientes para testimoniar mi cariño y agradecimiento.

A todos ustedes, mi mayor reconocimiento y gratitud.

Contents

List of figures

RESUMEN

Esta tesis presenta modelos para establecer estrategias cooperativas y no cooperativas en diferentes problemas de teoría de juegos. De manera general, los jugadores pueden actuar de dos formas: jugando cooperativamente o no cooperativamente con respecto a los otros jugadores. También es importante considerar los juegos en donde los jugadores forman coaliciones, en este caso ellos pueden cooperar o no cooperar, o pueden hacer una combinación de estos comportamientos, los jugadores cooperan dentro de la coalición pero el juego entre coaliciones es no cooperativo.

El concepto de colaboración implica que los jugadores interactúan con los otros jugadores con el fin de alcanzar una estabilidad cooperativa. Esta noción requiere que los jugadores seleccionen estrategias óptimas, condicionando su propio comportamiento al comportamiento de los demás para alcanzar la mejor estrategia en el futuro. En teoría de juegos, la estabilidad colectiva es una caso especial del equilibrio de Nash llamado equilibrio *strong Nash*.

Este trabajo presenta un método para calcular el equilibrio *strong* L_p-*Nash*. Este problema se resuelve en términos de la norma L_p: los jugadores seleccionan una estrategia que minimice la distancia a un mínimo utópico o ideal en el espacio euclidiano, es decir, no existe otra estrategia que mejore el comportamiento de la función de costo. Esto significa que existe una solución óptima que es un punto *strong Pareto optimal* que corresponde al equilibrio *strong Nash*. Además, se presenta un método para calcular el equilibrio *strong Stackelberg/Nash*. Este juego tipo líder-seguidor involucra a $\mathfrak{n}$ líderes jugando de manera cooperativa y $\mathfrak{m}$ seguidores que también juegan cooperativamente entre ellos, por lo tanto es necesario hacer uso del concepto de equilibrio *strong* L_p-*Nash*, es decir, la existencia de un equilibrio L_p-Stackelberg/Nash es caracterizada bajo estrategias que son *strong Pareto*.

Hay un creciente interés en aplicar juegos tipo Stackelberg para modelar asignación de recursos para problemas de patrullaje (seguridad), en los cuales los defensores tienen recursos limitados y deben asignarlos para proteger diferentes objetivos de posibles atacantes. En el

mundo real los atacantes son agentes sofisticados que emplean estrategias dinámicas. Sin embargo, la mayoría de los enfoques que existen en la literatura para calcular las estrategias de los defensores consideran que los atacantes tienen un comportamiento fijo y debido a esto, no aseguran que se tenga éxito en la realización del juego.

Para abordar esta deficiencia, presentamos un método para adaptar las estrategias de los atacantes y las estrategias de patrullaje de los defensores que son aplicadas en juegos de seguridad de tipo Stackelberg empleando un enfoque de aprendizaje por refuerzo. Se propone un marco común que combina tres paradigmas diferentes: conocimiento previo, imitación y el método de diferencia temporal. La arquitectura general de aprendizaje por refuerzo incluye dos componentes principales: una arquitectura de aprendizaje adaptivo primario y la arquitectura de actor crítico. Este trabajo considera que los defensores y los atacantes forman coaliciones en el juego de seguridad Stackelberg calculando el equilibrio L_p-Stackelberg/Nash.

Otra clase importante de juegos que incluye soluciones cooperativas y no cooperativas es el problema de negociación. El juego de negociación se refiere a una situación en la cual los jugadores tienen la oportunidad de concluir un acuerdo de beneficio mutuo. Sin embargo, en este tipo de juegos existe conflicto de intereses sobre cual acuerdo pactar, considerando que no se puede imponer un acuerdo a ningún jugador sin su aprobación. Cabe destacar que el problema de negociación y sus soluciones ha sido aplicado en contextos importantes como acuerdos corporativos, arbitraje, juegos de mercado de duopolio, protocolos de negociación, etc. El presente trabajo examina los juegos de negociación desde una perspectiva teórica y proporciona un método de solución para diferentes modelos: los modelos de negociación cooperativa presentados por Nash y por Kalai y Smorodinsky, quienes proponen un enfoque axiomático para resolver el problema dependiendo de diferentes principios de imparcialidad; y el modelo para una negociación no cooperativa que presenta Rubinstein, quien propone un juego de negociación con ofertas alternadas y factores de descuento.

En la presente tesis se consideran juegos con cadenas de Markov en tiempo continuo y discreto. Diseñamos un método para juegos estáticos en términos de problemas de programación no lineal implementando el principio de Lagrange. Además, se utiliza el método de regularización de Tikhonov con el fin de asegurar la convergencia de las funciones de costo a un

punto de equilibrio. El problema de programación no lineal es formulado considerando varias restricciones lineales empleando el método c-*variable* con la finalidad de hacer el problema manejable computacionalmente. Para calcular el punto de equilibrio se emplea un enfoque de programación de dos niveles implementado por el método extraproximal, el cual consiste de un procedimiento iterativo de dos pasos, el primer paso es una predicción que calcula una aproximación preliminar del punto de equilibrio y el segundo paso tiene como finalidad realizar un ajuste de la predicción calculada previamente. Cada ecuación en este método es un problema de optimización para el cual se resuelve la condición necesaria de un mínimo utilizando el método de gradiente. El método extraproximal conduce a una realización computacional simple y lógicamente justificada: en cada iteración de ambos pasos del procedimiento, el funcional del juego disminuye y converge a un punto de equilibrio.

Los métodos propuestos para cada uno de los problemas de teoría de juegos mencionados anteriormente son validados de manera teórica. Además, algunos ejemplos ilustran los resultados principales así como la efectividad de los métodos.

ABSTRACT

This thesis presents a model to establish cooperative and non-cooperative strategies for solving different problems within game theory. In general, players proceed in two different ways: as in a cooperative game or, a non-cooperative game (selecting their strategies not cooperatively among them) with respect to the other players. In the case when players form separate coalitions they can cooperate or do not cooperate or make a combination (players into coalition play cooperatively but the game between coalitions is non-cooperative).

The notion of collaboration implies that related players interact with each other looking for cooperative stability. This notion consents players to select optimal strategies and to condition their own behavior on the behavior of others in a strategic forward-looking manner. In game theory, collective stability is a special case of the Nash equilibrium called strong Nash equilibrium.

This work presents a novel method for computing the strong L_p-Nash equilibrium. The problem is solved in terms of the L_p-norm: players choose a strategy that minimizes the distance to the utopian minimum in the Euclidean space, i.e., no other strategy produces a smaller total expected loss. This means that there exists an optimal solution that is a strong Pareto optimal point and it is the closest solution to the minimum utopia point. The strong Pareto optimal solution corresponds to the strong Nash equilibrium. Moreover, an approach for computing the strong Stackelberg/Nash equilibrium is presented. This leader-follower game implies that $\mathfrak{n}-$leaders play cooperatively and $\mathfrak{m}-$followers play also do cooperatively employing the strong L_p-Nash equilibrium concept, i.e., the existence of the L_p-Stackelberg/Nash equilibrium is characterized as a strong Pareto policy.

There is a growing interest in applying Stackelberg games to model resource allocation for patrolling security problems in which defenders must allocate limited security resources to protect targets from attack by adversaries. In real-world adversaries are sophisticated presenting dynamic strategies. Most existing approaches for computing defender strategies calculate the game against fixed behavioral models of adversaries, and cannot ensure success in the realization of the game.

To address this shortcoming, we present a novel approach for adapting attackers and defenders preferred patrolling strategies in Stackelberg security games using a reinforcement learning (RL) approach based on average rewards. We propose a common framework that combines three different paradigms: prior knowledge, imitation and temporal-difference method. The overall RL architecture involves two highest components: the adaptive primary learning architecture and the actor-critic architecture. This work considers that defenders and attackers conform coalitions in the Stackelberg security game, these are reached by computing the strong L_p-Stackelberg/Nash equilibrium.

Another important class of games that includes cooperative and non-cooperative solutions is the bargaining problem. The bargaining game refers to a situation in which players have the possibility of concluding a mutually beneficial agreement. Here there is a conflict of interests about which agreement to conclude, and no-agreement may be imposed on any player without that player's approval. Remarkably, bargaining and its game-theoretic solutions have been applied in many important contexts, like corporate deals, arbitration, duopoly market games, negotiation protocols, etc. Among all these research applications, equilibrium computation serves as a basis. This work examines bargaining games from a theoretical perspective and provides a solution method for different game-theoretic models: the cooperative bargaining models presented by Nash and Kalai-Smorodinsky which propose an elegant axiomatic approach to solve the problem depending on different principles of fairness, and the non-cooperative bargaining solution presented by Rubinstein which propose a bargaining game with alternating offers and a cost by time.

In this work, we consider games in case of a metric state space for a class of continuous and discrete time ergodic controllable Markov chains games. We design a method for the static game in terms of nonlinear programming problems implementing the Lagrange principle. In addition, we make use of the Tikhonov's regularization method to ensure the convergence of the cost functions to an equilibrium point. We formulate the nonlinear programming problem considering several linear constraints employing the $c-$variable method for making the problem computationally tractable. For computing the equilibrium point we employ a bi-level programming approach implemented by the extraproximal method, which consists of a two-step

iterated procedure where the first step is a prediction that calculates the preliminary position approximation to the equilibrium point and the second step is designed to find a basic adjustment of the previous prediction. Each equation in this solver is an optimization problem for which the necessary condition of a minimum is solved using the gradient projection method. The extraproximal method leads to a simple and logically justified computational realization: at each iteration of both steps of the procedure, the functional of the game decrease and converges to an equilibrium point.

The proposed methods for the game theory problems mentioned above are validated theoretically. In addition, some examples in game theory illustrate the main results and the effectiveness of the methods.

PUBLICATIONS

Journal:

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2015). A Stackelberg security game with random strategies based on the extraproximal theoretic approach. *Engineering Applications of Artificial Intelligence*, 37:145-153.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2015). Computing the Stackelberg/Nash equilibria using the extraproximal method: Convergence analysis and implementation details for Markov chains games. *International Journal of Applied Mathematics and Computer Science*, 25(2):337-351.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2016). An optimal strong equilibrium solution for cooperative multi-leader-follower Stackelberg Markov chains games. *Kybernetika*, 52(2):258-279.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2017). Computing the strong L_p-Nash equilibrium for Markov chains games: Convergence and uniqueness. *Applied Mathematical Modelling*, 41:399-418.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2018). Adapting attackers and defenders patrolling strategies: A reinforcement learning approach for Stackelberg security games. *Journal of Computer and System Sciences*, 95:35-54.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2018). Computing the bargaining approach for equalizing the ratios of maximal gains in continuous-time Markov chains games. *Computational Economics*, DOI: 10.1007/s10614-018-9859-9.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2018). Proximal constrained optimization approach with time penalization. *Engineering Optimization*, DOI: 10.1080/0305215X.2018.1519072.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (Submitted). Computing the Nash bargaining solution for multiple players in discrete-time Markov chains games.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (Submitted). Manipulating power in non-cooperative bargaining games.

- Kristal K. Trejo, Ruben Juarez, Julio B. Clempner and Alexander. S. Poznyak (Submitted). Non-cooperative bargaining for unsophisticated agents.

Conference:

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2015). Computing the L_p-strong Nash equilibrium looking for cooperative stability in multiple agents Markov games. In *12th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, 309-314.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2016). Adapting strategies to dynamic environments in controllable Stackelberg security games. In *55th IEEE Conference on Decision and Control (CDC)*, 5484-5489.

- Kristal K. Trejo, Julio B. Clempner and Alexander. S. Poznyak (2017). Nash bargaining equilibria for controllable Markov chains games. In *The 20th World Congress of the International Federation of Automatic Control (IFAC)*, 50(1):12261-12266.

Chapter Book:

- Kristal K. Trejo and Julio B. Clempner (2018). Setting Nash vs. Kalai-Smorodinsky bargaining approach: Computing the continuous-time controllable Markov game, 335-369. In *New Perspectives and Applications of Modern Control Theory*. Springer International Publishing.

Chapter 1

Introduction

Nash [61] established the framework to study bargaining where the players should cooperate when non-cooperation leads to Pareto-inefficient results. The bargaining game is based on a model in which players are assumed to negotiate on a set of feasible payoffs. A fundamental element of the game is the disagreement point (status quo) which plays a role of a deterrent. A bargaining solution is a single-valued function that selects an outcome from the feasible payoffs for each bargaining problem which is the result of cooperation by the players involved in the game. The agreement reached in the game is the most preferred alternative within the set of feasible outcomes.

Nash [61] proposed this approach by presenting four axioms and showing that they characterize the Nash bargaining solution. In the classical bargaining game theory models, a bargainer has a positive interest in the other's welfare as well as in his own. The agreement will represent a situation that could not be improved on to both players' advantage. Rational players would not accept a given agreement if some alternative arrangement could make both parties better off or at least one better off with the other no worse off. Then, the resulting bargaining strategy is an outcome which is Pareto optimally.

Game theory analyses of bargaining assume one of two approaches: a) the axiomatic, originates in the characterization of the Nash solution [61] (extended by Kalai and Smorodinsky [46]), where the desired properties of a solution are satisfied and b) the strategic, exemplified by Rubinstein's solution [83], where the bargaining procedure is modeled in detail as a sequential game, this approach is also called the non-cooperative bargaining solution. When players

are patient, the equilibrium agreement of the Rubinstein's game approximates the agreement given by the Nash's axiomatic approach. In the bargaining problem, the players have a mutual interest in reaching an agreement, although in general there is a conflict of interest over the particular agreement to be reached.

However, Nash [62] then changed to the question of how the dynamics and the rationality proposed for this solution correspond with many real-world situations given the constraint that players are concerned only with maximizing their own welfare. As a result, Nash proposed a non-cooperative game in which the only equilibrium outcome is exactly the allocation suggested by the Nash solution.

1.1 The Stackelberg/Nash game

The Nash equilibrium [62], players always make a best-reply to what other players are doing, is a fundamental concept in game theory and the most widely used method of predicting the outcome of a strategic interaction of several decision makers in non-cooperative games. It describes a mathematical model in which all players simultaneously compete against each other in a game. It is concerned with a strategy profile such that no player can unilaterally change her/his strategy to increase her/his payoff. However, non-cooperative equilibrium has individually stability and the collective stability is a special case of the Nash equilibrium called strong Nash equilibrium (SNE).

The SNE was introduced by Aumann [7] for cooperative games. A SNE is a Nash equilibrium for which no coalition of players has a joint deviation that improves the payoff of each member of the coalition [7]. In cooperative games the players can find a strategy producing the smaller total expected loss, such a cooperative strategy leads to strong Pareto optimal solution of the game.

There are several proposals reported in the literature to search strong Nash equilibria for specific classes of games, however, these proposals and algorithms fail in establishing a proper formulation regarding existence, recognition, and computation for the Pareto optimality. Most of them find a Nash equilibrium and then verify the Pareto optimality.

On the other hand, the leader-follower solution in game theory was introduced by von Stackelberg [89], as an extension of the Cournot duopoly model [22], suggesting a firm with the power to commit to a number of production profits from a leadership position. The leader-follower game theory has been studied in depth in oligopoly theory [15, 58].

Stackelberg games are usually represented by a leader-follower problem which corresponds to a bi-level programming problem. In bi-level programming problems there are two competing decision-making parties [10]: a) one is upper-level decision makers and, b) the other is lower level decision makers. The two levels interact with each other as follows. The lower level is completely restricted by the upper level's decision and for each decision made by the upper level, the lower level will choose the best option according to their objectives. Instead, the upper-level objectives are restricted from below by the lower level: the upper level controls the lower level's decision in the way that lower level will react by choosing the best option.

Game-theoretic approaches have been used in multiple deployed applications. An important example is security games between a defender and an attacker: first the defender considers what the target (best-reply) of the attacker is; then, holding the attacked target fixed, the defender picks a quantity that minimizes its payoff; finally the attacker actually observes this and in equilibrium picks the expected quantity that maximizes its payoff as a response. Some applications [72, 43, 106, 73, 2] use the (two-players) leader-follower Stackelberg game-theoretic formulation for solving the security problem, providing a randomized strategy for the defender (leader) and the attacker (follower).

We describe a Stackelberg security game as follows. Let us consider a game with $\mathfrak{n} + \mathfrak{m}$ players. Let $\mathcal{N} = \{1, \ldots, \mathfrak{n}\}$ denote the set of players called defenders and let their strategy set be defined by U. The set $\mathcal{M} = \{1, \ldots, \mathfrak{m}\}$ of players are called attackers and, similarly, let the set of their strategy profiles be defined by V. Then, $U \times V$ is the set of full strategy profiles. The dynamics of the game is as follows: the defenders choose a strategy $u \in U$ considering the cost-function $\varphi(u|v)$ for a fixed strategy v of the attackers, the attackers are informed about the strategy u selected by the defenders and choose their strategies v considering the cost-function $\psi(v|u)$ for a fixed strategy u of the defenders. We understand $\psi(v|u)$ as the response of the attackers to the strategy u of the defenders, which is the best-reply in the original game. In the

security game framework, we suppose that defenders commit to a randomized strategy while attackers choose their best-reply to this strategy. The solution of the game is a Stackelberg equilibrium point.

There exists a growing interest in applying Stackelberg games to model resource allocation for patrolling security problems in which defenders must allocate limited security resources to protect targets from attack by adversaries. In real-world, adversaries are sophisticated presenting dynamic strategies. Most existing approaches for computing defender strategies calculate the game against fixed behavioral models of adversaries, and cannot ensure success in the realization of the game. In the original Stackelberg security games formulation on Markov chains, we usually assume fixed and static domains models not able to be adapted to the environment: fixing a state and an action, the cost/reward and transitions always remain the same. The reason is that the main goal is minimizing/maximizing the players' expected cost/reward that depends on the transitions at each state. However, it is an unrealistic assumption: the transitions matrices and the reward received for Stackelberg security games are commonly non-static. Producing always the same resulting behavior can be exploited by intelligent attackers that carry out surveillance before an attack, it is often desirable for the security agencies to have a system in which randomness is involved in allocating their resources. To address this shortcoming, we will consider the learning properties of the attackers and defenders interaction, and we will deal with the adaptation (estimation and assessment) of the payoff and strategies to dynamic environments based on the information available to them.

Reinforcement learning (RL) is a problem faced by an agent or multiple agents that must learn behavior through trial-and-error interactions with a dynamic environment [44, 87]. It does not assume the existence of a teacher that provides examples upon which learning of a task takes place [78]. Computationally, RL is intended to operate in a learning environment composed of two subjects: the learner and a dynamic process. At successive time steps, the learner makes an observation of the process state, selects an action and applies it back to the process. Its goal is to find out an action policy that controls the behavior of the dynamic process, guided by signals that indicate how badly or well it has been performing the required

task. These signals are usually associated with a dramatic condition, a reward or a punishment, and the learner tries to optimize its behavior [78].

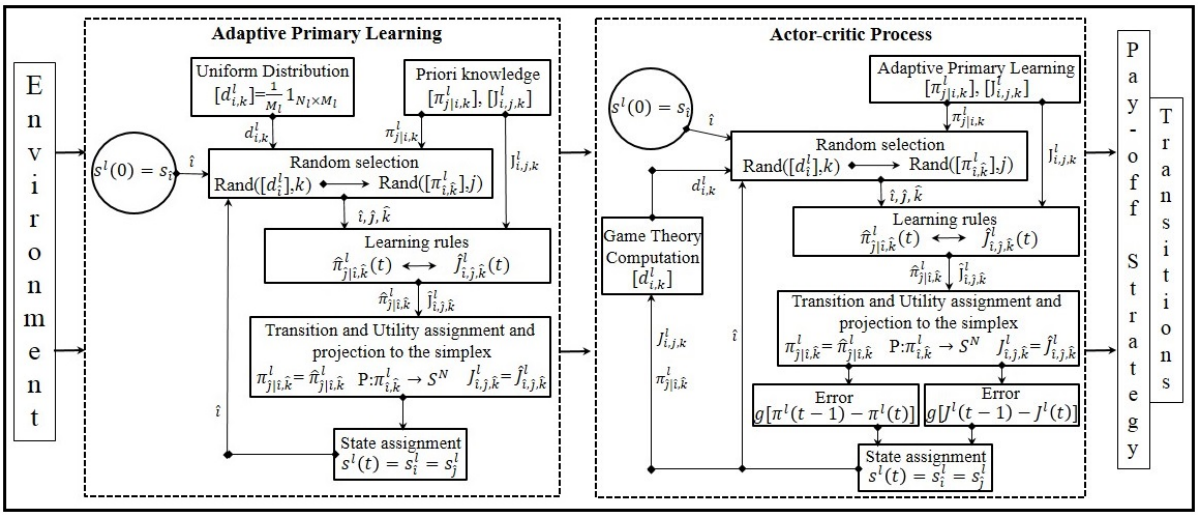


Figure 1.1 Reinforcement architecture.

Markov chains are a typical tool in the modeling of stochastic processes, specifically in the area of reinforcement learning, the environment is generally formulated as a Markov decision process. Reinforcement learning algorithms are strongly related to problems where a balance between the exploration of an unknown environment and the exploitation of previous knowledge and knowledge obtained during the exploration process is required. Reinforcement learning is especially appropriate for problems that include long-term vs. short-term reasoning, it is applicable in problems of game theory where there are scenarios with intelligent behaviors, situations where it is necessary to learn to decide what action to follow in a specific situation within a changing environment to achieve its goal.

1.2 The bargaining game

The bargaining model has attracted the attention of researchers from different disciplines and it is still, a relevant topic which is receiving a growing amount of attention for practitioners and academics in game theory. It has been applied in many important contexts including arbitration, supply chain contracts, duopoly market games, negotiation protocols, etc. It is

related to negotiation and group decision processes and introduces a solution concept for cooperative games. Cooperation concerns to coalitions of two or more players acting together with a specific common purpose taking into account the objective of maximizing their own individual payoffs. The bargaining game dynamics refers to a situation in which players have the possibility of concluding a mutually beneficial agreement. Here there is a conflict of interests about which agreement to conclude, and no-agreement may be imposed on any player without that player's approval. There are two theoretical perspectives that provide a solution for the cooperative game-theoretic bargaining models that employ the axiomatic method to evaluate bargaining: Nash [61] and Kalai-Smorodinsky [46]. It is important to note that the two bargaining solution approaches have the same feasible payoff set and disagreement point are considered to be the same bargaining problem in Nash's model.

The bargaining model was first presented as a game in John Nash's seminal 1950 paper [61], using the framework of game theory proposed by von Neumann and Morgenstern [65]. The von Neumann and Morgenstern theory supposes that when players form a coalition, they expect that a complementary coalition responds by damaging them in the worst way. This statement finds disapprovements in the literature. In this sense, Nash improved von Neumann and Morgenstern's work extending the idea by proposing axioms that characterize a unique result and a solution to the problem called the Nash bargaining solution. The formal description consists of two main components: a feasible set of utility allocations reached via cooperation, and the disagreement point occurring when players do not cooperate. A solution is a function that selects a feasible utility allocation for every problem. It is interesting to note that bargaining is one of the first situations of conflict of interest presented in the literature of game theory [45, 66].

The Kalai-Smorodinsky [46] bargaining solution differs from the Nash approach [61]. The fundamental difference between the two approaches resides in the fact that the Nash solution complies with the independence of irrelevant alternatives instead of the Kalai-Smorodinsky's solution fits monotonicity. Kalai and Smorodinsky argue that the entire set of alternatives must affect the agreement reached.

The most basic definition of bargaining refers to a socio-economic class of problems involving several players who can cooperate of not in terms of obtaining a better position of a desirable surplus whose distribution is in conflict. The features of the cooperation or non-cooperation of the players in terms of reaching an agreement and the initial situations of the players in the status-quo before an agreement has effect will determine how the surplus will be distributed. Several social, political and economic problems are related to the bargaining problem.

For instance, consider the case of selling a used car. When it comes to selling the car, the seller naturally wants to obtain the most money possible. It is practical to trade the car at a dealer or make a quick sale to a used car dealership, but these options usually leave the seller with significantly less than what the car is actually worth. Selling a car by himself allows the seller to get its full value. Then, the seller values his car at 3,000 which is the minimum price at which he would sell it. On the other hand, there is a buyer that values the car at 5,000 which is the maximum price at which he would buy it. If the trade occurs, the price lies between 3,000 and 5,000, then both the seller and the buyer would become better-off and a conflict of interests arises. In any trade, the seller and the buyer have the possibility of achieving a mutually beneficial agreement, or they can reach a non-cooperative agreement, by having conflicting interests over the terms of the trade.

1.2.1 Cooperative bargaining models

Following Nash [61], a solution to the bargaining problems $\mathcal{B}$ is a function f that takes as input any bargaining problem and returns a vector of utilities that belongs to the set of possible agreements Ψ. Several solutions can be proposed for solving the problem, but some of them can present inconsistencies. For example, one solution can go against symmetry by proposing a total improvement of the position of one player obtaining a point in the Pareto frontier of the utility and the other player receives no improvement. A different solution to the problem could be a disagreement point. The first solution violates symmetry, so the solution is unfair, and the second solution is not Pareto-efficient and does not take advantage of the cooperation related to an agreement situation. For solving the inconsistencies in the solution of the problem, Nash

[61] proposed several axioms: a) *Invariant to affine transformations* (or Invariant to equivalent utility representations): an affine transformation of the utility and disagreement point should not alter the outcome of the bargaining process; b) *Pareto optimality*: the solution selects a point of the Pareto frontier such that the players can be made better off without making other players worse off; c) *Symmetry*: if the players are indistinguishable, the solution should not discriminate between them; and d) *Independence of irrelevant alternatives*: if the solution is chosen from a feasible set which is an element of a subset of the original set but containing the point selected earlier by the solution, then the solution must still assign the same point chosen from the subset. As a result, Nash [61] proposed the Nash bargaining solution: we say that there is a unique solution to the bargaining problem that satisfies the four axioms (a to d) which is given by the point that maximizes the product of utilities of the players.

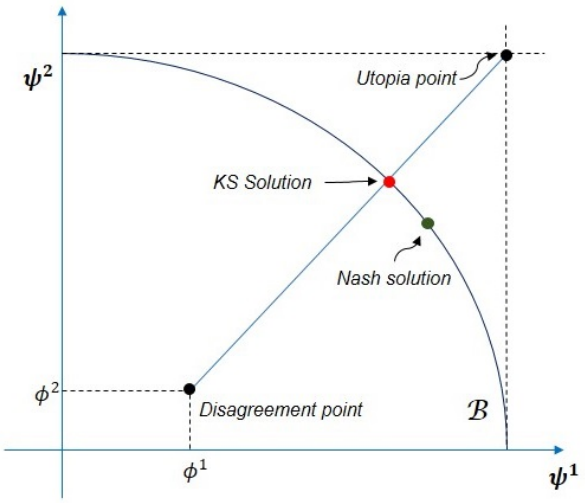


Figure 1.2 Cooperative bargaining models.

While three of Nash's axioms are quite uncontroversial, the fourth one (*Independence of irrelevant alternatives*) raised some criticism, which leads to a different line of research. Kalai and Smorodinsky [46] looked for characterizations of an alternative solution which do not use the controversial axiom. The solution idea can be represented geometrically in the following way. Let $\psi^*(\Psi)$ be the utopia point, typically not feasible, which gives the maximum payoff. Now, connect the point of disagreement ϕ and that ideal point $\psi^*(\Psi)$ by a line segment. The Kalai-Smorodinsky solution is the maximal point in Ψ on that line segment. They replaced

Nash arguable fourth axiom by e) *Monotonicity axiom*: if the set of possible agreements Ψ is enlarged such that the maximum utilities of the players remain unchanged, then neither of the players must suffer from it. Then, Kalai and Smorodinsky [46] proposed the following solution: we say that there is a unique solution ψ to the bargaining problem that satisfies the four axioms (a, b, c, and e) which is given by the intersection point of the Pareto frontier and the straight line segment connecting ϕ and the utopia point $\psi^*(\Psi)$. Figure 1.2 shows the cooperative solutions.

Nash [61] showed that there exists a unique standard independent solution for the bargaining model, while Kalai and Smorodinsky [46] showed that a different solution is the unique standard monotonic one.

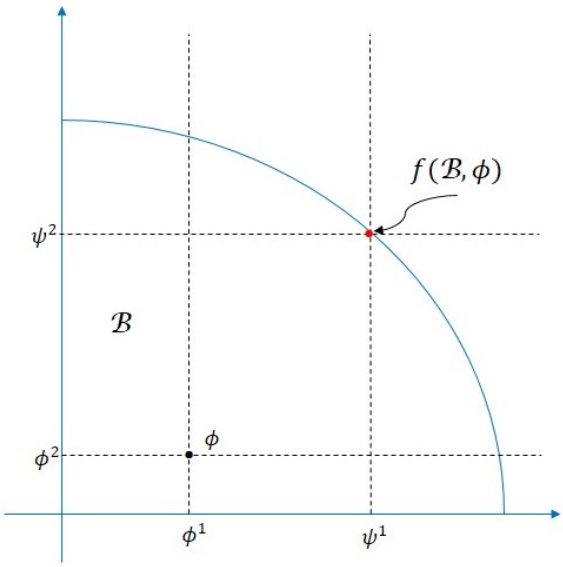


Figure 1.3 Bargaining model.

Consider two players $l = 1, 2$. A *bargaining* problem is a pair $\mathcal{B} = (\Psi, \phi)$ in the utility space were Ψ is a set of possible agreements in terms of utilities ψ that player 1 and player 2 can yield. The player's utility function ψ^l is strictly increasing and concave. The set of possible agreements is Ψ, which is a compact and convex set of $\mathbb{R}^2$. An element of Ψ is a pair $\psi = (\psi^1, \psi^2) \in \Psi$ and $\phi = (\phi^1, \phi^2)$ is called the disagreement utility point. Compactness arises from the assumptions related to closed production sets and bounded factor endowments.

Convexity is obtained from the fact that expected utility over outcomes. Also, the set Ψ involves points that dominate the disagreement point, i.e., there is a positive surplus to be enjoyed if agreement is reached. The function f takes as input any bargaining problem and returns a pair of utilities $\psi = (\psi^1, \psi^2) \in \Psi$. When we need to refer to the components of f, we write $\psi^1 = f^1(\mathcal{B})$ and $\psi^2 = f^2(\mathcal{B})$. The interpretation is that given a bargaining problem $\mathcal{B} = (\Psi, \phi)$ there exists an agreement $\psi = f(\Psi, \phi) \in \Psi$ such that $\psi^1 \geq \phi^1$ and $\psi^2 \geq \phi^2$ which ensures that there exists a mutually beneficial agreement. Figure 1.3 shows the bargaining problem.

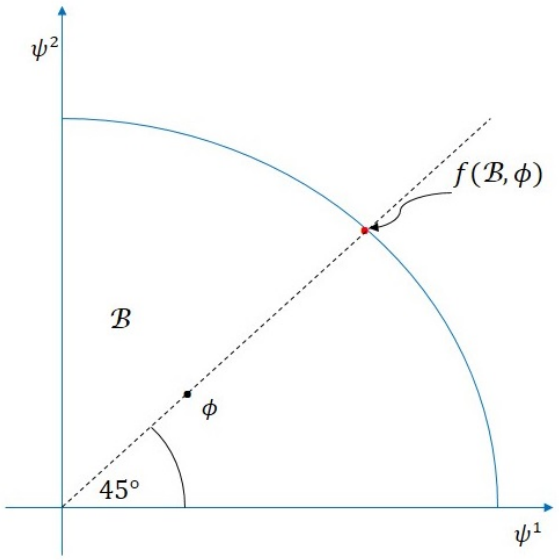


Figure 1.4 Bargaining axioms.

Two fundamental axioms impose the most important restrictions over the solution of the bargaining problem (see Figure 1.4). Pareto optimality: the function $f(\Psi, \phi)$ has the property that there does not exist a point $\psi = (\psi^1, \psi^2) \in \Psi$ such that $\psi^1 \geq f^1(\Psi, \phi)$ and $\psi^2 \geq f^2(\Psi, \phi)$ such that $(\psi^1, \psi^2) \neq f(\Psi, \phi)$. Symmetry: suppose that $\mathcal{B}$ is such that Ψ is symmetric around the 45° line and $\phi^1 = \phi^2$, then $f^1(\mathcal{B}) = f^2(\mathcal{B})$. The rest of the axioms will be presented in the formalization of the model.

1.2.2 The non-cooperative bargaining game

The Nash model does not allow for delayed agreements. In real situations, when the rules of the bargaining process are flexible, involving the facts that the time of starting negotiations and the moment of reaching agreement may be strategic variables, the Nash solution may be inappropriate.

There has been a large and growing literature in non-cooperative bargaining. Rubinstein [83] presented a bilateral non-cooperative bargaining process as an alternating offers game with a penalty according to the time taken by players in the decision making process, where it is proved that when every player bears fixed bargaining cost for each period, in this case, each player has a fixed discounting factor, the agreed contract is individual-rational and is Pareto optimal, i.e. it is no worse than disagreement, and there is no agreement which both would prefer. Such a model has been studied and extended for three or more players in a variety of papers and situations. The non-cooperative bargaining model and its game-theoretic solution have also been applied in many important contexts like market games, networks, apex games, union formation, and water management.

Consider a bargaining situation defined for two players who have to reach an agreement on the partition of a good. Each player takes turns to make an offer to the other agent on how it should be divided between them. After player 1 has made such an offer $(x, 1-x)$, player 2 must decide whether to accept it, in this case the bargaining game ends and the players divide the good according to the accepted offer, or to reject it and continue with the bargaining process. If player 2 rejects, then this player has to make a counteroffer $(y, 1-y)$ which player 1 would accept it, in this case the players divide the good according to the accepted offer but also considering a discount factor β associated to each player, or reject it and continue with the negotiation process. The bargaining game continues until an offer is accepted. Figure 1.5 shows the non-cooperative model.

Despite its wide applicability, crucial assumptions of the traditional Rubinstein bargaining model include that players have complete information about the characteristics of other agents (e.g., their discount factor or their utility) and that players are sophisticated in their behavior

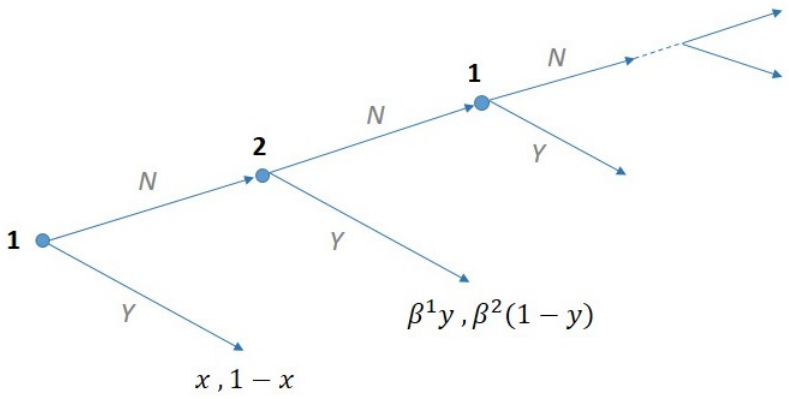


Figure 1.5 Non-cooperative bargaining model.

(e.g., they are forward-looking). As such, there is a need to develop a general theory of bargaining that is robust to work in the absence of sophisticated players or incomplete information about other players.

Then, we introduce an alternative approach to the traditional bargaining literature that aids unsophisticated players to reach the equilibrium as if they were forward-looking agents. The key element of the game is that players are penalized for their deviation from the previous best response strategy as well as their time taken for the decision-making at each step of the game.

1.3 Summary of the following chapters

The remainder of this thesis is organized as follows. Next Chapter presents the necessary notions and definitions related to continuous and discrete time Markov chains games to understand the rest of the work. The **Part I** is related to strong Stackelberg/Nash games. Chapter 3 establishes the definitions of the strong L_p-Nash equilibrium. We first present a general solution for the L_p-norm for computing the strong L_p-Nash equilibrium. Then, we suggest an explicit solution for the norms L_1, L_2 and L_∞. Chapter 4 describes and presents the solution method for computing the strong L_p-Stackelberg/Nash equilibrium. In Chapter 5 we suggest an approach for adapting attackers and defenders patrolling strategies in Stackelberg security

games. This chapter considers that defenders and attackers conform coalitions in the security game, these are reached by computing the strong L_p-Stackelberg/Nash equilibrium.

The **Part II** is related to the bargaining game. Chapter 6 presents a method for computing the Nash bargaining solution. Chapter 7 presents an approach for solving a bargaining problem employing a new equilibrium point for the game theory called the manipulation equilibrium point, this formulation employs the Nash bargaining and Stackelberg concepts. Chapter 8 presents a method to compute the Kalai-Smorodinsky bargaining solution. In Chapter 9 we suggest a novel method to find the equilibrium point in non-cooperative bargaining games.

All these chapters present numerical examples that validate the application of the method. Finally, Chapter 10 presents some final remarks.

Chapter 2

Mathematical background

This chapter presents some basic concepts and results about Markov chains games, needed to understand the rest of this work. For more information about these topics, please see [75, 76]. As usual, $\mathbb{R}$ and $\mathbb{N}$ stand for the sets of real numbers and non-negative integers, respectively.

2.1 Random sequences

2.1.1 Random variables

Let $\Omega = \{\omega\}$ be a set of elementary events ω which represents the occurrence or non-occurrence of a phenomenon.

Definition 2.1 *The system $\mathcal{F}$ of subsets of Ω is said to be the σ-algebra associated with Ω, if the following properties are fulfilled:*

1. $\Omega \in \mathcal{F}$
2. *for any sets* $A(n) \in \mathcal{F}$ $(n = 1, 2, ...)$

$$\bigcup_{n=1}^{\infty} A(n) \in \mathcal{F}, \quad \bigcap_{n=1}^{\infty} A(n) \in \mathcal{F};$$

3. *for any set* $A \in \mathcal{F}$

$$\overline{A} := \{\omega \in \Omega \mid \omega \notin A\} \in \mathcal{F}.$$

Definition 2.2 *The pair $(\Omega, \mathcal{F})$ represents the measurable space.*

Definition 2.3 *The function $P = P(A)$ of sets $A \in \mathcal{F}$ is called probability measure on $(\Omega, \mathcal{F})$ if it satisfies the following conditions:*

1. *for any $A \in \mathcal{F}$*
$$P(A) \in [0, 1];$$
2. *for any sequence $\{A(n)\}$ of sets $A(n) \in \mathcal{F}$ $(n = 1, 2, ...)$ such that*
$$A(n) \bigcap_{n \neq m} A(m) = \emptyset,$$
we have
$$P\left(\bigcup_{n=1}^{\infty} A(n)\right) = \sum_{n=1}^{\infty} P\left(A(n)\right).$$

The number $P(A)$ is called the probability of the event A. From a practical point of view, this probability is concerned with the occurrence of events.

Definition 2.4 *The triple $(\Omega, \mathcal{F}, P)$ is said to be the probability space.*

Definition 2.5 *A real function $\xi = \xi_\omega$, $\omega \in \Omega$ is called random variable defined on the probability space $(\Omega, \mathcal{F}, P)$, if it is $\mathcal{F}$-measurable, i.e., for any $s \in (-\infty, \infty)$*

$$\{\omega \mid \xi_\omega \leq s\} \in \mathcal{F}.$$

We say that two random variables $\xi_\omega(1)$ and $\xi_\omega(2)$ are equal with probability one (or, almost surely) if

$$P\left\{\omega \mid \xi_\omega(1) = \xi_\omega(2)\right\} = 1.$$

This fact can be expressed mathematically as follows

$$\xi_\omega(1) \stackrel{a.s.}{=} \xi_\omega(2).$$

Definition 2.6 *Let $\xi(1), \xi(2), ..., \xi(n)$ be random variables defined on $(\Omega, \mathcal{F}, P)$. The minimal σ-algebra $\mathcal{F}(n)$ which for any $s = (s(1), ..., s(n))^T \in \mathbb{R}^n$ contains the events*

$$\left\{\omega \mid \xi_\omega(1) \leq s(1), ..., \xi_\omega(n) \leq s(n)\right\},$$

is said to be the σ-algebra associated to the random variables $\xi(1), \xi(2), ..., \xi(n)$. It is denoted by

$$\mathcal{F}(n) = \sigma\left(\xi(1), \xi(2), ..., \xi(n)\right).$$

Definition 2.7 *The Lebesgue integral*

$$E\{\xi\} := \int\limits_{\omega\in\Omega} \xi_\omega P\{d\omega\},$$

is said to be the mathematical expectation of a random variable $\xi(\omega)$ given on $(\Omega, \mathcal{F}, P)$

Definition 2.8 *The random variable $E\{\xi \mid \mathcal{F}(0)\}$ is called the conditional mathematical expectation of the random variable $\xi(\omega)$ given on $(\Omega, \mathcal{F}, P)$ with respect to the σ-algebra $\mathcal{F}(0) \subseteq \mathcal{F}$ if*

1. *it is $\mathcal{F}(0)$-measurable, i.e.,*

$$\{\omega \mid E\{\xi \mid \mathcal{F}(0)\} \leq s\} \in \mathcal{F}(0) \; \forall \; s \in \mathbb{R}^1;$$

2. *for any set $A \in \mathcal{F}(0)$*

$$\int\limits_{\omega\in A} E\left\{\xi \mid \mathcal{F}(0)\right\} P\{d\omega\} = \int\limits_{\omega\in A} \xi(\omega) P\{d\omega\}.$$

Let $\xi = \xi(\omega)$ and $\theta = \theta(\omega)$ be two random variables given on $(\Omega, \mathcal{F}, P)$, θ an $\mathcal{F}(0)$-measurable ($\mathcal{F}(0) \subseteq \mathcal{F}$), then

1. $E\left\{\theta \mid \mathcal{F}(0)\right\} \stackrel{a.s.}{=} \theta$;

2. $E\{\theta\, \xi \mid \mathcal{F}(0)\} \stackrel{a.s.}{=} \theta E\{\xi \mid \mathcal{F}(0)\}$;

3. $E\{E\{\xi \mid \mathcal{F}(1)\} \mid \mathcal{F}(0)\} \stackrel{a.s.}{=} E\{\xi \mid \mathcal{F}(0)\}$ if $\mathcal{F}(0) \subseteq \mathcal{F}(1) \subseteq \mathcal{F}$.

Notice that if ξ is selected to be equal to the characteristic function of the event $A \in \mathcal{F}$, i.e.,

$$\xi(\omega) = \chi(\omega, A) := \begin{cases} 1 & \text{if the event } A \text{ has been realized} \\ 0 & \text{if not} \end{cases}$$

from the last definition we can define the conditional probability of this event under fixed $\mathcal{F}(0)$ as follows

$$P\{A \mid \mathcal{F}(0)\} := E\{\chi(\omega, A) \mid \mathcal{F}(0)\}.$$

2.1.2 Markov sequences and chains

Definition 2.9 *Any sequence* $\{s(n)\}$ *of random variables* $s(n) = s_\omega(n)$ $(n = 1, 2, ...)$ *given on* $(\Omega, \mathcal{F}, P)$ *and taking value in a set* S *is said to be a Markov sequence if for any set* $A \in B(S)$ *and for any time* n *the following property (Markov property) holds:*

$$P\{s(n+1) \in A \mid \sigma(s(n)) \wedge \mathcal{F}(n-1)\} \stackrel{a.s.}{=} P\{s(n+1) \in A \mid \sigma(s(n))\},$$

where $\sigma(s(n))$ *is the* σ*-algebra generated by* $s(n)$*,* $\mathcal{F}(n-1) = \sigma(s(1), ..., s(n-1))$ *and* $\sigma(s(n)) \wedge \mathcal{F}(n-1)$ *is the* σ*-algebra constructed from all events belonging to* $\sigma(s(n))$ *and* $\mathcal{F}(n-1)$*.*

This property means that any distribution on the future depends only on the value $s(n)$ realized at time n and is independent on the post values $s(1), ..., s(n-1)$; in other words, the present state of the system determines the probability for one step into the future.

Definition 2.10 *If the set* S*, defining any possible values of the random variables* $s(n)$*, is countable then the Markov sequence* $\{s(n)\}$ *is called a Markov chain. If, in addition, this set contains only finite number* N *of elements, i.e.,*

$$S = \{s_{(1)}, ..., s_{(N)}\},$$

then this Markov sequence is said to be a finite Markov chain.

2.2 Finite Markov chains

2.2.1 State space

Let $S = \left\{s_{(1)}, ..., s_{(N)}\right\}$ be a finite set of states. A state $s_{(i)} \in S$ is said to be

1. a non-return state if there exists a transition from this state to another one $s_{(j)} \in S$ but there is no way to return back to $s_{(i)}$;

2. an accessible (reachable) state from a state $s_{(j)} \in S$ if there exists a finite number n such that the probability for the random state $s(n)$ of a given finite Markov chain to be in the

state $s_{(i)} \in S$ starting from the state $s(1) = s_{(j)} \in S$ is more than zero, i.e.,

$$P\left\{s(n) = s_{(i)} \mid s(1) = s_{(j)}\right\} \overset{a.s.}{>} 0.$$

We will denote this fact as follows

$$s_{(j)} \Rightarrow s_{(i)}.$$

Otherwise we say that the considered state is inaccessible from the state $s_{(j)}$.

Definition 2.11 *Two states $s_{(j)}$ and $s_{(i)}$ are said to be a communicating states if each of them is accessible from the other one. We will denote this fact by*

$$s_{(j)} \Leftrightarrow s_{(i)}.$$

The class $S_{(i)}$ is said to be the j^{th} communicating class of states if it includes all communicating states of a given finite Markov, i.e., it includes all states such that

$$s_{(i)} \Leftrightarrow s_{(j)} \Leftrightarrow \cdots \Leftrightarrow s_{(m)} \Leftrightarrow s_{(k)}.$$

Definition 2.12 *A state $s_{(i)}$ is called recurrent if, when starting there, it will be visited infinitely often with probability one; otherwise the state is said to be transient.*

Definition 2.13 *A state $s_{(i)}$ is said to be an absorbing state if the probability to remain in state $s_{(i)}$ is positive, and the probability to move from any state $s_{(j)}$, $j \neq i$, to the state $s_{(i)}$ is equal to zero.*

2.2.2 Transition matrix

Definition 2.14 *Let $\Pi(n) \in \mathbb{R}^{N \times N}$ is said to be the transition matrix at time n of a given Markov chains with finite number N of states if it has the form*

$$\Pi(n) = \begin{bmatrix} \pi_{(1,1)}(n) & \pi_{(1,2)}(n) & \cdot & \cdot & \pi_{(1,N)}(n) \\ \pi_{(2,1)}(n) & \pi_{(2,2)}(n) & \cdot & \cdot & \pi_{(2,N)}(n) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \pi_{(N,1)}(n) & \pi_{(N,2)}(n) & \cdot & \cdot & \pi_{(N,N)}(n) \end{bmatrix}$$

where each element $\pi_{(i,j)}(n)$ represents the probability (one-step transition probability) for this finite Markov chain to go from the state $s(n) = s_{(i)}$ to the next state $s(n+1) = s_{(j)}$, i.e.,

$$\pi_{(j|i)}(n) := P\left\{s(n+1) = s_{(j)} \mid s(n) = s_{(i)}\right\}, \qquad i, j = 1, ..., N.$$

Each element $\pi_{(j|i)}(n)$ of the transition matrix $\Pi(n)$ is a probability of the corresponding event, then we conclude that

$$\pi_{(j|i)}(n) \in [0, 1], \qquad \sum_{j=1}^{N} \pi_{(j|i)}(n) = 1. \tag{2.1}$$

Definition 2.15 *Any matrix $\Pi(n) \in \mathbb{R}^{N\times N}$ with elements $\pi_{(j|i)}(n)$ satisfying the condition (2.1) is said to be a stochastic matrix.*

Any transition matrix of a finite Markov chains is a stochastic matrix. From condition (2.1), a stochastic matrix has the following properties:

1. the norm of a stochastic matrix is equal to one;
2. the modulus of the eigenvectors of a stochastic matrix are less or equal to one;
3. any stochastic matrix has l as an eigenvalue;
4. if λ is an eigenvalue of modulus equal to 1, and of multiplicity order equal to k, then the vector space generated by the eigenvectors associated with this eigenvalue (λ) is of dimension k.

A finite Markov chain is said to be:

1. a homogeneous (stationary or time homogeneous) chain if its associated transition matrix is stationary, i.e., $\Pi(n) = \Pi$;
2. a non-homogeneous chain if its associated transition matrix $\Pi(n)$, is non-stationary.

Example 2.16 *Consider 3 supermarkets denoted by s_1, s_2 and s_3. Each month the supermarket s_1 maintains 60% of its clients and losses the 20% that goes to s_2 and the rest to s_3; on the*

other hand, the supermarket s_2 maintains the 40% of its clients and losses the 50% that goes to s_1 and the 10% that goes to s_3; while supermarket s_3 retains the 80% of its clients and only losses the 20% that goes to s_2. The transition matrix obtained whit the previous information is as follows:

$$\Pi = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.4 & 0.1 \\ 0.0 & 0.2 & 0.8 \end{bmatrix}$$

Figure 2.1 shows the state transition diagram for this Markov chain.

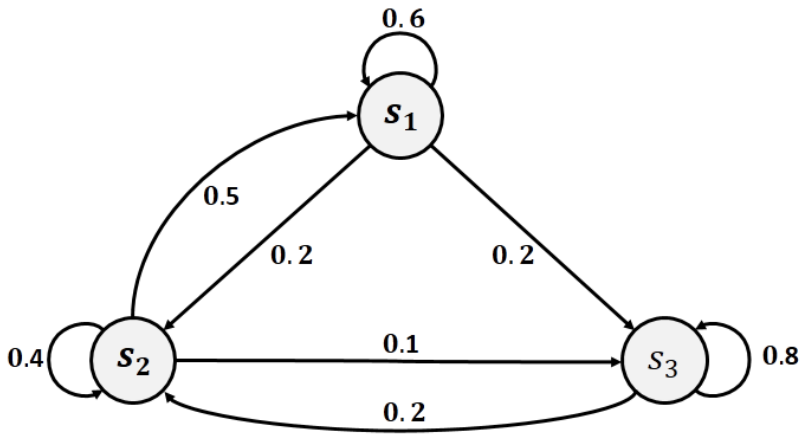


Figure 2.1 State transition diagram.

Definition 2.17 *For a homogeneous chain each l^{th} group $S(l)$ $(l = 1, ..., L)$ of communicating states is also said to be l^{th} ergodic subclass of states. The index L corresponds to the number of ergodic subclasses.*

An ergodic subclass (set of states) is a collection $S(l)$ of recurrent states with the probability that, when starting in one of the states in $S(l)$, all states will be visited with probability one. A Markov chain is ergodic if it has only one subclass, and that subclass is ergodic.

Definition 2.18 *If an homogeneous finite Markov chain has only one ergodic subclass and has no group of non-return states, i.e.,*

$$L = 1, \qquad S(0) = \emptyset,$$

it is said to be an ergodic homogeneous finite Markov chain.

Remark 2.19 *For any homogeneous finite Markov chain there exists a time n_0 such that the probabilities of transition from any initial states $s(1) = s_{(i)}$ to the state $s(n_0) = s_{(j)}$ are strictly positive, i.e.,*

$$\tilde{\pi}_{(j|i)}(n_0) > 0,$$

for $i, j = 1, ..., N$, where

$$\tilde{\pi}_{(j|i)}(n_0) := P\{s(n_0) = s_{(j)} \mid s(1) = s_{(i)}\} = \Pi^{n_0},$$

If there exists n_0 for a homogeneous Markov chain such that $\Pi^{n_0} > 0$ then, this Markov chain is ergodic.

2.3 Coefficient of ergodicity

In this section it is presented a non-traditional approach for ergodicity verification. The result shows that there exists the class of homogeneous Markov chains, called ergodic, which satisfy some additional conditions providing that after a long time such chains "forget" the initial states from which they have started.

For any time n and for any finite Markov chain with transition matrix

$$\Pi = \left[\pi_{(j|i)}\right]_{i,j=1,...,N}$$

containing N states, the following basic relation holds:

$$p(n+1) = \Pi^T p(n)$$

where $n = 1, 2, ...$ and the state distribution vector $p(n)$ is defined by

$$p^T(n) = \left(p_{(1)}(n), ..., p_{(N)}(n)\right), \quad \text{where} \quad p_{(i)}(n) = P\left\{s(n) = s_{(i)}\right\}.$$

Definition 2.20 *The state distribution vector*

$$(p^*)^T = \left(p^*_{(1)}, ..., p^*_{(N)}\right)$$

is called the stationary distribution of a homogeneous Markov chain with a given transition matrix

$$\Pi = \left[\pi_{(j|i)}\right]_{i,j=1,\ldots,N}$$

if it satisfies the following algebraic relations

$$p^*_{(j)} = \sum_{i=1}^{N} \pi_{(j|i)} p^*_{(i)}$$

Definition 2.21 *For an homogeneous finite Markov chain, the parameter $k_{erg}(n_0)$ defined by*

$$k_{erg}(n_0) := 1 - \frac{1}{2} \max_{i,j=1,\ldots,N} \sum_{m=1}^{N} \left|\tilde{\pi}_{(m|i)}(n_0) - \tilde{\pi}_{(m|j)}(n_0)\right| \in [0,1)$$

is said to be the coefficient of ergodicity of this Markov chain at time n_0, where

$$\tilde{\pi}_{(m|i)}(n_0) = P\left\{ s(n_0) = s_{(m)} \;\middle|\; s(1) = s_{(i)} \right\} = \left[\Pi_{(m|i)}\right]^{n_0}$$

is the probability to evolve from the initial state $s(1) = s_{(i)}$ to the state $s(n_0) = s_{(m)}$ after n_0 transitions.

Lemma 2.22 *The coefficient of ergodicity $k_{erg}(n_0)$ can be calculated as*

$$k_{erg}(n_0) \geq \min_{i,j=1,\ldots,N} \sum_{m=1}^{N} \min\left\{\tilde{\pi}_{(m|i)}(n_0), \tilde{\pi}_{(m|j)}(n_0)\right\}.$$

Its lower estimate is given by

$$k_{erg}(n_0) \geq \min_{i=1,\ldots,N} \max_{j=1,\ldots,N} \tilde{\pi}_{(j|i)}(n_0)$$

For the proof see [76].

If all the elements $\tilde{\pi}_{(j|i)}(n_0)$ of the transition matrix Π^{n_0} are positive, then the coefficient of ergodicity $k_{\text{erg}}(n_0)$ is also positive. Notice that there exist ergodic Markov chains with elements $\tilde{\pi}_{(j|i)}(n_0)$ equal to zero, but with a positive coefficient of ergodicity $k_{\text{erg}}(n_0)$.

Theorem 2.23 *The lower bound estimate of the ergodicity coefficient for a given finite homogeneous Markov chain*

$$\chi_{erg} := \min_{n_0} \max_{j=1,\ldots,N} \min_{i=1,\ldots,N} \tilde{\pi}_{(j|i)}(n_0)$$

is strictly positive, that is, $\chi_{erg} > 0$, then the following properties hold:

1. *there exists a unique stationary distribution*

$$\lim_{n\to\infty} p_{(i)}(n) := p^*_{(i)}$$

where $i = 1, ..., N$ *and the vector* p^* *describes a stationary distribution with positive components;*

2. *the convergence of the current-state distribution to the stationary one is exponential, then, for any initial state distribution* $p(1)$

$$\sup_{p(1)} \left| p_{(i)}(n) - p^*_{(i)} \right| \leq C \exp\{-D\, n\}$$

where

$$C = \frac{1}{1 - \chi_{erg}} \quad \textit{and} \quad D = \frac{1}{n_0^*} \ln C$$

and

$$n_0^* = \arg\min_{n_0} \left[\max_{j=1,...,N} \min_{i=1,...,N} \tilde{\pi}_{(j|i)}(n_0) \right]$$

For the proof see [76].

2.4 Controlled Markov chains

2.4.1 Discrete time Markov chains

The behavior of a controlled Markov chain can be described as follows: at each time n the system is observed to be in one state $s(n)$, whenever the system is in the state $s(n)$ one decision $a(n)$ (control action) is chosen according to some rule to achieve the desired control objective; in other words, the decision is selected to guarantee that the resulting state process performs satisfactorily. Then, at the next time $n + 1$ the system goes to the state $s(n + 1)$. In the case when the state and action sets are finite, and the transition from one state to another is random according to a fixed distribution, we deal with controlled finite Markov chains.

Consider the usual partial order for n-vectors x and y, the inequality $x \leq y$ means that $x^l \leq y^l$ for all $l = \overline{1, \mathcal{N}}$ ($l = 1, ..., \mathcal{N}$). We have that

$$x < y \Leftrightarrow x \leq y \text{ and } x \neq y$$

$$x << y \Leftrightarrow x^l < y^l \text{ for all } l = 1, ..., \mathcal{N}$$

A sequence $\{x^n\} \subset \mathbb{R}^n$ converging to x is said to converge in the direction $y \in \mathbb{R}^n$ if there is a sequence of positive numbers i_n such that $i_n \to 0$ and

$$\lim_{n\to\infty} (x^n - x)/i_n = y$$

Let S be a finite set, called the state space, consisting of a finite set of states $\{s_{(1)}, ..., s_{(N)}\}$, $N \in \mathbb{N}$. A Stationary Markov chain is a sequence of S-valued random variables $s(n)$, $n \in \mathbb{N}$, satisfying the Markov condition:

$$\begin{aligned} P\left(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}, s(n-1) = s_{(i_{n-1})}, ..., s(1) = s_{(i_1)}\right) \\ = P\left(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}\right) =: \pi_{(j|i)} \end{aligned} \tag{2.2}$$

The Markov chain can be represented by a complete graph whose nodes are the states, where each edge $(s_{(i)}, s_{(j)}) \in S^2$ is labeled by the transition probability (2.2). The matrix $\Pi = (\pi_{(j|i)})_{(s_{(i)}, s_{(j)}) \in S} \in [0,1]^{N\times N}$ determines the evolution of the chain: for each $m \in \mathbb{N}$, the power Π^m has in each entry $(s_{(i)}, s_{(j)})$ the probability of going from state $s_{(i)}$ to state $s_{(j)}$ in exactly m steps.

Definition 2.24 *A controlled homogeneous finite Markov chain is described by a 4-tuple*

$$MC = \{S, A, \mathbb{K}, \Pi\}$$

where:

- *S is a finite set of states, $S \subset \mathbb{N}$, endowed with a discrete topology;*
- *A is the finite set of actions. For each $s \in S$, $A(s) \subset A$ is the non-empty set of admissible actions at state $s \in S$. Without loss of generality we may take $A = \cup_{s\in S} A(s)$;*
- *$\mathbb{K} = \{(s, a) \mid s \in S, a \in A(s)\}$ is the set of admissible state-action pairs, which is a measurable subset of $S \times A$;*
- *$\Pi_{(k)} = \left[\pi_{(j|i,k)}\right]$ is a stationary controlled transition matrix, where*

  $$\pi_{(j|i,k)} \equiv P(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}, a(n) = a_{(k)})$$

  represents the probability associated with the transition from state $s_{(i)}$ to state $s_{(j)}$, $i, j = \overline{1, N}$ ($i = 1, ..., N$), under an action $a_{(k)} \in A(s_{(i)})$, $k = \overline{1, M}$ ($k = 1, ..., M$).

We say that a controlled homogeneous finite Markov chain is a communicating chain, if for any two states $s_{(i)}$ and $s_{(j)}$ of this chain, there exists a deterministic causal strategy $\{a(n)\}$ such that

$$a(n) = a(n)(s(1), a(1); ...; s(n-1), a(n-1); s(n))$$

such that for some n the conditional probability corresponding to the transition from $s_{(i)}$ to $s_{(j)}$ would be positive, i.e.,

$$P\left\{s(n) = s_{(j)} \mid s(1) = s_{(i)} \wedge \sigma(s(1), a(1); ...; s(n-1), a(n-1))\right\} \overset{a.s.}{>} 0$$

Definition 2.25 *A Markov Decision Process is a pair*

$$MDP = \{MC, J\}$$

where:

- *MC is a controlled homogeneous finite Markov chain (Definition 2.24);*
- $J : \mathbb{K} \to \mathbb{R}$ *is a cost/utility function, associating to each state a real value.*

The Markov property of the decision process (Definition 2.25) is said to be fulfilled if

$$\begin{aligned} &P\left(s(n+1) = s_{(j)} \mid (s(1), s(2), ..., s(n-1)), s(n) = s_{(i)}, a(n) = a_{(k)}\right) \\ &= P\left(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}, a(n) = a_{(k)}\right) \end{aligned}$$

The strategy (policy)

$$d_{(k|i)}(n) \equiv P\left(a(n) = a_{(k)} \mid s(n) = s_{(i)}\right)$$

represents the probability associated with the occurrence of an action $a(n)$ from state $s(n) = s_{(i)}$. The elements of the transition matrix for the Markov chain can be expressed as

$$\begin{aligned} &P\left(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}\right) = \\ &\sum_{k=1}^{M} P\left(s(n+1) = s_{(j)} \mid s(n) = s_{(i)}, a(n) = a_{(k)}\right) d_{(k|i)}(n) \end{aligned}$$

Let us denote the collection $\left\{d_{(k|i)}(n)\right\}$ by D_n as follows

$$D_n = \left\{d_{(k|i)}(n)\right\}_{k=\overline{1,M},\ i=\overline{1,N}}$$

A policy $\left\{d_n^{loc}\right\}_{n\geq 0}$ is said to be local optimal if for each $n \geq 0$ minimizes (or maximizes if we have a utility function) the conditional mathematical expectation of the cost function $J\left(s(n+1)\right)$ under the condition that the history of the process

$$\mathcal{F}_n := \left\{ D_0, P\left\{s(0) = s_{(j)}\right\}_{j=\overline{1,N}}; ...; D_{n-1}, P\left\{s(n) = s_{(j)}\right\}_{j=\overline{1,N}} \right\}$$

is fixed and can not be changed hereafter, i.e., it realizes the "one-step ahead" conditional optimization rule

$$d_n^{loc} := \arg \min_{d_n \in D_n} \mathrm{E}\left\{J(s(n+1)) \mid \mathcal{F}_n\right\}$$

where $J(s(n+1))$ is the cost function at the state $s(n+1)$. Notice that if $J(s(n+1))$ is a utility function then we have a $\max$ problem.

2.4.1.1 Discrete time Markov chains games

Definition 2.26 *A discrete-time Markov game is a pair*

$$\mathcal{G} = \{\mathcal{N}, MDP\}$$

where:

- MDP *is a discrete-time Markov decision process (Definition 2.25); and*
- $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ *is the set of players, each player is indexed by* $l = \overline{1, \mathfrak{n}}$.

The game for Markov chains consists of $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ players (denoted by $l = 1, ..., \mathfrak{n} = \overline{1, \mathfrak{n}}$) and begins at the initial state $s^l(0)$ which (as well as the states further realized by the process) is assumed to be measurable. Each of the players l is allowed to randomize, with distribution $d^l_{(k|i)}(n)$, over the action choices $a^l_{(k)} \in A^l\left(s^l_{(i)}\right)$, $i = \overline{1, N}$ and $k = \overline{1, M}$. From now on, consider only stationary strategies $d^l_{(k|i)}(n) = d^l_{(k|i)}$. These choices induce the state distribution dynamics

$$P^l\left(s^l(n+1)=s_{(j)}\right) = \sum_{i=1}^{N} \left(\sum_{k=1}^{M} \pi^l_{(j|i,k)} d^l_{(k|i)} \right) P^l\left(s^l(n) = s_{(i)}\right)$$

In the ergodic case when all Markov chains are ergodic for any stationary strategy $d^l_{(k|i)}$ the distributions $P^l(s^l(n+1) = s_{(j)})$ exponentially quickly converge to their limits $P^l\left(s^l = s_{(i)}\right)$

satisfying

$$P^l\left(s^l = s_{(j)}\right) = \sum_{i=1}^{N}\left(\sum_{k=1}^{M}\pi^l_{(j|i,k)}d^l_{(k|i)}\right)P^l\left(s^l{=}s_{(i)}\right)$$

2.4.2 Continuous time Markov chains

This section follows the concepts and definitions presented in [35]. Suppose that $s(t)$ (for each fixed $t \geq 0$) is a random variable on a probability space $(\Omega, \mathcal{F}, P)$ and takes values in the set S, i.e., $s(t)$ is an $\mathcal{F}$-measurable function on Ω and takes values in S. Then we call the set $\{s(t), t \geq 0\}$ a stochastic process with the state space S.

Definition 2.27 *A stochastic process $\{s(t), t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, with values in a countable set S (the state space of the process), is called a continuous-time Markov chain if, for any finite sequence of "times" $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$ and a corresponding set of states $s_{(i_1)}, s_{(i_2)}, ..., s_{(i_{n-1})} \in S$, it holds the Markov property*

$$P\left(s(t_{n+1}) = s_{(j)} \mid s(t_1) = s_{(i_1)}, ..., s(t_{n-1}) = s_{(i_{n-1})}, s(t_n) = s_{(i)}\right) =$$
$$P\left(s(t_{n+1}) = s_{(j)} \mid s(t_n) = s_{(i)}\right)$$

whenever $P\left(s(t_1) = s_{(i_1)}, ..., s(t_{n-1}) = s_{(i_{n-1})}, s(t_n) = s_{(i)}\right) > 0$

The probability

$$p(r, i, t, j) := P\left(s(t) = s_{(j)} \mid s(r) = s_{(i)}\right)$$

for $0 \leq r \leq t$ is called the chain's transition (probability) function. Note that $p(r, i, t, j)$ denotes the transition probability of the process being in state $s_{(j)}$ at time t starting from $s_{(i)}$ at time r.

Proposition 2.28 *Suppose that $p(r, i, t, j)$ is the transition function of a Markov chain. Then, for all $s_{(i)}, s_{(j)} \in S$ and $0 \leq r \leq t$:*

1. *$p(r, i, t, j) \geq 0$ and $\sum_{s_{(j)} \in S} p(r, i, t, j) \leq 1$.*
2. *The Kronecker delta: $p(r, i, t, j) = \delta_{(i,j)}$.*

3. *The Chapman-Kolmogorov equation:* $p(r,i,t,j) = \sum_{s_{(h)} \in S} p(r,i,v,h)p(v,h,t,j)$ *for all* $s_{(i)}, s_{(j)} \in S$ *and* $0 \leq r \leq v \leq t$.

4. $p(r,i,t,j)$ *is continuous in* $r \in [0,t]$, *right-continuous at* 0 *and left-continuous at* t.

5. $p(r,i,t,j)$ *is continuous in* $t \in [r,+\infty)$, *right-continuous at* r *and uniformly continuous in* $s_{(j)} \in S$

Definition 2.29 *A controllable continuous-time Markov chain is a 4-tuple*

$$CTMC = \{S, A, \mathbb{K}, Q\}$$

where:

- *The state space* S *is a finite set of states* $\left\{s_{(1)}, ..., s_{(N)}\right\}$, $N \in \mathbb{N}$, *endowed with the discrete topology;*

- *The set of actions* A *is a finite action (or control) space, for each* $s \in S$, $A(s) \subset A$ *is the non-empty set of admissible actions at state* $s \in S$ *and we shall suppose that is compact;*

- $\mathbb{K} = \{(s,a) | s \in S, a \in A(s)\}$ *is the class of admissible state-action pairs, which is considered a subspace of* $S \times A$;

- Q *is the matrix of the transition rates* $\left[q_{(j|i,k)}\right]$, *the transition from state* $s_{(i)}$ *to state* $s_{(j)}$ *under an action* $a_{(k)} \in A\left(s_{(i)}\right)$, $k = 1, ..., M$; *satisfying* $q_{(j|i,k)} \geq 0$ *for all* $(s,a) \in \mathbb{K}$ *and* $i \neq j$ *such that*

$$q_{(j|i,k)} = \begin{cases} -\sum_{i \neq j}^{N} \lambda_{(i,j)}(a_{(k)}), & \text{if } \ i = j \\ \\ \lambda_{(i,j)}(a_{(k)}), & \text{if } \ i \neq j \end{cases}$$

where $\lambda_{(i,j)}$ *is a transition rate between state* $s_{(i)}$ *and* $s_{(j)}$, $\lambda_{(i)} = \sum_{i \neq j}^{N} \lambda_{(i,j)}$. *This matrix is assumed to be conservative, i.e.,* $\sum_{j=1}^{N} q_{(j|i,k)} = 0$, *and stable, which means that*

$$q^*_{(i)} := \sup_{a \in A} q_{(i)}(a) < \infty \quad \forall\, i \in S$$

where $q_{(i)}(a) := -q_{(i,i)}(a) \geq 0$ *for all* $a \in A$.

Definition 2.30 *A continuous-time Markov Decision Process is a pair*

$$CTMDP = \{CTMC, U\}$$

where:

- $CTMC$ *is a controllable continuous-time Markov chain (Definition 2.29); and*
- $J : \mathbb{K} \to \mathbb{R}$ *is the (measurable) one stage cost/ utility function, associating to each state a real value.*

Now, we denote the probability transition matrix by

$$\Pi(t) = [\pi_{(r,i,\tau,j,k)}]_{i,j,k}, \quad \tau \geq r$$

such that, $\pi_{(r,i,\tau,j,k)} = \pi_{(0,i,t,j,k)},\ t = \tau - r\ \forall\ i, j \in S$ and where $\sum_{j=1}^{N} \pi_{(j|i,k)} = 1$. The Kolmogorov forward equations can be written as the matrix differential equation as follows:

$$\Pi'(t) = \Pi(t)\, Q; \quad \Pi(0) = I$$

$\Pi(t) \in \mathbb{R}^{N\times N}$, $I \in \mathbb{R}^{N\times N}$ is the identity matrix. This system can be solved by

$$\Pi(t) = \Pi(0)\, e^{Qt} = e^{Qt} := \sum_{t=0}^{\infty} \frac{t^n Q^n}{n!}$$

and at the stationary state, the probability transition matrix is defined as

$$\Pi^* = \lim_{t\to\infty} \Pi(t)$$

Definition 2.31 *The vector $P \in \mathbb{R}^N$ is called stationary distribution vector if*

$$\left(\Pi^\top\right)^* P = P$$

where $\sum_{i=1}^{N} P_{(i)} = 1$.

This vector can be seen as the long-run proportion of time that the process is in state $s_{(i)} \in S$.

Theorem 2.32 *The following statements are equivalent:*

- $Q^\top\, P = 0$
- $\Pi^\top(t)\, P = P;\ \ \forall\, t \geq 0$

The proof of this fact is easy in the case of a finite state space, recalling the Kolmogorov backward equation.

2.4.2.1 Continuous time Markov chains games

Definition 2.33 *A continuous-time Markov game is a pair*

$$\mathcal{G} = \{\mathcal{N}, CTMDP\}$$

where:

- $CTMDP$ *is a continuous-time Markov decision process (Definition 2.30); and*
- $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ *is the set of players, each player is indexed by* $l = \overline{1, \mathfrak{n}}$.

A strategy for player l is then defined as a sequence $d^l = \left\{d^l(t), t \geq 0\right\}$ of stochastic kernels $d^l(t)$ such that:

a. for each time $t \geq 0$, $d^l_{(k|i)}(t)$ is a probability measure on A^l such that $d^l_{(A^l(s_{(i)})|i)}(t) = 1$ and,

b. for every $E^l \in \mathcal{B}(A^l)$, $d^l_{(E^l|i)}(t)$ is a Borel-measurable function in $t \geq 0$.

We denoted by D^l the family of all strategies for player l. A multistrategy is a vector $\mathbf{d} = (d^1, ..., d^{\mathcal{N}}) \in D := \otimes_{l=1}^{\mathcal{N}} D^l$. From now on, we will consider only stationary strategies $d^l_{(k|i)}(t) = d^l_{(k|i)}$. For each strategy $d^l_{(k|i)}$ the associated transition rate matrix is defined as:

$$Q^l(d^l) := \left[q^l_{(i,j)}(d^l)\right] = \sum_{k=1}^{M} q^l_{(j|i,k)} d^l_{(k|i)}$$

such that on a stationary state distribution for all $d^l_{(k|i)}$ and $t \geq 0$ we have that

$$\Pi^{l*}(d) = \lim_{t \to \infty} e^{Q^l(d^l)t}$$

where $\Pi^{l*}\left(d^l\right)$ is a stationary transition controlled matrix.

2.5 Formulation of Markov chains games

Considering the utility matrix $U^l_{(i,j,k)}$ and the transition matrix $\pi^l_{(j|i,k)}$, the utility function that describes the behavior of each player is defined as

$$W^l_{(i,k)} = \sum_{j=1}^{N} U^l_{(i,j,k)} \pi^l_{(j|i,k)} \tag{2.3}$$

so that the average utility function $\mathbf{J}^l$ in the stationary regime can be expressed as

$$\mathbf{J}^l\left(c^1,...,c^{\mathcal{N}}\right) := \sum_{i=1}^{N}\sum_{k=1}^{M} W^l_{(i,k)} \prod_{l=1}^{\mathcal{N}} d^l_{(k|i)} P^l\left(s^l{=}s_{(i)}\right)$$

Given that $c^l := \left[c^l_{(i,k)}\right]_{i=\overline{1,N};k=\overline{1,M}}$ is a matrix with elements

$$c^l_{(i,k)} = d^l_{(k|i)} P^l\left(s^l{=}s_{(i)}\right) \tag{2.4}$$

it follows that

$$\mathbf{J}^l\left(c^1,...,c^{\mathcal{N}}\right) := \sum_{i=1}^{N}\sum_{k=1}^{M} W^l_{(i,k)} \prod_{l=1}^{\mathcal{N}} c^l_{(i,k)} \tag{2.5}$$

Notice that by (2.4) it follows that

$$P^l\left(s^l{=}s_{(i)}\right) = \sum_{k=1}^{M} c^l_{(i,k)}, \qquad d^l_{(k|i)} = \frac{c^l_{(i,k)}}{\sum\limits_{k=1}^{M} c^l_{(i,k)}} \tag{2.6}$$

The variable $c^l_{(i,k)}$ satisfies the following restrictions:

1. Each vector from the matrix $c^l_{(i,k)}$ represents a stationary mixed-strategy that belongs to the simplex

$$c^l \in C^l_{\text{adm}} = \left\{ c^l_{(i,k)} \in \mathbb{R}^{N\times M} : c^l_{(i,k)} \ \geq\ 0,\ \sum_{i=1}^{N}\sum_{k=1}^{M} c^l_{(i,k)}{=}1 \right\} \tag{2.7}$$

2. The variable $c^l_{(i,k)}$ satisfies the ergodicity constraints, and belongs to the convex, closed and bounded set defined as follows:

$$c^l \in C^l_{\text{adm}} = \left\{ h^l_{(j)}(c^l) = \sum_{i=1}^{N}\sum_{k=1}^{M} \pi^l_{(j|i,k)} c^l_{(i,k)} - \sum_{k=1}^{M} c^l_{(j,k)}{=}\,0 \right\} \tag{2.8}$$

3. And, in case of continuous time Markov games, the variable $c^l_{(i,k)}$ satisfies the continuous time condition:

$$c^l \in C^l_{\text{adm}} = \left\{ \sum_{i=1}^{N}\sum_{k=1}^{M} q^l_{(j|i,k)} c^l_{(i,k)} = 0 \right\} \tag{2.9}$$

In the ergodic case $\sum_{k=1}^{M} c^l_{(i,k)} > 0$ for all $l = \overline{1,\mathcal{N}}$. The individual aim of each player is $\min\limits_{c^l\in C^l_{\text{adm}}} \mathbf{J}^l(c^l)$ or $\max\limits_{c^l\in C^l_{\text{adm}}} \mathbf{J}^l(c^l)$ depending on whether $\mathbf{J}^l(c^l)$ is a cost or utility function.

Part I

The L_p-Stackelberg/Nash game

Chapter 3

The Strong $\mathbf{L_p}-$Nash Equilibrium

3.1 Introduction

Nash equilibrium [62] is a fundamental concept in game theory and the most widely used method of predicting the outcome of a strategic interaction of several decision makers in non-cooperative games. It is concerned with a strategy profile such that no player can unilaterally change her/his strategy to increase her/his payoff. However, non-cooperative equilibrium has individually stability and the collective stability is a special case of the Nash equilibrium called strong Nash equilibrium (SNE).

The SNE was introduced by Aumann [7] for cooperative games. It may benefit from establishing coalitions with other players and there is no coalition that can definitely improve their payoffs by a collective deviation. A SNE is a Nash Equilibrium for which no coalition of players has a joint deviation that improves the payoff of each member of the coalition. In cooperative games the players can find a strategy producing the smaller total expected loss, such cooperative strategy leads to strong Pareto optimal solution of the game.

The difference between the non-cooperative and cooperative Nash equilibrium can be exemplified by the following version of the Prisoner's dilemma [82, 84].

Example 3.1 ***Prisoner's dilemma***. *Consider the following two-person game with two possible strategies:* a_l *and* b_l *(*$l = 1, 2$*) and the utilities represented by the following matrix*

$Player1 \backslash \mathbf{Player2}$	$\mathbf{a_2}$	$\mathbf{b_2}$
a_1	$(2, \mathbf{2})$	$(0, \mathbf{3})$
b_1	$(3, \mathbf{0})$	$(1, \mathbf{1})$

The following interpretation of the game explains its name. The column $\mathbf{Player2}$ *is society, the row* $Player1$ *is a citizen. When free, the citizen can behave well* (a_1) *or commit a crime* (b_1). *Society can jail him* (b_2) *or let him free* (a_2). *Commission of a crime benefits the individual but damages society; punishing the criminal is damaging to both. The players are planning to play* (a_1, a_2) *unless one of them deviates. If player 1 deviates* (b_1, a_2), *player 2 punishes him by forcing* (b_1, b_2). *Clearly, the punishing player profits from the punishing arrangement and he has no motivation to forgive the deviant. It is easily verified that* (a_1, a_2) *is indeed a strong Nash equilibrium. The only Nash equilibrium is* (b_1, b_2).

There are several proposals reported in the literature to search strong Nash equilibria for specific classes of games, e.g., congestion games, connection games, maxcut games, voting models, coalition formation and other fields. Proving the existence of SNE is a difficult problem [64] and there are a small number of computational tools available for finding the SNE.

In order to solve the problem many refinements of Nash equilibrium were proposed to have a better model of the real world. Ichiishi [41] proposed a social coalition equilibrium where an abstract model of society in which each member can cooperate with others by forming a coalition, but at the same time can be influenced by the members outside the coalition. Greenberg and Weber [34] investigated the existence and proposed a partial characterization of a "strong Tiebout equilibrium" consisting of an endogenously formed partition of the individuals into disjoint jurisdictions with each jurisdiction producing and financing its own public goods where no group of individuals can benefit by establishing their own community. Demange and Henriet [25] proved that in a sustainable oligopoly each consumer chooses the firm which proposes the price-quality schedule he prefers, firms earn non-negative profits and no new firm could attract consumers and make profits. Demange [24] proposed two forces that are at work to explain the formation of coalitions that partition the society in a stable way: the

increasing power of the coalitions which incites to cooperate, the heterogeneity of the agents which leads to the formation of subgroups. Konishi et al. [49] proved that a non-cooperative game with a finite set of players and common finite strategy sets possesses a strong Nash equilibrium in pure strategies whenever individuals' preferences satisfy independence of irrelevant choices, anonymity, and partial rivalry. Then, he [50] examined the conditions which guarantee that the set of coalition-proof Nash equilibria coincides with the set of strong Nash equilibria in the normal form games without spillovers. Hotzman [38] obtained conditions for the existence of a strong equilibrium in congestion games, as well for the equivalence of Nash and strong equilibria, giving conditions for uniqueness and for Pareto optimality of the Nash equilibrium. Rozenfeld [81] dealt with possible deviations by coalitions of players in congestion games studying the existence of strong and correlated strong equilibria in monotone congestion games. Gatti et al. [31] suggested that in order for a n–agent game to have at least one non–pure–strategy SNE, the agents' payoffs restricted to the agents' supports must lie on an $(n-1)$–dimensional space. Gatti et al. [30] provided a nonlinear program in which a strategy profile is forced to be Pareto efficient with respect to coalition correlated strategies. It is a sufficient, but non–necessary, condition for the existence of an SNE that can be used to search for an SNE. Kubica and Wozniak [51] provided an interval method approach to verify the existence of equilibria in certain points and proposed an algorithm for finding the SNE.

However, these proposal and algorithms fail in establishing a proper formulation regarding existence, recognition, and computation for the Pareto optimality. Most of them find a Nash equilibrium and then verify the Pareto optimality.

This chapter presents a method for computing the strong L_p-Nash equilibrium in case of discrete time Markov chains games [101, 96]. The problem is solved in terms of the L_p-norm: players choose a strategy that minimizes the distance to the utopian minimum and no other strategy produces a smaller total expected loss. This means that there exists an optimal solution that is a strong Pareto optimal point and it is the closest solution to the minimum utopia point. The strong Pareto optimal solution corresponds to the strong Nash equilibrium. First, a general solution is presented for the L_p-norm for computing the strong L_p-Nash equilibrium. Then, an explicit solution is suggested for the norms L_1, L_2 and L_∞. For solving the problem,

the extraproximal method is used [5]: a natural extension of the proximal and the gradient optimization methods used for solving the more difficult problems for finding an equilibrium point in game theory. The extraproximal method is defined by a two-step iterated procedure consisting of a prediction step that calculates the preliminary position approximation to the equilibrium point, and a basic adjustment of the previous step. The method is designed for the static strong Nash game in terms of nonlinear programming problems implementing the Lagrange principle; then, employing the Tikhonov's regularization method ensures the convergence of the cost-functions to a unique strong L_p-Nash equilibrium. The nonlinear programming problem is formulated considering several linear constraints employing the c-variable method for making the problem computationally tractable. For solving each equation of the extraproximal optimization approach the projection gradient method is used. It is proved that the proposed method converges in exponential time to a strong L_p-Nash equilibrium.

3.2 Formulation of the problem

To study the existence of Pareto policies it is necessary first follow the well-known "scalarization" approach. Thus, given a n-vector $\lambda > 0$ consider the cost-function $\mathbf{J}$. Let

$$u^l := \text{col}\ \left(c^l\right),\ U^l := C^l_{\text{adm}}\ ,\ U := \bigotimes_{l=1}^{\mathcal{N}} U^l$$

for $l = \overline{1, \mathfrak{n}}$, where col is the column operator.

The Pareto set can be defined as [32, 33]

$$\mathcal{P} := \left\{ u^*\left(\lambda\right) := \arg\min_{u \in U} \left[\sum_{l=1}^{\mathfrak{n}} \lambda^l\ \mathbf{J}^l(u) \right],\ \lambda \in \mathcal{S}^{\mathfrak{n}} \right\}$$

such that

$$\mathcal{S}^{\mathfrak{n}} := \left\{ \lambda \in \mathbb{R}^{\mathfrak{n}} : \lambda \in [0,1]\,,\ \sum_{l=1}^{\mathfrak{n}} \lambda^l = 1 \right\}$$

for

$$\mathbf{J}\left(u^*(\lambda)\right) = \left(\mathbf{J}^1\left(u^*\left(\lambda\right)\right), \mathbf{J}^2\left(u^*\left(\lambda\right)\right), ..., \mathbf{J}^{\mathfrak{n}}\left(u^*\left(\lambda\right)\right)\right)$$

The vector u^* is called a Pareto optimal solution for $\mathcal{P}$. The Pareto front is defined as the image of $\mathcal{P}$ under $\mathbf{J}$ as follows

$$\mathbf{J}(\mathcal{P}):=\left\{\left(\mathbf{J}^1\left(u^*\left(\lambda\right)\right),\mathbf{J}^2\left(u^*\left(\lambda\right)\right),...,\mathbf{J}^{\mathfrak{n}}\left(u^*\left(\lambda\right)\right)\right)\mid u^*\in\mathcal{P}\right\}$$

A **Nash equilibrium** is a strategy $u^*=(u^{1*},...,u^{\mathfrak{n}*})$ such that

$$\mathbf{J}\left(u^{1*},...,u^{\mathfrak{n}*}\right)\leq\mathbf{J}\left(u^{1*},...,u^l,...,u^{\mathfrak{n}*}\right)$$

for any $u^l\in U$.

A **strong Nash equilibrium** is a strategy $u^{**}=(u^{1**},...,u^{\mathfrak{n}**})$ such that there does not exist any $u^l\in U$, $u^l\neq u^{l**}$ such that

$$\mathbf{J}\left(u^{1**},...,u^l,...,u^{\mathfrak{n}**}\right)\leq\mathbf{J}\left(u^{1**},...,u^{\mathfrak{n}**}\right)$$

for any $u^l\in U$.

Remark 3.2 *The game problem is to find a policy u^* that minimizes $\mathbf{J}(u^1,...,u^{\mathfrak{n}})$ in the sense of Pareto.*

Let $\mathcal{P}$ be a subset of $\mathbb{R}^{\mathfrak{n}}$. The tangent cone to $\mathcal{P}$ at $u\in\mathcal{P}$ is the set of all the directions $u'\in\mathbb{R}^{\mathfrak{n}}$ in which some sequence in $\mathcal{P}$ converges to u. A vector $u^*\in\mathcal{P}$ in $\mathbb{R}^{\mathfrak{n}}$ is said to be

1. a *Pareto point* of $\mathcal{P}$ if there is no $u\in\mathcal{P}$ such that $u<u^*$;
2. a *weak Pareto* point of $\mathcal{P}$ if there is no $u\in\mathcal{P}$ such that $u<<u^*$;
3. a *proper Pareto* point of $\mathcal{P}$ if u^* is a Pareto point and, in addition, the tangent cone to $\mathcal{P}$ at u^* does not contain vectors $u'<0$.

A policy u^* is said to be a Pareto policy (or Pareto optimal) if there is no policy u such that $\mathbf{J}(u)<\mathbf{J}(u^*)$, and similarly for weak or proper Pareto policies.

Problem Formulation. Let

$$\mathbf{J}^{l*}=\inf_{u\in U}\mathbf{J}^l(u)$$

and define the utopia minimum as $\mathbf{J}^* = (\mathbf{J}^{1*}, ..., \mathbf{J}^{\mathfrak{n}*})$ (infeasible in general), then the resulting problem is to find the values of

$$\lambda^* = \arg\min_{\lambda \in \mathcal{S}^{\mathfrak{n}}} \sum_{l=1}^{\mathfrak{n}} \lambda^l \, \mathbf{J}^l(u^*(\lambda))$$

in order to find the strong Nash equilibrium $u^*(\lambda)$ whose cost vector $\mathbf{J}(u^*(\lambda))$ is the "closest" to $\mathbf{J}^*$ in the usual Euclidean norm. Let $\|.\|$ be the Euclidean norm in $\mathbb{R}^{\mathfrak{n}}$ and let $\varrho : D \to \mathbb{R}_+$ be the map defined as

$$\varrho(u) := \|\mathbf{J}(u) - \mathbf{J}^*\|$$

$\mathbf{J}^*$ is also known as the utopian or the ideal or the shadow minimum [91, 92]. This is a utility function (or a strongly monotonically increasing function [42]) for the Markov chains game in the sense that if u and u' are such that $\mathbf{J}(u) < \mathbf{J}(u')$, then $\varrho(u) < \varrho(u')$.

A policy u^* is said to be *strong Pareto optimal* (or a strong Pareto policy) if it minimizes the function ϱ that is,

$$\varrho(u^*) = \inf \{\varrho(u) \mid u^* \in D\} =: \varrho^*$$

As ϱ is a utility function, it is clear that a strong Pareto policy is Pareto optimal, but of course the converse is not true.

3.3 The strong L_p-Nash equilibrium

Consider a game with $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ players with strategies $u^l \in U^l \ \left(l = \overline{1, \mathfrak{n}}\right)$ where U is a convex and compact set. Denote by $u = (u^1, ..., u^{\mathfrak{n}})^\top \in U$ the joint strategy of the players and $u^{\hat{l}}$ is a strategy of the rest of the players adjoint to u^l, namely,

$$u^{\hat{l}} := \left(u^1, ..., u^{l-1}, u^{l+1}, ..., u^{\mathfrak{n}}\right)^\top \in U^{\hat{l}} := \bigotimes_{m=1,\, m \neq l}^{\mathfrak{n}} U^m$$

such that $u = (u^l, u^{\hat{l}}) \ \left(l = \overline{1, \mathfrak{n}}\right)$.

Players try to reach the one of Nash equilibria, that is, to find a joint strategy $u^* = (u^{1*}, ..., u^{\mathfrak{n}*}) \in U$ satisfying for any admissible $u^l \in U^l$ and any $l = \overline{1, \mathfrak{n}}$

$$G_{L_p}(u, \hat{u}(u)) := \left(\sum_{l=1}^{\mathfrak{n}} \left| \left(\min_{u^l \in U^l} \varphi_l \left(u^l, u^{\hat{l}}\right) \right) - \varphi_l \left(u^l, u^{\hat{l}}\right) \right|^p \right)^{1/p} \tag{3.1}$$

where $\hat{u}(u) = (u^{\hat{1}\top}, ..., u^{\hat{\mathfrak{n}}\top})^\top \in \hat{U} \subseteq \mathbb{R}^{\mathfrak{n}(\mathfrak{n}-1)}$ and $p \geq 1$ [92, 91]. Here $\varphi_l\left(u^l, u^{\hat{l}}\right)$ is the cost-function of the player l which plays the strategy $u^l \in U^l$ and the rest of the players the strategy $u^{\hat{l}} \in U^{\hat{l}}$.

If we consider the utopia point

$$\bar{u}^l := \arg \min_{u^l \in U^l} \varphi_l\left(u^l, u^{\hat{l}}\right) \tag{3.2}$$

then, we can rewrite eq. (3.1) as follows

$$G_{L_p}(u, \hat{u}(u)) := \left(\sum_{l=1}^{\mathfrak{n}} \left|\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right)\right|^p\right)^{1/p}$$

The functions $\varphi_l\left(u^l, u^{\hat{l}}\right)$ $\left(l = \overline{1, \mathfrak{n}}\right)$ are assumed to be convex in all their arguments.

Remark 3.3 *The function $G_{L_p}(u, \hat{u}(u))$ satisfies the Nash property*

$$\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right) \leq 0 \tag{3.3}$$

for any $u^l \in U^l$ and all $l = \overline{1, \mathfrak{n}}$

Remark 3.4 *Following restrictions (2.7) and (2.8), the set U admissible (U_{adm}) is defined as follows*

$$U_{adm} = C^1_{adm} \times \cdots \times C^{\mathfrak{n}}_{adm}$$

Definition 3.5 *A strategy $u^* \in U_{adm}$ is said to be a L_p-Nash equilibrium if*

$$u^*_{L_p} \in Arg \min_{u \in U_{adm}} \left\{G_{L_p}(u, \hat{u}(u))\right\}$$

Remark 3.6 *If $G_{L_p}(u, \hat{u}(u))$ is strictly convex then*

$$u^*_{L_p} = \arg \min_{u \in U_{adm}} \left\{G_{L_p}(u, \hat{u}(u))\right\}$$

Definition 3.7 *A strategy $u^{**} \in U_{adm}$ is said to be a strong L_p-Nash equilibrium if*

$$u^{**}_{L_p} \in Arg \min_{u \in U_{adm}, \lambda \in \mathcal{S}^{\mathfrak{n}}} \left\{G_{L_p}(u(\lambda), \hat{u}(u, \lambda))\right\}$$

where

$$G_{L_p}(u(\lambda), \hat{u}(u, \lambda) := \left(\sum_{l=1}^{\mathfrak{n}} \lambda^l \left|\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right)\right|^p\right)^{1/p} \tag{3.4}$$

Remark 3.8 *If $G_{L_p}(u(\lambda), \hat{u}(u,\lambda))$ is strictly convex then*

$$u^{**}_{L_p} = \arg \min_{u \in U_{adm}, \lambda \in \mathcal{S}^{\mathfrak{n}}} \left\{ G_{L_p}(u(\lambda), \hat{u}(u,\lambda)) \right\}$$

3.3.1 The strong Nash equilibrium for norms L_1 and L_2

If $\varphi_l\left(u^l, u^{\hat{l}}\right) = \varphi_l\left(\bar{u}^l, u^{\hat{l}}\right)$ in eq. (3.4), then u^l achieves the minimum of $\mathbf{J}(u)$ giving the optimal solution to each player l. But such cases are rarely and hardly ever take place. Then, players choose a strategy that minimizes the distance from $\varphi_l\left(u^l, u^{\hat{l}}\right)$ to $\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right)$ where $\bar{u}^l$ satisfies the utopia (3.2), this corresponds to the strong Nash equilibrium. That is, no other strategy produces a smaller total expected loss in the sense of the distance given by eq. (3.4). This means that there exists an optimal solution u^l that is a strong Pareto optimal solution and it is the closest solution to the utopia point $\bar{u}^l$.

To find the strong L_p-Nash equilibrium (Definition 3.7) of this minimization L_p-norm problem, we propose the following solutions:

Definition 3.9 *The strong L_1-Nash equilibrium $u^{**}_{L_p} \in U$ can be expressed for L_1 norm as follows*

$$u^{**}_{L_p} = \arg \min_{u \in U, \lambda \in \mathcal{S}^{\mathfrak{n}}} G_{L_p}(u(\lambda), \hat{u}(u,\lambda))$$

$$G_{L_p}(u(\lambda), \hat{u}(u,\lambda)) := \sum_{l=1}^{\mathfrak{n}} \lambda^l \left| \varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right) \right|$$

$$\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) := \min_{z^l \in U^l} \varphi_l\left(z^l, u^{\hat{l}}\right)$$

Definition 3.10 *The strong L_2-Nash equilibrium $u^{**}_{L_p} \in U$ can be expressed for L_2 norm as follows*

$$u^{**}_{L_p} = \arg \min_{u \in U, \lambda \in \mathcal{S}^{\mathfrak{n}}} G_{L_p}(u(\lambda), \hat{u}(u,\lambda))$$

$$G_{L_p}(u(\lambda), \hat{u}(u,\lambda)) := \left(\sum_{l=1}^{\mathfrak{n}} \lambda^l \left| \varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right) \right|^2 \right)^{1/2}$$

$$\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) := \min_{z^l \in U^l} \varphi_l\left(z^l, u^{\hat{l}}\right)$$

Applying the Lagrange principle (see, for example, [76]) for Definitions 3.9 and 3.10, we may conclude

$$u_{L_p}^{**} = \arg \min_{u\in U,\hat{u}(u)\in\hat{U},\lambda\in\mathcal{S}^{\mathfrak{n}}} \max_{\xi\geq 0} \mathcal{L}_\delta(u,\hat{u}(u),\lambda,\xi)$$

$$\mathcal{L}_\delta(u,\hat{u}(u),\lambda,\xi) := G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) + \sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N}\xi_j^l h_j^l(c) - \tfrac{\delta}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N}\left(\xi_j^l\right)^2 \tag{3.5}$$

where

$$G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) = \left(\sum_{l=1}^{\mathfrak{n}}\lambda^l\left|\varphi_l\left(\bar{u}^l,u^{\hat{l}}\right)-\varphi_l\left(u^l,u^{\hat{l}}\right)\right|^p\right)^{1/p} + \frac{\delta}{2}\left(\|u\|^2+\|\hat{u}(u)\|^2+\|\lambda\|^2\right)$$

Now, the function $G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right)$ is strictly convex if the Hessian matrix is positive semi-definite, then $G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right)$ attains a minimum at $(u(\lambda),\hat{u}(u,\lambda))$ if

$$\nabla^2 G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) =$$

$$\begin{bmatrix} \frac{\partial^2}{(\partial u_1)^2}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) & \cdots & \frac{\partial^2}{\partial u_1\partial u_{\mathfrak{n}}}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) \\ \frac{\partial^2}{\partial u_2\partial u_1}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) & \cdots & \frac{\partial^2}{\partial u_2\partial u_{\mathfrak{n}}}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) \\ \cdots & \cdots & \cdots \\ \frac{\partial^2}{\partial u_{\mathfrak{n}}\partial u_1}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) & \cdots & \frac{\partial^2}{(\partial u_{\mathfrak{n}})^2}G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right) \end{bmatrix} =$$

$$\begin{bmatrix} \delta I_{n_1\times n_1} & \mathcal{DG}_{1,2}(\hat{u}_{1,2}) & \cdots & \mathcal{DG}_{1,\mathfrak{n}}(\hat{u}_{1,\mathfrak{n}}) \\ \mathcal{DG}_{2,1}(\hat{u}_{2,1}) & \delta I_{n_2\times n_2} & \cdots & \mathcal{DG}_{3,2}(\hat{u}_{3,2}) \\ \cdots & \cdots & \cdots & \cdots \\ \mathcal{DG}_{3,1}(\hat{u}_{3,1}) & \mathcal{DG}_{3,2}(\hat{u}_{3,2}) & \cdots & \delta I_{n_{\mathfrak{n}}\times n_{\mathfrak{n}}} \end{bmatrix} > 0$$

or, equivalently, δ should provide the inequality

$$\min_{u\in U,\hat{u}\in\hat{U}}\left[\Lambda_{\min}\left(\nabla^2 G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\right)\right)\right] > 0$$

where $\Lambda_{\min}$ is the minimum eigenvalue. Here, $\hat{u}_{ik}$ is independent of $u^{(i)}$ and $u^{(k)}$, that is, $\frac{\partial}{\partial u^{(i)}}\hat{u}_{ik} = 0$ and $\frac{\partial}{\partial u^{(k)}}\hat{u}_{ik} = 0$.

With sufficiently large δ, the considered functions provide the uniqueness of the conditional optimization problem (3.5). Notice also that the Lagrange function in (3.5) satisfies the saddle-point condition [75], namely, for all $u \in U$, $\hat{u} \in \hat{U}$, $\lambda \in \mathcal{S}^{\mathrm{n}}$ and $\xi \geq 0$ we have

$$\mathcal{L}_\delta(u^*_\delta, \hat{u}^*_\delta(u), \lambda^*_\delta, \xi_\delta) \leq \mathcal{L}_\delta(u^*_\delta, \hat{u}^*_\delta(u), \lambda^*_\delta, \xi^*_\delta) \leq \mathcal{L}_\delta(u_\delta, \hat{u}_\delta(u), \lambda_\delta, \xi^*_\delta)$$

3.3.2 Strong $L_\infty-$Nash equilibrium

Definition 3.11 *The strong Nash equilibrium $u^{**} \in U$ can be expressed for L_∞ as follows*

$$u^{**}_{L_p} \in \arg \min_{u \in U, \lambda \in \mathcal{S}^{\mathrm{n}}} G_{L_p}\left(u(\lambda), \hat{u}(u, \lambda)\right)$$

$$G_{L_p}\left(u(\lambda), \hat{u}(u, \lambda)\right) := \max_l \left| \lambda^l \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \right|$$

$$\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) := \min_{z^l \in U^l} \varphi_l \left(z^l, u^{\hat{l}} \right)$$

That implies

$$\left| \lambda^l \left[\varphi_l \left(u^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \right| \leq t, \qquad t \to \min_{t,u,\hat{u}}$$

Then, applying the Lagrange principle we have

$$\begin{aligned} L(t, u, \hat{u}(u), \lambda, \theta) :&= t + \sum_{l=1}^{\mathrm{n}} \theta^l \left(\left| \lambda^l \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \right| - t \right) \\ &= t \left(1 - \sum_{l=1}^{\mathrm{n}} \theta^l \right) + \sum_{l=1}^{\mathrm{n}} \theta^l \left(\left| \lambda^l \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \right| \right) \end{aligned}$$

It has a minimum if and only if θ belongs to the simplex, i.e., $\theta \in \mathcal{S}^{\mathrm{n}}$

$$\mathcal{S}^{\mathrm{n}} := \left\{ \theta \in \mathbb{R}^{\mathrm{n}} \ : \ \theta \in [0, 1] \, , \ \sum_{l=1}^{\mathrm{n}} \theta^l = 1 \right\}$$

Then, the L_∞-norm problem is reduced to the form

$$L(u, \hat{u}(u), \lambda, \theta) = \sum_{l=1}^{\mathrm{n}} \theta^l \left| \lambda^l \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \right| \to \min_{u, \hat{u}(u), \lambda \in \mathcal{S}^{\mathrm{n}}} \max_{\theta \in \mathcal{S}^{\mathrm{n}}}$$

Remark 3.12 *Applying the Lagrange principle for Definition 3.11, we may conclude that eq. (3.5) can be rewritten as follows*

$$u_{L_p}^{**} = \arg \min_{u\in U,\hat{u}(u)\in\hat{U},\lambda\in\mathcal{S}^{\mathfrak{n}}} \max_{\xi\geq 0,\theta\in\mathcal{S}^{\mathfrak{n}}} \mathcal{L}_\delta(u,\hat{u}(u),\lambda,\xi,\theta)$$

$$\mathcal{L}_\delta(u,\hat{u}(u),\lambda,\xi,\theta) := L_\delta(u,\hat{u}(u),\lambda,\theta) + \sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N} \xi_j^l h_j^l(c) - \tfrac{\delta}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N}\left(\xi_j^l\right)^2 \tag{3.6}$$

where

$$L_\delta(u,\hat{u}(u),\lambda,\theta) = \sum_{l=1}^{\mathfrak{n}} \theta^l \left|\lambda^l\left[\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right)\right]\right| + \frac{\delta}{2}\left(\|u\|^2 + \|\hat{u}(u)\|^2 + \|\lambda\|^2 - \|\theta\|^2\right)$$

3.4 The proximal format

In the proximal format (see [5]) the relation (3.5) can be expressed as

$$\begin{gathered}
\xi_\delta^* = \arg\max_{\xi\geq 0}\left\{-\tfrac{1}{2}\|\xi-\xi_\delta^*\|^2 + \gamma\mathcal{L}_\delta(u_\delta^*,\hat{u}_\delta^*(u),\lambda_\delta^*,\xi)\right\} \\
u_\delta^* = \arg\min_{u\in U}\left\{\tfrac{1}{2}\|u-u_\delta^*\|^2 + \gamma\mathcal{L}_\delta(u,\hat{u}_\delta^*(u),\lambda_\delta^*,\xi_\delta^*)\right\} \\
\hat{u}_\delta^*(u) = \arg\min_{\hat{u}\in\hat{U}}\left\{\tfrac{1}{2}\|\hat{u}(u)-\hat{u}_\delta^*(u)\|^2 + \gamma\mathcal{L}_\delta(u_\delta^*,\hat{u}(u),\lambda_\delta^*,\xi_\delta^*)\right\} \\
\lambda_\delta^* = \arg\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\left\{\tfrac{1}{2}\|\lambda-\lambda_\delta^*\|^2 + \gamma\mathcal{L}_\delta(u_\delta^*,\hat{u}_\delta^*(u),\lambda,\xi_\delta^*)\right\}
\end{gathered} \tag{3.7}$$

Given (3.6) the proximal format for the L_∞-norm problem will be extended with the following equation

$$\theta_\delta^* = \arg\max_{\theta\in\mathcal{S}^{\mathfrak{n}}}\left\{-\tfrac{1}{2}\|\theta-\theta_\delta^*\|^2 + \gamma\mathcal{L}_\delta(u_\delta^*,\hat{u}_\delta^*(u),\lambda_\delta^*,\xi_\delta^*,\theta)\right\}$$

where the solutions u_δ^*, $\hat{u}_\delta^*(u)$, λ_δ^*, θ_δ^* and ξ_δ^* depend on the parameters $\delta,\gamma>0$.

3.5 The extraproximal method

The Extraproximal Method for the conditional optimization problems (3.5) was suggested by Antipin [5]. The general format iterative version $(n=0,1,...)$ of the extraproximal method with some fixed admissible initial values $(u_0\in U,\ \hat{u}_0(u)\in\hat{U},\ \lambda_0\in[0,1]$, and $\xi_0\geq 0)$ is as follows

1. The *first half-step* (prediction):

$$\begin{aligned}
\bar{\xi}_n &= \arg\min_{\xi\geq 0}\left\{\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),\lambda_n,\xi)\right\}\\
\bar{u}_n &= \arg\min_{u\in U}\left\{\tfrac{1}{2}\|u-u_n\|^2+\gamma\mathcal{L}_\delta(u,\hat{u}_n(u),\lambda_n,\bar{\xi}_n)\right\}\\
\overline{\hat{u}}_n(u) &= \arg\min_{\hat{u}\in\hat{U}}\left\{\tfrac{1}{2}\|\hat{u}(u)-\hat{u}_n(u)\|^2+\gamma\mathcal{L}_\delta(u_n,\hat{u}(u),\lambda_n,\bar{\xi}_n)\right\}\\
\bar{\lambda}_n &= \arg\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\left\{\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),\lambda,\bar{\xi}_n)\right\}
\end{aligned} \tag{3.8}$$

2. The *second half-step* (basic)

$$\begin{aligned}
\xi_{n+1} &= \arg\min_{\xi\geq 0}\left\{\tfrac{1}{2}\|\mu-\mu_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{\lambda}_n,\xi)\right\}\\
u_{n+1} &= \arg\min_{u\in U}\left\{\tfrac{1}{2}\|u-u_n\|^2+\gamma\mathcal{L}_\delta(u,\overline{\hat{u}}_n(u),\bar{\lambda}_n,\bar{\xi}_n)\right\}\\
\hat{u}_{n+1}(u) &= \arg\min_{\hat{u}\in\hat{U}}\left\{\tfrac{1}{2}\|\hat{u}(u)-\hat{u}_n(u)\|^2+\gamma\mathcal{L}_\delta(\bar{u}_n,\hat{u}(u),\bar{\lambda}_n,\bar{\xi}_n)\right\}\\
\lambda_{n+1} &= \arg\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\left\{\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\lambda,\bar{\xi}_n)\right\}
\end{aligned} \tag{3.9}$$

Then, given (3.6) the extraproximal method will be extended with the following equations

1. The *first half-step* (prediction):

$$\bar{\theta}_n = \arg\min_{\theta\in\mathcal{S}^{\mathfrak{n}}}\left\{\tfrac{1}{2}\|\theta-\theta_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),\lambda_n,\bar{\xi}_n,\theta)\right\}$$

2. The *second half-step* (basic)

$$\theta_{n+1} = \arg\min_{\theta\in\mathcal{S}^{\mathfrak{n}}}\left\{\tfrac{1}{2}\|\theta-\theta_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{\lambda}_n,\bar{\xi}_n,\theta)\right\}$$

3.6 Convergence analysis

The following theorem presents the convergence conditions of (3.8 - 3.9) and gives the estimate of its rate of convergence for the L_p- Nash equilibrium and the strong L_p- Nash equilibrium. As well, we prove that the extraproximal method converges to an equilibrium point. Let us define the following extended vectors

$$\tilde{u}:=\begin{pmatrix} u\\ \hat{u}\\ \lambda\end{pmatrix}\in\tilde{U}:=U\times\hat{U}\times\mathbb{R}^+,\quad \tilde{z}:=\xi\in\tilde{Z}:=\mathbb{R}^+$$

Then, the regularized Lagrange function can be expressed as

$$\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}) := \mathcal{L}_\delta(u_\delta,\hat{u}_\delta,\lambda_\delta,\xi_\delta)$$

The equilibrium point that satisfies (3.7) can be expressed as

$$\tilde{u}^*_\delta = \arg\min_{\tilde{u}\in\tilde{U}}\left\{\tfrac{1}{2}\|\tilde{u}-\tilde{u}^*_\delta\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}^*_\delta)\right\}$$

$$\tilde{z}^*_\delta = \arg\max_{\tilde{z}\in\tilde{Z}}\left\{-\tfrac{1}{2}\|\tilde{z}-\tilde{z}^*_\delta\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{u}^*_\delta,\tilde{z})\right\}$$

Now, introducing the following variables

$$\tilde{w} = \begin{pmatrix}\tilde{w}_1\\ \tilde{w}_2\end{pmatrix} \in \tilde{U}\times\tilde{Z}, \quad \tilde{v} = \begin{pmatrix}\tilde{v}_1\\ \tilde{v}_2\end{pmatrix} \in \tilde{U}\times\tilde{Z}$$

and let define the Lagrangian in term of the previous variables

$$L_\delta(\tilde{w},\tilde{v}) := \tilde{\mathcal{L}}_\delta(\tilde{w}_1,\tilde{v}_2) - \tilde{\mathcal{L}}_\delta(\tilde{v}_1,\tilde{w}_2)$$

For $\tilde{w}_1=\tilde{u}$, $\tilde{w}_2=\tilde{z}$, $\tilde{v}_1=\tilde{v}^*_1=\tilde{u}^*_\delta$ and $\tilde{v}_2=\tilde{v}^*_2=\tilde{z}^*_\delta$ we have

$$L_\delta(\tilde{w},\tilde{v}^*) := \tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}^*_\delta) - \tilde{\mathcal{L}}_\delta(\tilde{u}^*_\delta,\tilde{z})$$

In these variables the relation (3.7) can be represented as follows

$$\tilde{v}^* = \arg\min_{\tilde{w}\in\tilde{U}\times\tilde{Z}}\left\{\tfrac{1}{2}\|\tilde{w}-\tilde{v}^*\|^2+\gamma L_\delta(\tilde{w},\tilde{v}^*)\right\} \tag{3.10}$$

Finally, we have that the extraproximal method can be expressed by

1. First step

$$\hat{v}_n = \arg\min_{\tilde{w}\in\tilde{U}\times\tilde{Z}}\left\{\tfrac{1}{2}\|\tilde{w}-\tilde{v}_n\|^2+\gamma L_\delta(\tilde{w},\tilde{v}_n)\right\} \tag{3.11}$$

2. Second step

$$\tilde{v}_{n+1} = \arg\min_{\tilde{w}\in\tilde{U}\times\tilde{Z}}\left\{\tfrac{1}{2}\|\tilde{w}-\tilde{v}_n\|^2+\gamma L_\delta(\tilde{w},\hat{v}_n)\right\} \tag{3.12}$$

Once the proximal and extraproximal method has been defined in terms of these new variables, we can follow the convergence theorems and proofs presented in Appendix C.

3.7 Numerical example

Consider a Nash game with 3 players. Let the number of states $N = 4$ and the actions $M = 2$ for each player. Then, the transition matrices for each player be defined as follows

$$\pi^1_{(i,j,1)=}\begin{bmatrix} 0.9535 & 0.0055 & 0.0120 & 0.0289 \\ 0.3971 & 0.2711 & 0.3097 & 0.0221 \\ 0.0778 & 0.0309 & 0.0431 & 0.8482 \\ 0.9398 & 0.0485 & 0.0095 & 0.0022 \end{bmatrix} \quad \pi^1_{(i,j,2)=}\begin{bmatrix} 0.2981 & 0.0405 & 0.3335 & 0.3279 \\ 0.3522 & 0.2399 & 0.1467 & 0.2612 \\ 0.0882 & 0.7087 & 0.1557 & 0.0474 \\ 0.2406 & 0.1503 & 0.3903 & 0.2189 \end{bmatrix}$$

$$\pi^2_{(i,j,1)=}\begin{bmatrix} 0.7543 & 0.2206 & 0.0162 & 0.0089 \\ 0.1495 & 0.2411 & 0.2547 & 0.3547 \\ 0.0076 & 0.1221 & 0.0527 & 0.8176 \\ 0.3773 & 0.3154 & 0.0372 & 0.2701 \end{bmatrix} \quad \pi^2_{(i,j,2)=}\begin{bmatrix} 0.2809 & 0.3226 & 0.3149 & 0.0815 \\ 0.9486 & 0.0091 & 0.0409 & 0.0014 \\ 0.0495 & 0.3111 & 0.2543 & 0.3851 \\ 0.2649 & 0.0297 & 0.2508 & 0.4547 \end{bmatrix}$$

$$\pi^3_{(i,j,1)=}\begin{bmatrix} 0.1368 & 0.3923 & 0.0525 & 0.4184 \\ 0.0598 & 0.3454 & 0.3366 & 0.2582 \\ 0.1450 & 0.3711 & 0.1858 & 0.2981 \\ 0.4365 & 0.4967 & 0.0350 & 0.0318 \end{bmatrix} \quad \pi^3_{(i,j,2)=}\begin{bmatrix} 0.1888 & 0.2770 & 0.2159 & 0.3184 \\ 0.2945 & 0.3463 & 0.1538 & 0.2054 \\ 0.0504 & 0.2463 & 0.3467 & 0.3566 \\ 0.3766 & 0.2250 & 0.2691 & 0.1292 \end{bmatrix}$$

The individual utility for each player is defined by

$$U^1_{(i,j,1)} = \begin{bmatrix} 7 & 17 & 3 & 55 \\ 1 & 10 & 6 & 7 \\ 0 & 16 & 17 & 4 \\ 0 & 15 & 9 & 1 \end{bmatrix} \quad U^2_{(i,j,1)=}\begin{bmatrix} 0 & 17 & 9 & 11 \\ 0 & 17 & 9 & 7 \\ 0 & 11 & 12 & 4 \\ 0 & 15 & 11 & 1 \end{bmatrix} \quad U^3_{(i,j,1)=}\begin{bmatrix} 5 & 8 & 7 & 8 \\ 1 & 4 & 3 & 5 \\ 4 & 6 & 17 & 1 \\ 12 & 0 & 7 & 3 \end{bmatrix}$$

$$U^1_{(i,j,2)} = \begin{bmatrix} 37 & 6 & 8 & 3 \\ 4 & 10 & 8 & 6 \\ 4 & 6 & 8 & 10 \\ 2 & 5 & 9 & 0 \end{bmatrix} \quad U^2_{(i,j,2)=}\begin{bmatrix} 10 & 18 & 80 & 9 \\ 4 & 5 & 1 & 3 \\ 4 & 6 & 7 & 0 \\ 12 & 7 & 8 & 0 \end{bmatrix} \quad U^3_{(i,j,2)=}\begin{bmatrix} 9 & 13 & 7 & 9 \\ 5 & 0 & 70 & 4 \\ 11 & 2 & 19 & 6 \\ 3 & 10 & 14 & 5 \end{bmatrix}$$

Applying the extraproximal method for Markov chains: **For** $\mathbf{L_2}-$**norm** the resulting strategies $c_{(i,k)}$ for each player (see Figures 3.1, 3.2 and 3.3) are as follows

$$c^1 = \begin{bmatrix} 0.1737 & 0.1963 \\ 0.1012 & 0.0856 \\ 0.0878 & 0.1330 \\ 0.0050 & 0.2174 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.4272 & 0.0050 \\ 0.1935 & 0.0050 \\ 0.0453 & 0.0838 \\ 0.0602 & 0.1801 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.0894 & 0.0856 \\ 0.3310 & 0.0050 \\ 0.1332 & 0.1055 \\ 0.1084 & 0.1419 \end{bmatrix}$$

As well Figure 3.4 shows the convergence of the parameter λ

$$\lambda^l = \begin{bmatrix} 0.3691 \\ 0.2843 \\ 0.3467 \end{bmatrix}$$

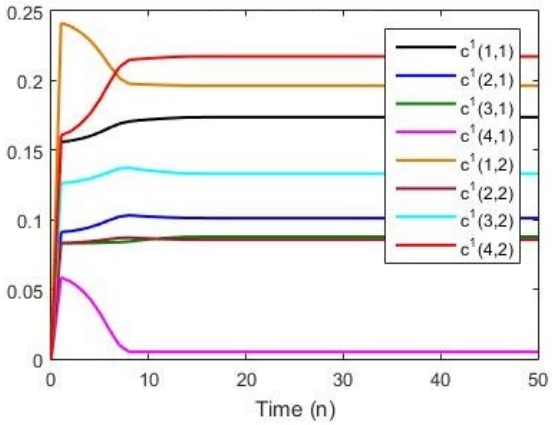


Figure 3.1 Strategies for Player 1, norm $p = 2$.

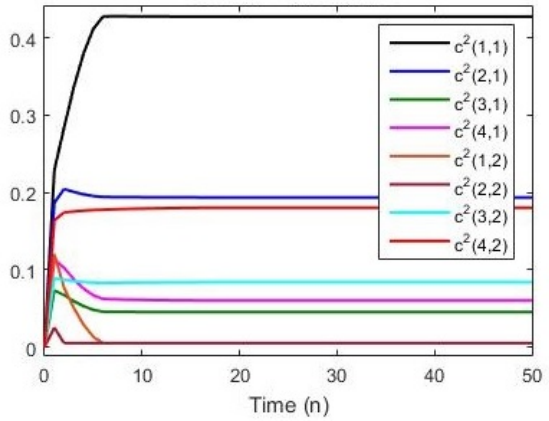


Figure 3.2 Strategies for Player 2, norm $p = 2$.

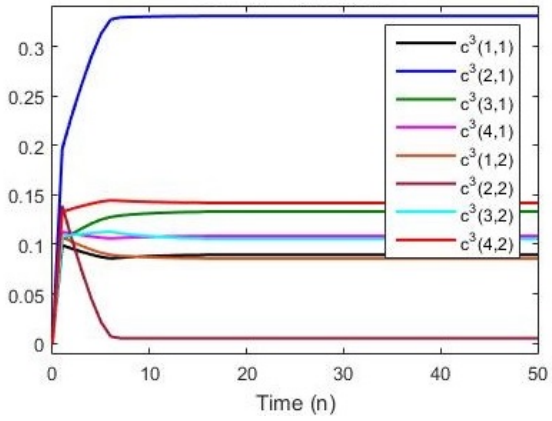


Figure 3.3 Strategies for Player 3, norm $p = 2$.

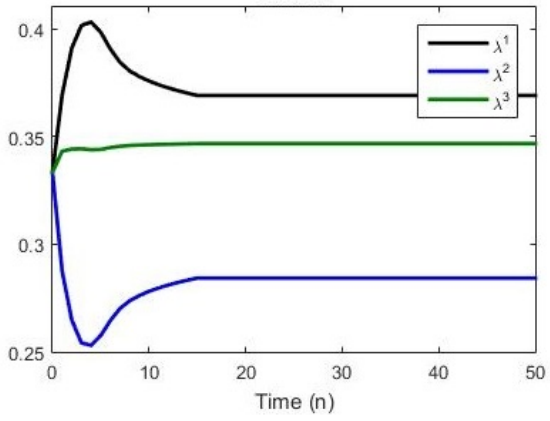


Figure 3.4 Convergence of λ^l, norm $p = 2$.

Then, applying (2.6) the strategies needed to converge to a strong Nash equilibrium are as follows:

$$d^1 = \begin{bmatrix} 0.4693 & 0.5307 \\ 0.5416 & 0.4584 \\ 0.3977 & 0.6023 \\ 0.0225 & 0.9775 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.9884 & 0.0116 \\ 0.9748 & 0.0252 \\ 0.3506 & 0.6494 \\ 0.2504 & 0.7496 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.5109 & 0.4891 \\ 0.9851 & 0.0149 \\ 0.5579 & 0.4421 \\ 0.4330 & 0.5670 \end{bmatrix}$$

Finally, the resulting individual utilities are as follows:

$$J^1 = 48.9213$$
$$J^2 = 22.1660$$
$$J^3 = 22.6872$$

For $\mathbf{L_\infty}-$norm the strategies $c_{(i,k)}$ for the players (see Figures 3.5, 3.6 and 3.7) are as follows:

$$c^1 = \begin{bmatrix} 0.1669 & 0.2082 \\ 0.1004 & 0.0863 \\ 0.0837 & 0.1350 \\ 0.0177 & 0.2018 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.4278 & 0.0050 \\ 0.1940 & 0.0050 \\ 0.0456 & 0.0828 \\ 0.0622 & 0.1775 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.0860 & 0.0887 \\ 0.3293 & 0.0050 \\ 0.1285 & 0.1120 \\ 0.1062 & 0.1443 \end{bmatrix}$$

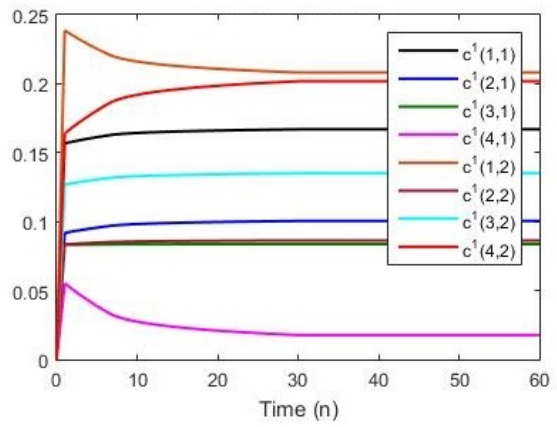


Figure 3.5 Strategies for Player 1, norm $p = \infty$.

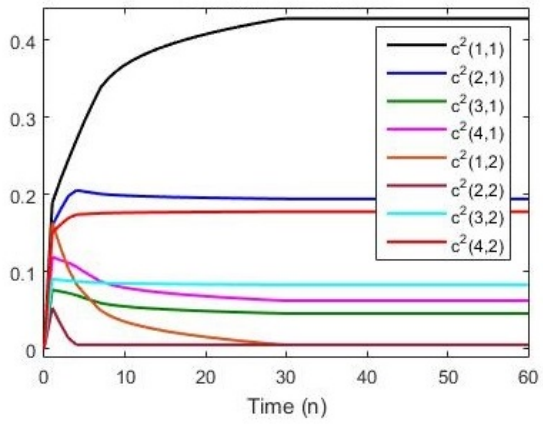


Figure 3.6 Strategies for Player 2, norm $p = \infty$.

As well Figures 3.8 and 3.9 show the convergence of the parameter λ and θ.

$$\lambda^l = \begin{bmatrix} 0.3981 \\ 0.2475 \\ 0.3544 \end{bmatrix} \quad \theta^l = \begin{bmatrix} 0.2869 \\ 0.3817 \\ 0.3314 \end{bmatrix}$$

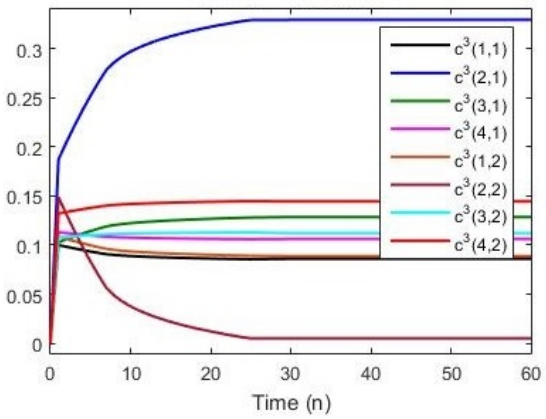


Figure 3.7 Strategies for Player 3, norm $p = \infty$.

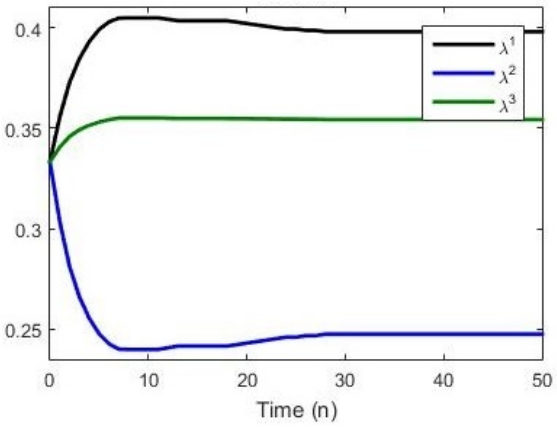


Figure 3.8 Convergence of λ, norm $p = \infty$.

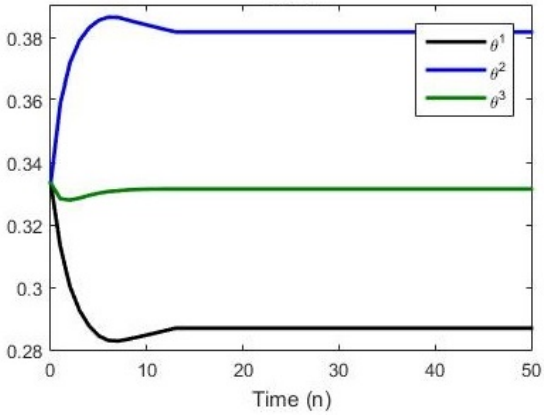


Figure 3.9 Convergence of θ, norm $p = \infty$.

Then, applying (2.6) the strategies needed to converge to a strong Nash equilibrium are as follows:

$$d^1 = \begin{bmatrix} 0.4450 & 0.5550 \\ 0.5378 & 0.4622 \\ 0.3827 & 0.6173 \\ 0.0805 & 0.9195 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.9884 & 0.0116 \\ 0.9749 & 0.0251 \\ 0.3553 & 0.6447 \\ 0.2595 & 0.7405 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.4921 & 0.5079 \\ 0.9850 & 0.0150 \\ 0.5344 & 0.4656 \\ 0.4239 & 0.5761 \end{bmatrix}$$

Then, the resulting individual utilities are as follows:

$$J^1 = 54.2461, \quad J^2 = 19.3209, \quad J^3 = 23.4588$$

Remark 3.13 *In order to demonstrate the effectiveness of the solution we prove that* $G_{L_\infty} = 4.1598 < G_{L_2} = 5.0202$

Chapter 4

The Strong L_p-Stackelberg game

4.1 Introduction

The notion of collaboration implies that related players interact with each other looking for cooperative stability. This notion consents players to select optimal strategies and to condition their own behavior on the behavior of others in a strategic forward-looking manner. This chapter examines the formation of coalitions within a class of hierarchical problems called Stackelberg games [89]. The complexity analysis of the Stackelberg equilibrium plays a central role in game theory and it has been analyzed to determine whether the concept is reasonable from a computational point of view.

Stackelberg games are usually represented by a leader-follower problem which corresponds to a bi-level programming problem. In bi-level programming problems there are two competing decision-making parties [10]: a) one is upper-level decision makers and, b) the other is lower-level decision makers. The two levels interact with each other as follows. The lower-level is completely restricted by the upper-level's decision and for each decision made by the upper-level, the lower-level will choose the best option according to their objectives. Instead, the upper-level objectives are restricted from below by the lower-level: the upper-level controls the lower-level's decision in the way that the lower-level will react by choosing the best option.

In a Stackelberg game, the leader's optimization problem is represented by the upper-level, restricted by the follower's optimization mission at the lower-level. The dynamics of a Stackelberg game is as follows: the leader considers the best-reply of the follower. Then, he/she

commits to a mixed strategy (a probability distribution over deterministic schedules) that minimizes the cost, anticipating the predicted best-reply of the follower. Then, taking into the account the adversary's mixed strategy selection, the follower in equilibrium selects the expected best-reply that minimizes the cost (maximizes the utility).

Bi-level programming models have vast theoretical studies and applications in the real world. The traditional methods employed to solve these problems include penalty functions [1], the Karush-Kuhn-Tucker method [13, 37] and branch-and-bound procedures [11]. Applications were presented into the security domain by ([20, 86]) suggesting an upper-level that represents defenders trying to minimize risk and a lower-level that represents attackers trying maximizing destruction for a given target. There are several applications implemented in different areas: transportation, agriculture, network, management.

This chapter presents an approach for computing the strong Stackelberg/Nash equilibrium for Markov chains games [100, 97]. The cooperative $\mathfrak{n}$-leaders and $\mathfrak{m}$-followers Markov game is solved considering the minimization of the L_p-norm. The existence of the L_p-Stackelberg/Nash equilibrium is characterized as a strong Pareto policy, which is the closest in the Euclidean norm to the virtual minimum (utopia point). Then, the optimization problem is reduced to find a Pareto optimal solution. A bi-level programming model implemented by the extraproximal optimization approach is designed for computing the static strong Stackelberg/Nash equilibrium. We design the method for the static strong Stackelberg/Nash game in terms of nonlinear programming problems implementing the regularized Lagrange principle to ensure the convergence of the cost-functions to a unique strong L_p-Stackelberg/Nash equilibrium. We formulate the nonlinear programming problem considering several linear constraints employing the c-variable method. The proposed method approaches in exponential time to a strong L_p-Stackelberg/Nash equilibrium. The usefulness of the proposed solution is proved theoretically and by an application example related to the effectiveness of relationship marketing strategies within the department store sector of the retail industry (supermarkets).

4.2 The strong Stackelberg/Nash game

Let us introduce the variables

$$v^m := \text{col}\,(c^m),\ V^m := C^m_{\text{adm}}\ V := \bigotimes_{m=1}^{\mathfrak{m}} V^m$$

for $m = \overline{1,\mathfrak{m}}$, where col is the column operator. Consider a Stackelberg game with $\mathcal{N} = \{1,...,\mathfrak{n}\}$ leaders whose strategies are denoted by $u^l \in U^l\ \left(l = \overline{1,\mathfrak{n}}\right)$ where U is a convex and compact set. Denote by $u = (u^1,...,u^{\mathfrak{n}})^\top \in U$ the joint strategy of the players and $u^{\hat{l}}$ is a strategy of the rest of the leaders adjoint to u^l, namely,

$$u^{\hat{l}} := \left(u^1,...,u^{l-1},u^{l+1},...,u^{\mathfrak{n}}\right)^\top \in U^{\hat{l}} := \bigotimes_{h=1,\ h\neq l}^{\mathfrak{n}} U^h$$

such that $u = (u^l, u^{\hat{l}})\ \left(l = \overline{1,\mathfrak{n}}\right)$. As well, consider $\mathcal{M} = \{1,...,\mathfrak{m}\}$ followers with strategies $v^m \in V^m\ \left(m = \overline{1,\mathfrak{m}}\right)$ and V is also a convex and compact set. Denote by $v = (v^1,...,v^{\mathfrak{m}}) \in V$ the joint strategy of the followers and $v^{\hat{m}}$ is a strategy of the rest of the followers adjoint to v^m, namely,

$$v^{\hat{m}} := \left(v^1,...,v^{m-1},v^{m+1},...,v^{\mathfrak{m}}\right)^\top \in V^{\hat{m}} := \bigotimes_{q=1,\ q\neq m}^{\mathfrak{m}} V^q$$

such that $v = (v^m, v^{\hat{m}})\ \left(m = \overline{1,\mathfrak{m}}\right)$.

4.2.1 The strong Nash equilibria

Following the concepts presented in Chapter 3. In the dynamics of the game leaders play cooperatively and they are assumed to anticipate the reactions of the followers trying to reach the strong Nash equilibria. For reaching the goal of the game leaders first try to find a joint strategy $u^* = (u^{1*},...,u^{\mathfrak{n}*}) \in U$ satisfying for any admissible $u^l \in U^l$ and any $l = \overline{1,\mathfrak{n}}$

$$G_{L_p}(u,\hat{u}(u)) := \left(\sum_{l=1}^{\mathfrak{n}} \lambda^l \left|\varphi_l\left(\bar{u}^l, u^{\hat{l}}\right) - \varphi_l\left(u^l, u^{\hat{l}}\right)\right|^p\right)^{1/p}$$

where $\hat{u}(u) = (u^{\hat{1}\top},...,u^{\hat{\mathfrak{n}}\top})^\top \in \hat{U} \subseteq \mathbb{R}^{\mathfrak{n}(\mathfrak{n}-1)}$ and $\bar{u}^l$ is the utopia point (3.2). Here $\varphi_l\left(u^l, u^{\hat{l}}\right)$ is the cost-function of the leader l which plays the strategy $u^l \in U^l$ and the rest of the leaders play the strategy $u^{\hat{l}} \in U^{\hat{l}}$, these functions are assumed to be convex in all their arguments.

Condition 4.1 *The function $G_{L_p}(u, \hat{u}(u))$ satisfies the Nash condition*

$$g(u, \hat{u}(u)) = \sum_{l=1}^{\mathfrak{n}} \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}} \right) - \varphi_l \left(u^l, u^{\hat{l}} \right) \right] \leq 0$$

for any $u^l \in U^l$ and all $l = \overline{1, \mathfrak{n}}$.

As well, in this process the followers try to reach one of the strong Nash equilibria trying to find a joint strategy $v^* = (v^{1*}, ..., v^{\mathfrak{m}*}) \in V$ satisfying for any admissible $v^m \in V^m$ and any $m = \overline{1, \mathfrak{m}}$

$$F_{L_p}(v, \hat{v}(v)) := \left(\sum_{m=1}^{\mathfrak{m}} \theta^m \left| \psi_m \left(\bar{v}^m, v^{\hat{m}} \right) - \psi_m \left(v^m, v^{\hat{m}} \right) \right|^p \right)^{1/p}$$

where $\hat{v}(v) = (v^{\hat{1}\top}, ..., v^{\hat{\mathfrak{m}}\top})^\top \in \hat{V} \subseteq \mathbb{R}^{\mathfrak{m}(\mathfrak{m}-1)}$ and $\bar{v}^m$ is defined as the utopia point (3.2). Here $\psi_m \left(v^m, v^{\hat{m}} \right)$ is the cost-function of the follower m which plays the strategy $v^m \in V^m$ and the rest of the followers play the strategy $v^{\hat{m}} \in V^{\hat{m}}$, these functions are assumed to be convex in all their arguments.

Condition 4.2 *The function $F_{L_p}(v, \hat{v}(v))$ satisfies the Nash condition*

$$f(v, \hat{v}(v)) = \sum_{m=1}^{\mathfrak{m}} \left[\psi_m \left(\bar{v}^m, v^{\hat{m}} \right) - \psi_m \left(v^m, v^{\hat{m}} \right) \right] \leq 0$$

for any $v^m \in V^m$ and all $m = \overline{1, \mathfrak{m}}$.

4.2.2 The Stackelberg game

Leaders and followers together are in a Stackelberg game: the model involves two cooperative Nash games restricted by a Stackelberg game defined as follows.

Definition 4.3 *A game with $\mathfrak{n}$ leaders and $\mathfrak{m}$ followers said to be a cooperatively Stackelberg-Nash game if*

$$G_{L_p}(u(\lambda), \hat{u}(u, \lambda)|v) := \left(\sum_{l=1}^{\mathfrak{n}} \lambda^l \left| \varphi_l \left(\bar{u}^l, u^{\hat{l}}|v \right) - \varphi_l \left(u^l, u^{\hat{l}}|v \right) \right|^p \right)^{1/p}$$

given $\lambda \in \mathcal{S}^{\mathfrak{n}}$ such that

$$g(u, \hat{u}(u)|v) = \sum_{l=1}^{\mathcal{N}} \left[\varphi_l \left(\bar{u}^l, u^{\hat{l}}|v \right) - \varphi_l \left(u^l, u^{\hat{l}}|v \right) \right] \leq 0$$

and for the followers,

$$f_{L_p}\left(v(\theta),\hat{v}(v,\theta)|u\right):=\left(\sum_{m=1}^{\mathfrak{m}}\theta^m\left|\psi_m\left(\bar{v}^m,v^{\hat{m}}|u\right)-\psi_m\left(v^m,v^{\hat{m}}|u\right)\right|^p\right)^{1/p} \tag{4.1}$$

where $\theta\in\mathcal{S}^{\mathfrak{m}}$.

Remark 4.4 *In the case of the bi-level approach introduced in Definition (4.3) we employ the restriction* $f_{L_p}\left(v(\theta),\hat{v}(v,\theta)|u\right)$ *in (4.1) for ensuring the followers play cooperatively.*

Definition 4.5 *Let* $G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)|v\right)$ *be the cost functions of the leaders* $\left(l=\overline{1,\mathfrak{n}}\right)$. *A strategy* $u^*\in U$ *of the leaders together with the collection* $v^*\in V$ *of the followers is said to be a cooperatively Stackelberg-Nash equilibrium if*

$$(u^*,v^*)\in \mathit{Arg}\min_{u\in U}\min_{\hat{u}(u)\in\hat{U}}\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\max_{v\in V}\max_{\hat{v}(v)\in\hat{V}}\max_{\theta\in\mathcal{S}^{\mathfrak{m}}}$$

$$\left\{G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)|v\right)\ |\ g\left(u,\hat{u}(u)|v\right)\leq 0,\ f_{L_p}\left(v(\lambda),\hat{v}(v,\lambda)|u\right)\leq 0\right\}$$

Remark 4.6 *If* $G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)|v\right)$ *is strictly convex then*

$$(u^*,v^*)=\arg\min_{u\in U}\min_{\hat{u}(u)\in\hat{U}}\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\max_{v\in V}\max_{\hat{v}(v)\in\hat{V}}\max_{\theta\in\mathcal{S}^{\mathfrak{m}}}$$

$$\left\{G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)|v\right)\ |\ g\left(u,\hat{u}(u)|v\right)\leq 0,\ f_{L_p}\left(v(\theta),\hat{v}(v,\theta)|u\right)\leq 0\right\}$$

Applying the Lagrange principle we may conclude that Definition 4.5 can be rewritten as

$$(u^*,v^*)\in \mathrm{Arg}\min_{u\in U}\min_{\hat{u}(u)\in\hat{U}}\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\max_{v\in V}\max_{\hat{v}(v)\in\hat{V}}\max_{\theta\in\mathcal{S}^{\mathfrak{m}}}\max_{\omega\geq 0}\max_{\xi\geq 0}\mathcal{L}(u,\hat{u}(u),v,\hat{v}(v),\lambda,\theta,\omega,\xi)$$

where

$$\mathcal{L}(u,\hat{u}(u),v,\hat{v}(v),\lambda,\theta,\omega,\xi):=$$

$$G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)\mid v\right)+\omega g\left(u,\hat{u}(u)\mid v\right)+\xi f_{L_p}\left(v(\theta),\hat{v}(v,\theta)\mid u\right)$$

The approximative solution obtained by the Tikhonov's regularization is given by

$$(u^*,v^*)=\arg\min_{u\in U}\min_{\hat{u}(u)\in\hat{U}}\min_{\lambda\in\mathcal{S}^{\mathfrak{n}}}\max_{v\in V}\max_{\hat{v}(v)\in\hat{V}}\max_{\theta\in\mathcal{S}^{\mathfrak{m}}}\max_{\omega\geq 0}\max_{\xi\geq 0}\mathcal{L}_\delta(u,\hat{u}(u),v,\hat{v}(v),\lambda,\theta,\omega,\xi)$$

such that

$$\mathcal{L}_\delta(u,\hat{u}(u),v,\hat{v}(v),\lambda,\theta,\omega,\xi) := G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\mid v\right)+\omega g_\delta\left(u,\hat{u}(u)\mid v\right)+ \\ \xi f_{L_p,\delta}\left(v(\theta),\hat{v}(v,\theta)\mid u\right)-\tfrac{\delta}{2}(\omega^2+\xi^2) \tag{4.2}$$

where

$$G_{L_p,\delta}\left(u(\lambda),\hat{u}(u,\lambda)\mid v\right)=\left(\sum_{l=1}^{\mathfrak{n}}\lambda^l\left|\varphi_l\left(\bar{u}^l,u^{\hat{l}}\mid v\right)-\varphi_l\left(u^l,u^{\hat{l}}\mid v\right)\right|^p\right)^{1/p}+ \\ \tfrac{\delta}{2}\left(\|u\|^2+\|\hat{u}(u)\|^2+\|\lambda\|^2\right)$$

$$g_\delta\left(u,\hat{u}(u)\mid v\right)=\sum_{l=1}^{\mathfrak{n}}\left[\varphi_l\left(\bar{u}^l,u^{\hat{l}}\mid v\right)-\varphi_l\left(u^l,u^{\hat{l}}\mid v\right)\right]+\tfrac{\delta}{2}\left(\|u\|^2+\|\hat{u}(u)\|^2\right)$$

$$f_{L_p,\delta}\left(v(\theta),\hat{v}(v,\theta)\mid u\right)=\left(\sum_{m=1}^{\mathfrak{m}}\theta^m\left|\psi_m\left(\bar{v}^m,v^{\hat{m}}\mid u\right)-\psi_m\left(v^m,v^{\hat{m}}\mid u\right)\right|^p\right)^{1/p}+ \\ \tfrac{\delta}{2}\left(\|v\|^2+\|\hat{v}(v)\|^2+\|\theta\|^2\right)$$

Now, the function $G_\delta\left(u,\hat{u}(u)\mid v\right)$ is strictly convex if the Hessian matrix is positive semi-definite, then $G_\delta\left(u,\hat{u}(u)'\mid v\right)$ attains a minimum at $\left(u,\hat{u}(u)\mid v\right)$ if

$$\nabla^2 G_\delta\left(u,\hat{u}(u)|v\right)=\begin{bmatrix}\delta I_{n_1\times n_1} & \mathcal{DG}_{1,2}(\hat{u}_{1,2}) & \dots & \mathcal{DG}_{1,\mathfrak{n}}(\hat{u}_{1,\mathfrak{n}})\\ \mathcal{DG}_{2,1}(\hat{u}_{2,1}) & \delta I_{n_2\times n_2} & \dots & \mathcal{DG}_{3,2}(\hat{u}_{3,2})\\ \dots & \dots & \dots & \dots\\ \mathcal{DG}_{3,1}(\hat{u}_{3,1}) & \mathcal{DG}_{3,2}(\hat{u}_{3,2}) & \dots & \delta I_{n_{\mathfrak{n}}\times n_{\mathfrak{n}}}\end{bmatrix}>0$$

or, equivalently, δ should provide the inequality

$$\min_{u\in U,\hat{u}\in\hat{U}}\left[\Lambda_{\min}\left(\nabla^2 G_\delta\left(u,\hat{u}(u)|v\right)\right)\right]>0$$

Here, $\hat{u}_{ik}$ is independent of $u^{(i)}$ and $u^{(k)}$, that is, $\frac{\partial}{\partial u^{(i)}}\hat{u}_{ik}=0$ and $\frac{\partial}{\partial u^{(k)}}\hat{u}_{ik}=0$. As well as, the function $f_\delta\left(v,\hat{v}(v)|u\right)$ is strictly concave if the Hessian matrix is negative semi-definite, then $f_\delta\left(v,\hat{v}(v)|u\right)$ attains a maximum at $\left(v,\hat{v}(v)|u\right)$ if

$$\max_{v\in V,\hat{v}\in\hat{V}}\left[\Lambda_{\max}\left(\nabla^2 f_\delta\left(v,\hat{v}(v)|u\right)\right)\right]<0$$

With sufficiently large δ, the considered functions provide the uniqueness of the conditional optimization problem (4.2). Notice also that the Lagrange function (4.2) satisfies the saddle-point condition, namely, for all $u\in U$, $\hat{u}\in\hat{U}$, $v\in V$, $\hat{v}(v)\in\hat{V}$, $\lambda\in\mathcal{S}^{\mathfrak{n}}$, $\theta\in\mathcal{S}^{\mathfrak{m}}$, $\omega\geq 0$ and

$\xi \geq 0$ we have

$$\mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta, \hat{v}_\delta(v), \lambda_\delta^*, \theta_\delta, \omega_\delta, \xi_\delta) \leq \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*) \leq$$

$$\mathcal{L}_\delta(u_\delta, \hat{u}_\delta(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*)$$

4.3 The proximal format

In the proximal format the relation (4.2) can be expressed as

$$\begin{aligned}
\omega_\delta^* &= \arg\max_{\omega \geq 0} \left\{ -\tfrac{1}{2}\|\omega - \omega_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta, \xi_\delta^*) \right\} \\
\xi_\delta^* &= \arg\max_{\xi \geq 0} \left\{ -\tfrac{1}{2}\|\xi - \xi_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta) \right\} \\
u_\delta^* &= \arg\min_{u \in U} \left\{ \tfrac{1}{2}\|u - u_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*) \right\} \\
\hat{u}_\delta^* &= \arg\min_{\hat{u} \in \hat{U}} \left\{ \tfrac{1}{2}\|\hat{u} - \hat{u}_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*) \right\} \\
v_\delta^* &= \arg\max_{v \in V} \left\{ -\tfrac{1}{2}\|v - v_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*) \right\} \\
\hat{v}_\delta^* &= \arg\max_{\hat{v} \in \hat{V}} \left\{ -\tfrac{1}{2}\|\hat{v} - \hat{v}_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*, \xi_\delta^*) \right\} \\
\lambda_\delta^* &= \arg\min_{\lambda \in \mathcal{S}^N} \left\{ \tfrac{1}{2}\|\lambda - \lambda_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta, \theta_\delta^*, \omega_\delta^*, \xi_\delta^* \right\} \\
\theta_\delta^* &= \arg\max_{\theta \in \mathcal{S}^N} \left\{ -\tfrac{1}{2}\|\theta - \theta_\delta^*\|^2 + \gamma \mathcal{L}_\delta(u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta, \omega_\delta^*, \xi_\delta^* \right\}
\end{aligned} \tag{4.3}$$

where the solutions $u_\delta^*, \hat{u}_\delta^*(u), v_\delta^*, \hat{v}_\delta^*(v), \lambda_\delta^*, \theta_\delta^*, \omega_\delta^*$ and ξ_δ^* depend on the parameters $\delta, \gamma > 0$.

4.4 The Extraproximal method

We design the extraproximal method for the static Stackelberg-Nash game in a general format iterative version $(n = 0, 1, ...)$ with some fixed admissible initial values ($u_0 \in U$, $\hat{u}_0 \in U$, $v_0 \in V$, $\hat{v}_0 \in \hat{V}$, $\omega_0 \geq 0$, $\xi_0 \geq 0$, $\lambda_0 \in \mathcal{S}^{\mathrm{n}}$ and $\theta_0 \in \mathcal{S}^{\mathrm{m}}$) as follows:

1. The *first half-step* (prediction):

$$\begin{aligned}
\bar{\omega}_n &= \arg\min_{\omega\geq 0}\left\{\tfrac{1}{2}\|\omega-\omega_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v_n,\hat{v}_n(v),\lambda_n,\theta_n,\omega,\bar{\xi}_n)\right\}\\
\bar{\xi}_n &= \arg\min_{\xi\geq 0}\left\{\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v_n,\hat{v}_n(v),\lambda_n,\theta_n,\bar{\omega}_n,\xi)\right\}\\
\bar{u}_n &= \arg\min_{u\in U}\left\{\tfrac{1}{2}\|u-u_n\|^2+\gamma\mathcal{L}_\delta(u,\hat{u}_n(u),v_n,\hat{v}_n(v),\lambda_n,\theta_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\overline{\hat{u}}_n &= \arg\min_{\hat{u}\in\hat{U}}\left\{\tfrac{1}{2}\|\hat{u}-\hat{u}_n\|^2+\gamma\mathcal{L}_\delta(u_n,\hat{u}(u),v_n,\hat{v}_n(v),\lambda_n,\theta_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\bar{v}_n &= \arg\min_{v\in V}\left\{\tfrac{1}{2}\|v-v_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v,\hat{v}_n(v),\lambda_n,\theta_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\overline{\hat{v}}_n &= \arg\min_{\hat{v}\in\hat{V}}\left\{\tfrac{1}{2}\|\hat{v}-\hat{v}_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v_n,\hat{v}(v),\lambda_n,\theta_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\bar{\lambda}_n &= \arg\min_{\lambda\in\mathcal{S}^N}\left\{\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v_n,\hat{v}_n(v),\lambda,\theta_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\bar{\theta}_n &= \arg\min_{\theta\in\mathcal{S}^N}\left\{\tfrac{1}{2}\|\theta-\theta_n\|^2-\gamma\mathcal{L}_\delta(u_n,\hat{u}_n(u),v_n,\hat{v}_n(v),\lambda_n,\theta,\bar{\omega}_n,\bar{\xi}_n)\right\}
\end{aligned}\tag{4.4}$$

2. The *second half-step* (basic):

$$\begin{aligned}
\omega_{n+1} &= \arg\min_{\omega\geq 0}\left\{\tfrac{1}{2}\|\omega-\omega_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{v}_n,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\bar{\theta}_n,\omega,\bar{\xi}_n)\right\}\\
\xi_{n+1} &= \arg\min_{\xi\geq 0}\left\{\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{v}_n,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\bar{\theta}_n,\bar{\omega}_n,\xi)\right\}\\
u_{n+1} &= \arg\min_{u\in U}\left\{\tfrac{1}{2}\|u-u_n\|^2+\gamma\mathcal{L}_\delta(u,\overline{\hat{u}}_n(u),\bar{v}_n,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\bar{\theta}_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\hat{u}_{n+1} &= \arg\min_{\hat{u}\in\hat{U}}\left\{\tfrac{1}{2}\|\hat{u}-\hat{u}_n\|^2+\gamma\mathcal{L}_\delta(\bar{u}_n,\hat{u}(u),\bar{v}_n,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\bar{\theta}_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
v_{n+1} &= \arg\min_{v\in V}\left\{\tfrac{1}{2}\|v-v_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),v,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\bar{\theta}_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\hat{v}_{n+1} &= \arg\min_{\hat{v}\in\hat{V}}\left\{\tfrac{1}{2}\|\hat{v}-\hat{v}_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{v}_n,\hat{v}(v),\bar{\lambda}_n,\bar{\theta}_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\lambda_{n+1} &= \arg\min_{\lambda\in\mathcal{S}^N}\left\{\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{v}_n,\overline{\hat{v}}_n(v),\lambda,\bar{\theta}_n,\bar{\omega}_n,\bar{\xi}_n)\right\}\\
\theta_{n+1} &= \arg\min_{\theta\in\mathcal{S}^N}\left\{\tfrac{1}{2}\|\theta-\theta_n\|^2-\gamma\mathcal{L}_\delta(\bar{u}_n,\overline{\hat{u}}_n(u),\bar{v}_n,\overline{\hat{v}}_n(v),\bar{\lambda}_n,\theta,\bar{\omega}_n,\bar{\xi}_n)\right\}
\end{aligned}\tag{4.5}$$

4.5 Convergence analysis

The following theorem presents the convergence conditions of (4.4) - (4.5) and gives the estimate of its rate of convergence for the strong L_p- Stackelberg/Nash equilibrium. As well, it is proved that the extraproximal method converges to an equilibrium point.

Let us define the following extended vectors

$$\tilde{u}:=\begin{pmatrix} u \\ \hat{u} \\ \lambda \end{pmatrix} \in \tilde{U}:=U\times\hat{U}\times\mathbb{R}^+,\ \tilde{z}:=\begin{pmatrix} v \\ \hat{v} \\ \theta \\ \xi \\ \omega \end{pmatrix} \in \tilde{Z}:=V\times\hat{V}\times\mathbb{R}^+\times\mathbb{R}^+\times\mathbb{R}^+$$

Then, the regularized Lagrange function can be expressed as

$$\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}) := \mathcal{L}_\delta(u_\delta,\hat{u}_\delta,v_\delta,\hat{v}_\delta,\lambda_\delta,\theta_\delta,\xi_\delta,\omega_\delta)$$

The equilibrium point that satisfies (4.3) can be expressed as

$$\tilde{u}^*_\delta = \arg\min_{\tilde{u}\in\tilde{U}}\left\{\tfrac{1}{2}\|\tilde{u}-\tilde{u}^*_\delta\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}^*_\delta)\right\}$$

$$\tilde{z}^*_\delta = \arg\max_{\tilde{z}\in\tilde{Z}}\left\{-\tfrac{1}{2}\|\tilde{z}-\tilde{z}^*_\delta\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{u}^*_\delta,\tilde{z})\right\}$$

Now, introducing the following variables

$$\tilde{w}=\begin{pmatrix}\tilde{w}_1\\ \tilde{w}_2\end{pmatrix}\in\tilde{U}\times\tilde{Z},\quad \tilde{v}=\begin{pmatrix}\tilde{v}_1\\ \tilde{v}_2\end{pmatrix}\in\tilde{U}\times\tilde{Z}$$

and let define the Lagrangian in term of the previous variables

$$L_\delta(\tilde{w},\tilde{v}) := \tilde{\mathcal{L}}_\delta(\tilde{w}_1,\tilde{v}_2)-\tilde{\mathcal{L}}_\delta(\tilde{v}_1,\tilde{w}_2)$$

For $\tilde{w}_1=\tilde{u}$, $\tilde{w}_2=\tilde{z}$, $\tilde{v}_1=\tilde{v}^*_1=\tilde{u}^*_\delta$ and $\tilde{v}_2=\tilde{v}^*_2=\tilde{z}^*_\delta$ we have

$$L_\delta(\tilde{w},\tilde{v}^*) := \tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z}^*_\delta)-\tilde{\mathcal{L}}_\delta(\tilde{u}^*_\delta,\tilde{z})$$

In these variables the relation (4.3) can be represented by

$$\tilde{v}^* = \arg\min_{\tilde{w}\in\tilde{U}\times\tilde{Z}}\left\{\tfrac{1}{2}\|\tilde{w}-\tilde{v}^*\|^2+\gamma L_\delta(\tilde{w},\tilde{v}^*)\right\} \tag{4.6}$$

Finally, we have that the extraproximal method can be expressed by

1. First step

$$\hat{v}_n = \arg \min_{\tilde{w} \in \tilde{U} \times \tilde{Z}} \left\{ \tfrac{1}{2} \|\tilde{w} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{w}, \tilde{v}_n) \right\} \tag{4.7}$$

2. Second step

$$\tilde{v}_{n+1} = \arg \min_{\tilde{w} \in \tilde{U} \times \tilde{Z}} \left\{ \tfrac{1}{2} \|\tilde{w} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{w}, \hat{v}_n) \right\} \tag{4.8}$$

Theorem 4.7 (Convergence and Rate of Convergence) *Let $\tilde{\mathcal{L}}_\delta(\tilde{u}, \tilde{z})$ be differentiable in $\tilde{u}$ and $\tilde{z}$, whose partial derivative with respect to $\tilde{z}$ satisfies the Lipschitz condition with positive constant C. Then, for some δ and*

$$C_0^l = \sum_{l=1}^{\mathfrak{n}} C_{0,l} \ \leq \ \mathfrak{n} \max_{l=\overline{1,\mathfrak{n}}} C_{0,l} \ = \ \mathfrak{n} C_0^{l+}$$

and

$$C_0^m = \sum_{m=1}^{\mathfrak{m}} C_{0,m} \ \leq \ \mathfrak{m} \max_{m=\overline{1,\mathfrak{m}}} C_{0,m} \ = \ \mathfrak{m} C_0^{m+}$$

there exists a small-enough

$$\gamma_0 = \gamma_0(\delta) < C :=$$

$$\max \left[\min \left\{ \frac{1}{\sqrt{2} C_0^{l+} \mathfrak{n}}, \frac{1+\sqrt{1+2\left(C_0^{l+}\right)^2}}{2\left(C_0^{l+}\right)^2 \mathfrak{n}} \right\}, \min \left\{ \frac{1}{\sqrt{2} C_0^{m+} \mathfrak{m}}, \frac{1+\sqrt{1+2\left(C_0^{m+}\right)^2}}{2\left(C_0^{m+}\right)^2 \mathfrak{m}} \right\} \right]$$

such that, for any $0 < \gamma \leq \gamma_0$, sequence $\{\tilde{v}_n\}$, which generated by the equivalent extraproximal procedure (4.7) - (4.8), monotonically converges with exponential rate $q \in (0, 1)$ to a unique equilibrium point $\tilde{v}^*$, i.e.,

$$\|\tilde{v}_n - \tilde{v}^*\|^2 \leq e^{n \ln q} \|\tilde{v}_0 - \tilde{v}^*\|^2$$

where

$$q = 1 + \frac{4(\delta\gamma)^2}{1 + 2\delta\gamma - 2\gamma^2 C^2} - 2\delta\gamma < 1$$

and $q_{\min}$ is given by

$$q_{\min} = 1 - \frac{2\delta\gamma}{1 + 2\delta\gamma} = \frac{1}{1 + 2\delta\gamma}.$$

For the proof, follow the convergence theorems and proofs presented in Appendix C.

4.6 Application examples

4.6.1 The pursuit problem

The pursuit evasion problem is among the oldest and most elegant problems in game theory. In our case, the game involves two pursuers whose goal is to capture two evaders, whose goal is to avoid capture. Capture is occurring when a pursuer occupies the same position as a prey. The following are fixed assumptions:

1. There are two evaders ($m = 1, 2$) and two pursuers ($l = 1, 2$).
2. For each time $n \in \mathbb{N}$, the evaders (and pursuers) jump from state a $s^m(n)$ to $s^m(n+1)$ ($s^l(n)$ to $s^l(n+1)$), a point within the Manhattan distance (just adjacent points are allowed).
3. The pursuers win the game if they occupy the same position as preys.
4. When a pursuer captures a prey, the pursuer continues in the game.
5. Each pursuer has no information about the current position of the evader, however strategically evaders and pursuers cooperate with themselves.

The principal result of the realization is to show that the pursuers' strategies win the game, regardless of evader strategy. Our choice for selecting a strategy is given by the Max Entropy [75] $H = d^{l*}_{(k|i)} \log d^{l*}_{(k|i)}$ between the computed distribution $d^{l}_{(k|i)}$ and the optimal distribution $d^{l*}_{(k|i)}$. This approach can be expressed as

$$d^{l*}_{(k^*|i)} = \delta_{(k^*(i),i)}$$

where $\delta_{(k^*(i),i)}$ is the Kronecker symbol, $k^*(i)$ is an index for which

$$k^*(i) = \max_{k \in M} H$$

Similarly, we have a starting point for the evaders with a stationary Markov policy given by equation (2.6) and a Max Entropy given by $H = d^{m*}_{(k|i)} \log d^{m*}_{(k|i)}$.

The capture condition at time n is determined by the fact that a pursuer and a evader are located at the same state and can be formalized mathematically as follows:

$$\sum_{j=1}^{N}\left[\chi(\alpha : s^l(n) = s_{(j)} \wedge s^m(n) = s_{(j)})\right] = \sum_{j=1}^{N}\left[\chi(\alpha : s^l(n) = s_{(j)})\ \chi(\alpha : s^m(n) = s_{(j)})\right]$$

where $\alpha \in \Omega$ is a trajectory. Now, the capture event of all the attackers is given by

$$\sum_{l=1}^{\mathfrak{n}}\sum_{m=1}^{\mathfrak{m}}\sum_{j=1}^{N}\left[\chi(\alpha : s^l(n) = s_{(j)})\ \chi(\alpha : s^m(n) = s_{(j)})\right]$$

A fixed Markov transition matrix $\pi_{(j|i,k)}$ is given. Then, the state transitions induced by the strategy $d^*_{(k^*|i)}$ are governed by the conditional probability law for pursuers and evaders as follows

$$\Pi^{l*}_{(j|i)}(d) = \sum_{k=1}^{M}\pi^{l}_{(j|i,k)}d^{l*}_{(k^*|i)}, \qquad \Pi^{m*}_{(j|i)}(d) = \sum_{k=1}^{M}\pi^{m}_{(j|i,k)}d^{m*}_{(k^*|i)}$$

Let $N = 4$, $M = 2$. The individual utility for each player are defined by

$$U^1_{(i,j,1)} = \begin{bmatrix} 70 & 17 & 30 & 55 \\ 19 & 1 & 100 & 6 \\ 20 & 60 & 16 & 17 \\ 0 & 15 & 15 & 30 \end{bmatrix} \quad J^1_{(i,j,2)} = \begin{bmatrix} 37 & 6 & 8 & 3 \\ 40 & 10 & 0 & 17 \\ 4 & 6 & 43 & 10 \\ 0 & 2 & 15 & 100 \end{bmatrix}$$

$$J^2_{(i,j,1)} = \begin{bmatrix} 0 & 17 & 9 & 11 \\ 13 & 4 & 0 & 17 \\ 9 & 0 & 11 & 46 \\ 9 & 0 & 25 & 11 \end{bmatrix} \quad J^2_{(i,j,2)} = \begin{bmatrix} 5 & 29 & 12 & 0 \\ 9 & 16 & 54 & 29 \\ 74 & 1 & 4 & 0 \\ 9 & 16 & 42 & 0 \end{bmatrix}$$

$$J^3_{(i,j,1)} = \begin{bmatrix} 9 & 13 & 70 & 9 \\ 13 & 5 & 0 & 70 \\ 11 & 2 & 16 & 19 \\ 16 & 49 & 3 & 10 \end{bmatrix} \quad J^3_{(i,j,2)} = \begin{bmatrix} 50 & 6 & 9 & 10 \\ 0 & 16 & 59 & 1 \\ 16 & 48 & 2 & 9 \\ 110 & 46 & 28 & 9 \end{bmatrix}$$

$$J^4_{(i,j,1)} = \begin{bmatrix} 17 & 7 & 13 & 5 \\ 19 & 43 & 11 & 0 \\ 1 & 16 & 20 & 6 \\ 27 & 4 & 0 & 25 \end{bmatrix} \quad J^4_{(i,j,2)} = \begin{bmatrix} 30 & 0 & 18 & 23 \\ 14 & 10 & 28 & 16 \\ 4 & 0 & 14 & 6 \\ 4 & 8 & 9 & 15 \end{bmatrix}$$

The transition matrices for each player are defined as follows

$$\pi^1_{(i,j,1)} = \begin{bmatrix} 0.6144 & 0.3856 & 0 & 0 \\ 0.4061 & 0.2772 & 0.3167 & 0 \\ 0 & 0.1208 & 0.1688 & 0.7104 \\ 0 & 0 & 0.7696 & 0.2304 \end{bmatrix} \quad \pi^1_{(i,j,2)} = \begin{bmatrix} 0.5535 & 0.4465 & 0 & 0 \\ 0.7197 & 0.1596 & 0.1207 & 0 \\ 0 & 0.6374 & 0.1401 & 0.2225 \\ 0 & 0 & 0.6407 & 0.3593 \end{bmatrix}$$

$$\pi^2_{(i,j,1)} = \begin{bmatrix} 0.6420 & 0.3580 & 0 & 0 \\ 0.2317 & 0.3736 & 0.3947 & 0 \\ 0 & 3269 & 0.1621 & 0.5110 \\ 0 & 0 & 0.1211 & 0.8789 \end{bmatrix} \quad \pi^2_{(i,j,2)} = \begin{bmatrix} 0.7040 & 0.2960 & 0 & 0 \\ 0.4417 & 0.1015 & 0.4568 & 0 \\ 0 & 0.3273 & 0.2675 & 0.4052 \\ 0 & 0 & 0.3555 & 0.6445 \end{bmatrix}$$

$$\pi^3_{(i,j,1)} = \begin{bmatrix} 0.3341 & 0.6659 & 0 & 0 \\ 0.1898 & 0.4103 & 0.3999 & 0 \\ 0 & 0.4340 & 0.2173 & 0.3487 \\ 0 & 0 & 0.5240 & 0.4760 \end{bmatrix} \quad \pi^3_{(i,j,2)} = \begin{bmatrix} 0.4053 & 0.5947 & 0 & 0 \\ 0.3706 & 0.4358 & 0.1936 & 0 \\ 0 & 0.6393 & 0.1778 & 0.1829 \\ 0 & 0 & 0.6756 & 0.3244 \end{bmatrix}$$

$$\pi^4_{(i,j,1)} = \begin{bmatrix} 0.6299 & 0.3701 & 0 & 0 \\ 0.2932 & 0.5345 & 0.1723 & 0 \\ 0 & 0.2800 & 0.5336 & 0.1864 \\ 0 & 0 & 0.4102 & 0.5898 \end{bmatrix} \quad \pi^4_{(i,j,2)} = \begin{bmatrix} 0.2267 & 0.7733 & 0 & 0 \\ 0.2195 & 0.5162 & 0.2643 & 0 \\ 0 & 0.2081 & 0.6278 & 0.1641 \\ 0 & 0 & 5328 & 0.4672 \end{bmatrix}$$

Given δ and γ and applying the extraproximal method we obtain the convergence of the strategies in terms of the variable $c_{(i,k)}$ for the pursuers (see Figures 4.1 and 4.2) and for the evaders (see Figures 4.3 and 4.4). In addition, Figures 4.5 and 4.6 show the convergence of the parameters ξ and ω.

With final values $\lambda^1 = 0.77$ and $\lambda^2 = 0.23$ for the leaders (pursuers) (see Figure 4.7), the mixed strategies obtained for all the players are as follows

$$d^1 = \begin{bmatrix} 0.9741 & 0.0259 \\ 0.9722 & 0.0278 \\ 0.0543 & 0.9457 \\ 0.1438 & 0.8562 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.4627 & 0.5373 \\ 0.2462 & 0.7538 \\ 0.5897 & 0.4103 \\ 0.3923 & 0.6077 \end{bmatrix}$$

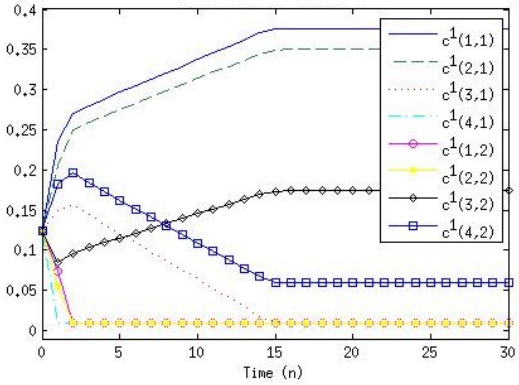


Figure 4.1 Strategies for pursuer 1.

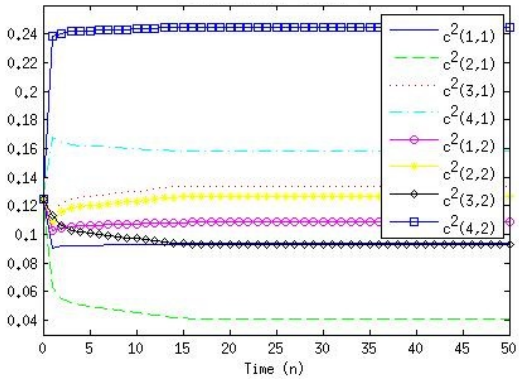


Figure 4.2 Strategies for pursuer 2.

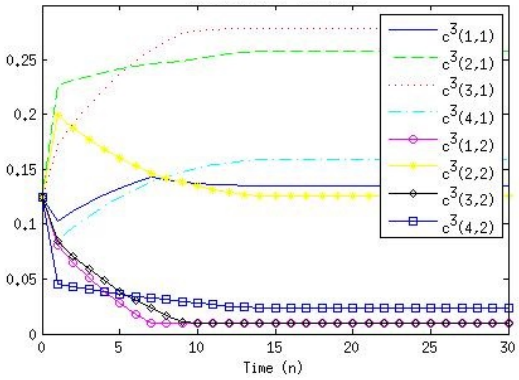


Figure 4.3 Strategies for evader 1.

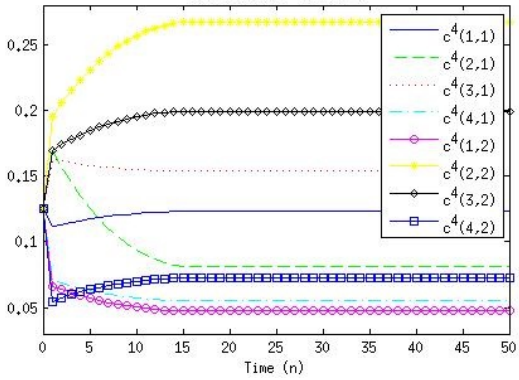


Figure 4.4 Strategies for evader 2.

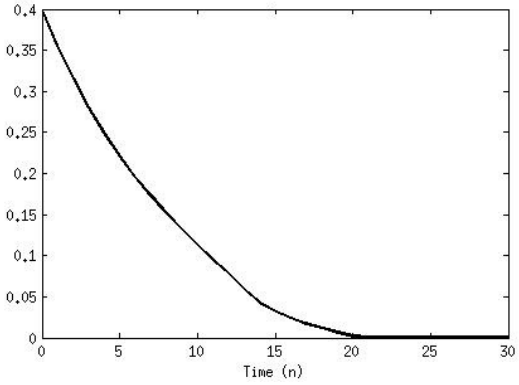


Figure 4.5 Convergence of the parameter ξ.

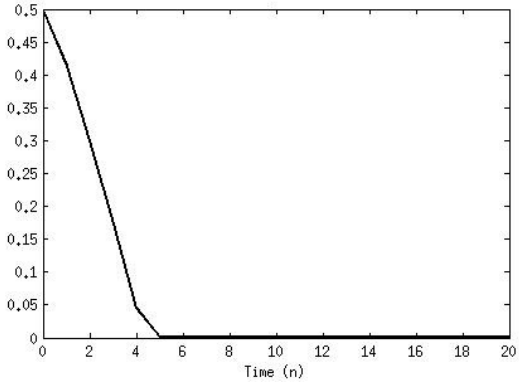


Figure 4.6 Convergence of the parameter ω.

$$d^3 = \begin{bmatrix} 0.9308 & 0.0692 \\ 0.6733 & 0.3267 \\ 0.9654 & 0.0346 \\ 0.8731 & 0.1269 \end{bmatrix} \qquad d^4 = \begin{bmatrix} 0.7214 & 0.2786 \\ 0.2340 & 0.7660 \\ 0.4354 & 0.5646 \\ 0.4318 & 0.5682 \end{bmatrix}$$

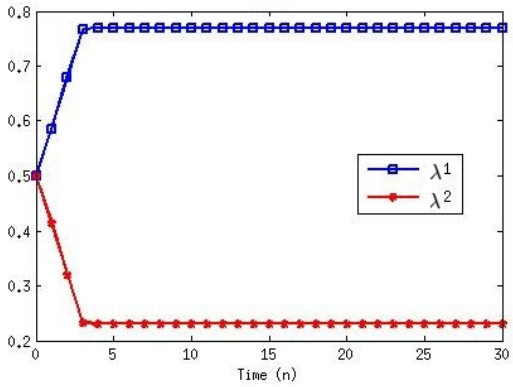


Figure 4.7 Convergence of λ.

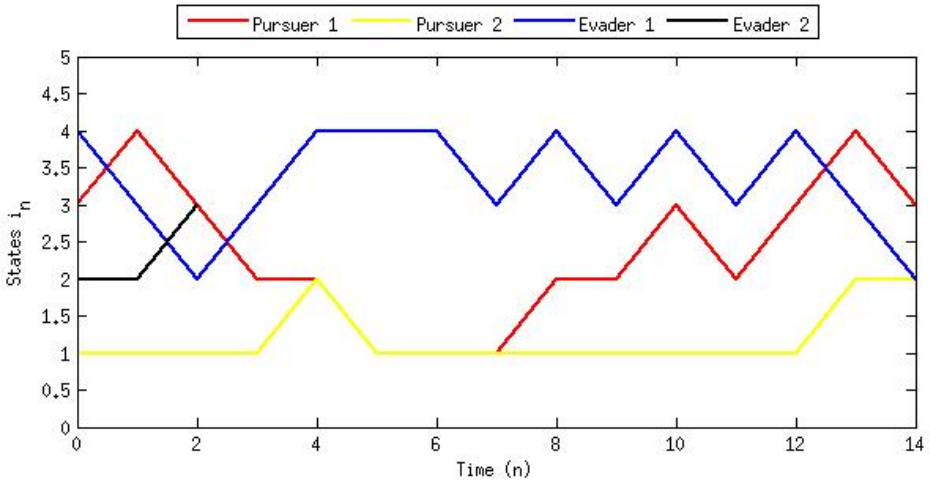


Figure 4.8 Realization of the game.

For the realization of the game we define the initial state $s_{(i)}$ of each player as follows: $s^1_{(i)}(0) = 3$, $s^2_{(i)}(0) = 1$, $s^3_{(i)}(0) = 4$ and $s^4_{(i)}(0) = 2$. As a result, we obtain that the evader 1 is caught at state $s_{(2)}$ and evader 2 is caught at state $s_{(3)}$, so the game is over (see Figure 4.8).

4.6.2 Marketing problem

This example analyzes the effectiveness of relationship marketing strategies within the department store sector of the retail industry considering two supermarket leaders with $l = 1, 2$ and two supermarkets followers with $m = 3, 4$. The four supermarkets are branching out into non-food items and they are also department stores in their own right, selling items like clothes,

entertainment products for example toys, books, cosmetics, non-prescription drugs and many other household goods. All the supermarkets offer loyalty cards having their own system with the purpose to attract customers, encourage customer loyalty and build strong customer relationships. As well, loyalty cards create an advantage for supermarkets developing profiles of individuals' personal shopping habits. When linked with the personal details that customers disclosed when signing up for the scheme, the store is in a position to target promotions that are tailored around specific customers shopping habits.

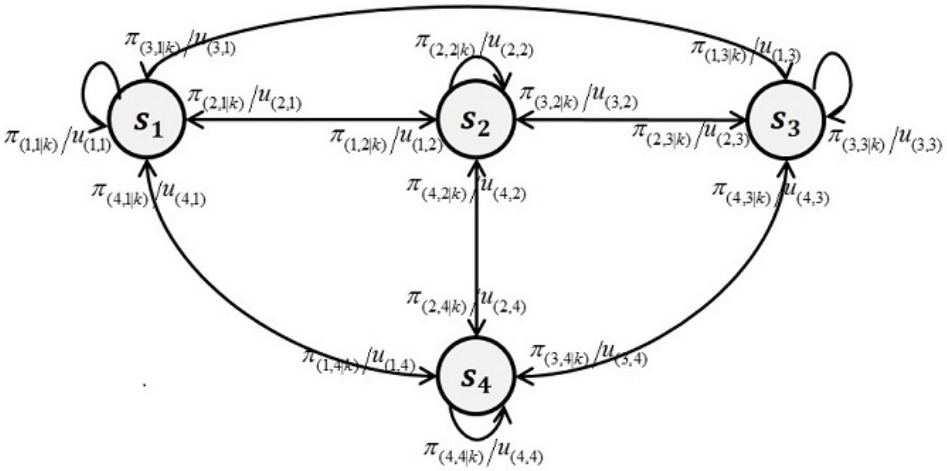


Figure 4.9 Supermarket Markov Chain.

Based on the available data, supermarkets discretize the client space in four sub-segments according to the regularly of purchasing, using frequency of the loyalty card and the revenue. Figure 4.9 describes the segments and promotions corresponding to the Markov chain of the marketing problem. Here a customer is said to be in state s_1 if he/she become a Potential customer. A Low-frequent customer corresponds with the state s_2 and a Regular customer is a frequent customer of the loyalty card that is said to be in state s_3. A Loyal customer corresponds with the state s_4 and he/she is a high-frequency user of the loyal card. The promotions (actions) offered by the supermarkets include two different benefits: 1) points and 2) discounts. We are interested in contrasting the strategies applied by the supermarkets defined over all possible combinations of states (i, j) and actions (k) given a fixed utility $U_{(i,j,k)}$.

Our goal is to analyze a four-player Stackelberg game for the norm $p = 1$ in a class of ergodic controllable finite Markov chains. Let $N = 4$, $M = 2$. The individual utility for each player are defined by

$$U^1_{(i,j,1)} = \begin{bmatrix} 567 & 822 & 733 & 830 \\ 261 & 896 & 85 & 568 \\ 30 & 996 & 634 & 261 \\ 288 & 90 & 806 & 785 \end{bmatrix} \quad U^1_{(i,j,2)} = \begin{bmatrix} 170 & 27 & 57 & 699 \\ 275 & 855 & 224 & 919 \\ 50 & 205 & 46 & 909 \\ 398 & 861 & 751 & 806 \end{bmatrix}$$

$$U^2_{(i,j,1)} = \begin{bmatrix} 810 & 36 & 27 & 9 \\ 63 & 90 & 567 & 72 \\ 81 & 0 & 9 & 45 \\ 855 & 594 & 441 & 9 \end{bmatrix} \quad U^2_{(i,j,2)} = \begin{bmatrix} 8 & 592 & 48 & 0 \\ 64 & 64 & 312 & 16 \\ 264 & 32 & 120 & 72 \\ 400 & 56 & 40 & 200 \end{bmatrix}$$

$$U^3_{(i,j,1)} = \begin{bmatrix} 22 & 7 & 11 & 6 \\ 10 & 0 & 19 & 8 \\ 23 & 28 & 23 & 9 \\ 90 & 5 & 12 & 1 \end{bmatrix} \quad U^3_{(i,j,2)} = \begin{bmatrix} 66 & 0 & 126 & 42 \\ 18 & 78 & 240 & 6 \\ 96 & 18 & 60 & 156 \\ 66 & 102 & 180 & 48 \end{bmatrix}$$

$$U^4_{(i,j,1)} = \begin{bmatrix} 0 & 60 & 2 & 26 \\ 10 & 26 & 36 & 48 \\ 14 & 56 & 28 & 24 \\ 8 & 12 & 16 & 38 \end{bmatrix} \quad U^4_{(i,j,2)} = \begin{bmatrix} 420 & 168 & 378 & 84 \\ 0 & 280 & 14 & 112 \\ 42 & 56 & 350 & 140 \\ 84 & 210 & 336 & 98 \end{bmatrix}$$

The transition matrices for each player are defined as follows

$$\pi^1_{(i,j,1)} = \begin{bmatrix} 0.2759 & 0.4886 & 0.0366 & 0.1989 \\ 0.1752 & 0.0953 & 0.3825 & 0.3470 \\ 0.1695 & 0.2629 & 0.4103 & 0.1574 \\ 0.2612 & 0.1665 & 0.4124 & 0.1600 \end{bmatrix} \quad \pi^1_{(i,j,2)} = \begin{bmatrix} 0.0863 & 0.3672 & 0.3201 & 0.2264 \\ 0.4339 & 0.1684 & 0.1919 & 0.2058 \\ 0.3856 & 0.2349 & 0.1324 & 0.2471 \\ 0.1475 & 0.3500 & 0.1903 & 0.3122 \end{bmatrix}$$

$$\pi^2_{(i,j,1)} = \begin{bmatrix} 0.1761 & 0.1204 & 0.3883 & 0.3151 \\ 0.2207 & 0.1632 & 0.2354 & 0.3807 \\ 0.0708 & 0.3708 & 0.1364 & 0.4219 \\ 0.0132 & 0.5169 & 0.4127 & 0.0572 \end{bmatrix} \quad \pi^2_{(i,j,2)} = \begin{bmatrix} 0.2033 & 0.2456 & 0.2667 & 0.2844 \\ 0.2732 & 0.1032 & 0.3046 & 0.3190 \\ 0.1207 & 0.0930 & 0.3997 & 0.3866 \\ 0.1032 & 0.6976 & 0.1609 & 0.0383 \end{bmatrix}$$

$$\pi^3_{(i,j,1)} = \begin{bmatrix} 0.4109 & 0.1654 & 0.0918 & 0.3319 \\ 0.3015 & 0.2201 & 0.1029 & 0.3756 \\ 0.1709 & 0.5673 & 0.0292 & 0.2326 \\ 0.1885 & 0.1491 & 0.3317 & 0.3307 \end{bmatrix} \quad \pi^3_{(i,j,2)} = \begin{bmatrix} 0.3046 & 0.2883 & 0.2573 & 0.1498 \\ 0.2470 & 0.0978 & 0.3060 & 0.3492 \\ 0.3006 & 0.0439 & 0.4387 & 0.2169 \\ 0.1141 & 0.3397 & 0.1855 & 0.3607 \end{bmatrix}$$

$$\pi^4_{(i,j,1)} = \begin{bmatrix} 0.2610 & 0.3145 & 0.2088 & 0.2158 \\ 0.3777 & 0.1968 & 0.1574 & 0.2681 \\ 0.2593 & 0.0308 & 0.5113 & 0.1986 \\ 0.3401 & 0.4638 & 0.1200 & 0.0761 \end{bmatrix} \quad \pi^4_{(i,j,2)} = \begin{bmatrix} 0.0316 & 0.4652 & 0.2221 & 0.2811 \\ 0.1624 & 0.3245 & 0.3691 & 0.1440 \\ 0.1448 & 0.5777 & 0.2087 & 0.0688 \\ 0.2536 & 0.1996 & 0.3231 & 0.2237 \end{bmatrix}$$

Given δ and γ and applying the extraproximal method we obtain the convergence of the strategies in terms of the variable $c_{(i,k)}$ for the leaders (see Figure 4.10) and for the followers (see Figure 4.11). In addition, the Figure 4.12 and Figure 4.13 show the convergence of the parameters ξ and ω.

With final values $\lambda^1 = 0.5063$ and $\lambda^2 = 0.4937$ for the leaders, and $\theta^1 = 0.5258$ and $\theta^2 = 0.4792$ for the followers (see Figure 4.14 and Figure 4.15), the mixed strategies obtained for determining the strong Stackelberg/Nash equilibrium for all the players applying (2.6) are as follows

$$d^1 = \begin{bmatrix} 0.8110 & 0.1890 \\ 0.1701 & 0.8299 \\ 0.7720 & 0.2280 \\ 0.2249 & 0.7751 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.6023 & 0.3977 \\ 0.8408 & 0.1592 \\ 0.8187 & 0.1813 \\ 0.8242 & 0.1758 \end{bmatrix}$$

$$d^3 = \begin{bmatrix} 0.6478 & 0.3522 \\ 0.7078 & 0.2922 \\ 0.6455 & 0.3545 \\ 0.6442 & 0.3558 \end{bmatrix} \quad d^4 = \begin{bmatrix} 0.7337 & 0.2663 \\ 0.7454 & 0.2546 \\ 0.7376 & 0.2624 \\ 0.6418 & 0.3582 \end{bmatrix} \tag{4.9}$$

The resulting utilities by segment are as follows:

$$J^1(s_i) = \begin{bmatrix} 129,130 \\ 92,790 \\ 84,590 \\ 121,520 \end{bmatrix} \quad J^2(s_i) = \begin{bmatrix} 13,102 \\ 22,635 \\ 1,113 \\ 64,809 \end{bmatrix} \quad J^3(s_i) = \begin{bmatrix} 551 \\ 1,295 \\ 746 \\ 1,494 \end{bmatrix} \quad J^4(s_i) = \begin{bmatrix} 3,914 \\ 2,113 \\ 2,158 \\ 3,467 \end{bmatrix} \tag{4.10}$$

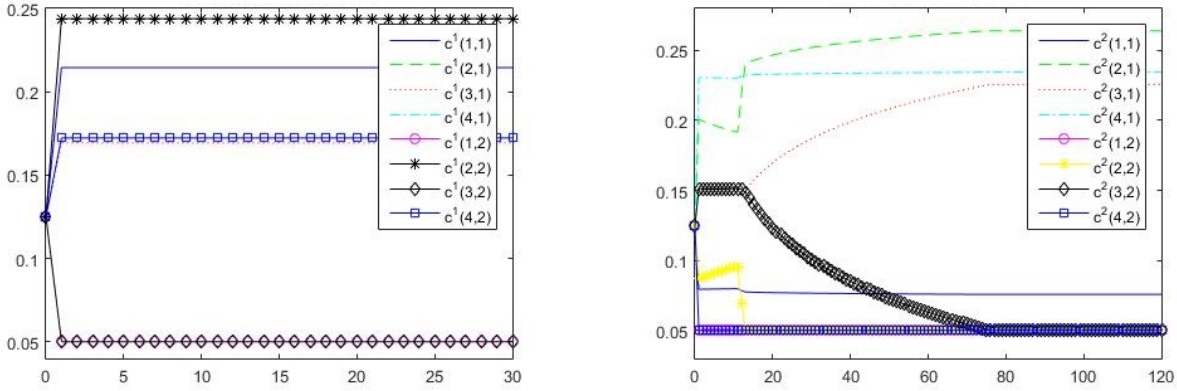


Figure 4.10 Convergence of the strategies for leader 1 (left) and leader 2 (right).

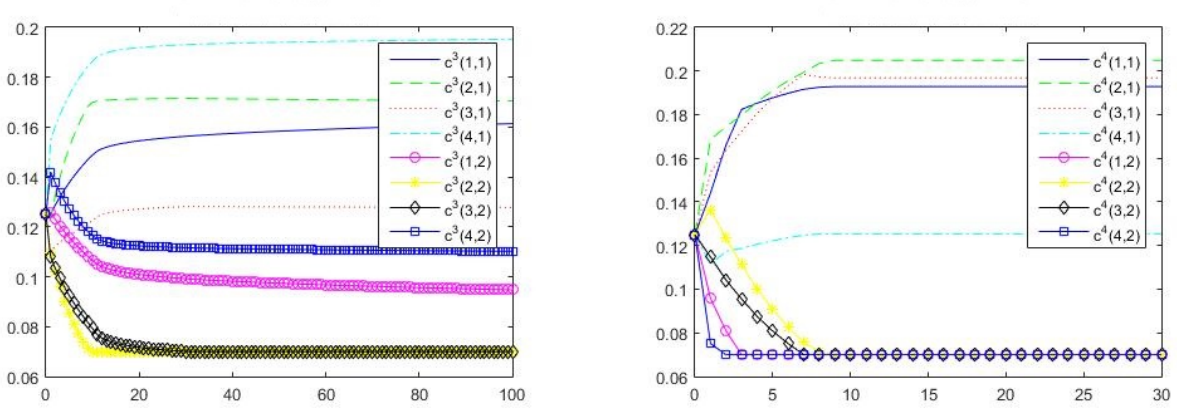


Figure 4.11 Convergence of the strategies for follower 1 (left) and follower 2 (right).

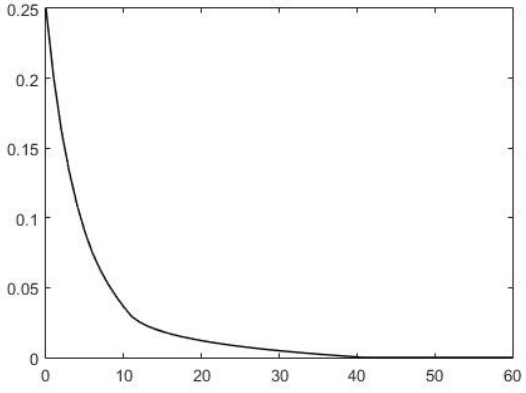


Figure 4.12 Convergence of the parameter ξ.

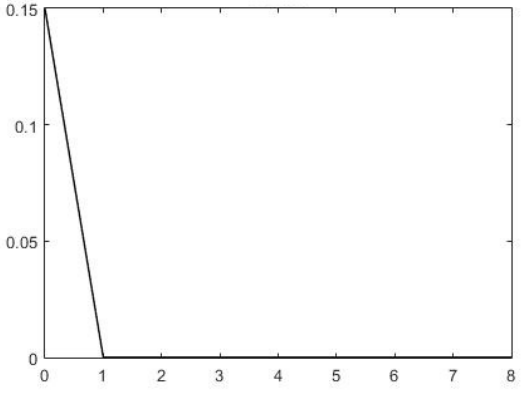


Figure 4.13 Convergence of the parameter ω.

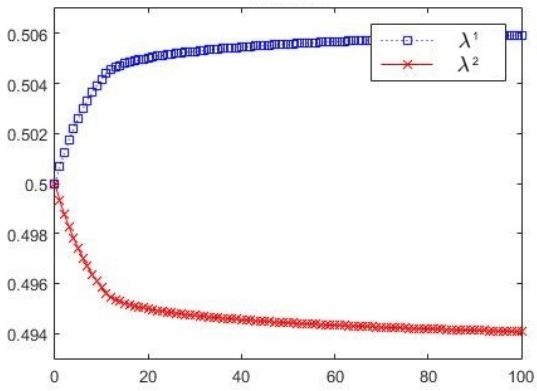


Figure 4.14 Convergence of the parameter λ.

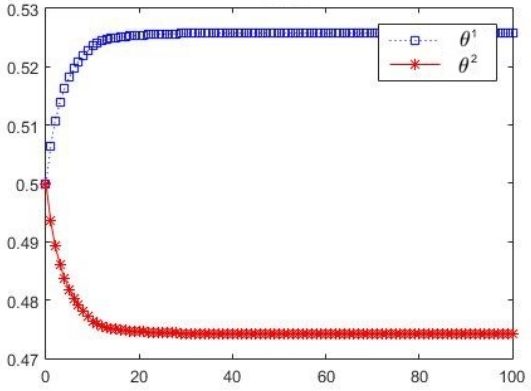


Figure 4.15 Convergence of the parameter θ.

And the resulting utilities by promotion are as follows:

$$\begin{aligned} J^1(k) &= \begin{bmatrix} 226,830 & 201,190 \end{bmatrix} \qquad J^2(k) = \begin{bmatrix} 93,930 & 7,729 \end{bmatrix} \\ J^3(k) &= \begin{bmatrix} 437 & 3,650 \end{bmatrix} \qquad J^4(k) = \begin{bmatrix} 609 & 11,045 \end{bmatrix} \end{aligned} \tag{4.11}$$

Relationship marketing recognizes that the focus of marketing is to build a relationship with existing customers. The main purpose of the game is to discover the extent to which customers use and are influenced by relationship marketing strategies. In addition, it is to analyze the impact that these strategies have on customer loyalty and the development of customer-department store relationship. The supermarket leaders (players 1 and 2) fix their strategies (4.9) to ensure high degrees of customer loyalty and retention as well utility by segment (4.10) and promotion (4.11). For segment 1, leader 1 made a strong emphasis on offering points (0.8110) for attracting Potential customers. Instead, the leader 2 made emphasis on offering points (0.6023) and discounts (0.3977) for the same segment. Looking at the utilities of the leaders, the follower1 decided for offering points (0.6478) and discounts (0.3522). Instead, the follower 2 resolved for competing for highlighting points (0.7337). For segment 2 corresponding to Low-Frequent customers the leader 1 promoted points (0.1701) and discounts (0.8299) and, the leader 2 chose offering points (0.8408) and discounts (0.1592). However, for competing with the leaders, follower 1 and follower 2 made emphasis on points (0.7078 and 0.7454 respectively). For Regular customers, the leader 1 focused on points (0.7720) and discounts (0.2280) and, the leader 2 made emphasis on points (0.8187). The follower 1 preferred offering

points (0.6455) and discounts (0.3545). Instead, follower 2 made emphasis on points (0.7376) and discounts (0.2624). For Loyal customers, the leader 1 made emphasis on points (0.2249) and discounts (0.7751), leader 2 focus on points (0.8242) and discounts (0.1758) as well, follower 1 chose the same strategies – points (0.6442) and discounts (0.3558) –. The follower 2 made emphasis on points (0.6418) and discounts (0.3582). For the leaders, the most profitable segments are the Potential customers and the Loyal customers (see 4.10). An insight into the mind of the consumer is obvious from the findings the importance that is placed on a given policy: the utilities obtained by action for the leaders and followers are shown in (4.11).

Chapter 5

A Reinforcement Learning Approach for Stackelberg Security Games

5.1 Introduction

There exists a growing interest in applying Stackelberg games to model resource allocation for patrolling security problems in which defenders must allocate limited security resources to protect targets from attack by adversaries [20, 28]. In real-world adversaries are sophisticated presenting dynamic strategies. In the original Stackelberg security games formulation on Markov chains, we usually assume fixed and static domains models not able to be adapted to the environment: fixing a state and action the reward and transitions always remain the same. The reason is that the main goal is minimizing/maximizing the players' expected cost/reward that depends on the transitions at each state. However, it is an unrealistic assumption: the transitions matrices and the reward received for Stackelberg security games are commonly non-static. Producing always the same resulting behavior can be exploited by intelligent attackers that carry out surveillance before an attack, it is often desirable for the security agencies to have a system in which randomness is involved in allocating their resources. To address this shortcoming, we will consider the learning properties of the attackers and defenders interaction, and we will deal with the adaptation (estimation and assessment) of the payoff and strategies to dynamic environments based on the information available to them.

Game-theoretic approaches have been used in multiple deployed applications. These games are security games between a defender and an attacker: first, the defender considers what the

target (best-reply) of the attacker is; then, holding the attacked target fixed, the defender picks a quantity that minimizes its payoff; finally, the attacker actually observes this and in equilibrium picks the expected quantity that maximizes its payoff as a response. These applications [72, 43, 106, 73, 2] use the (two-players) leader-follower Stackelberg game-theoretic formulation for solving the security problem, providing a randomized strategy for the defender (leader) and the attacker (follower).

Reinforcement learning (RL) is a problem faced by an agent or multiple agents that must learn behavior through trial-and-error interactions with a dynamic environment [44]. It does not assume the existence of a teacher that provides examples upon which learning of a task takes place [78]. Computationally, RL is intended to operate in a learning environment composed of two subjects: the learner and a dynamic process. At successive time steps, the learner makes an observation of the process state, selects an action and applies it back to the process. Its goal is to find out an action policy that controls the behavior of the dynamic process, guided by signals that indicate how badly or well it has been performing the required task. These signals are usually associated with a dramatic condition, a reward or a punishment, and the learner tries to optimize its behavior [78].

Motivated by the importance of game-theoretic Markov chains solutions, this chapter considers a RL process for Stackelberg security games [103, 99] that involves two components: the Adaptive Primary Learning architecture and the Actor-critic architecture. The Adaptive Primary Learning architecture proposes a connection between prior knowledge learning and imitative learning. The main goal of the Adaptive Primary Learning architecture is dramatically to accelerate the reinforcement learning process. The Actor-critic architecture is a temporal-difference method responsible for evaluating the new state and determine if rewards are better or worse than expected based on a game theory solution. The Stackelberg game is solved in terms of the L_p-norm: players choose a strategy that minimizes the distance to the utopian minimum and no other strategy produces a smaller total expected loss. The notion of collaboration implies that related players interact with each other looking for cooperative stability. This notion consents players to select optimal strategies and to condition their own behavior on the behavior of others in a strategic forward-looking manner.

The overall RL architecture presents several benefits. The Adaptive Primary Learning architecture can be viewed as a process for enhancing learning for multiple players. It allows players to use prior knowledge of the security problem. This is given in terms of the Markov assumptions and a uniform distribution that represent a simple solution of the security game. In the short term, learning according to a uniform distribution helps by focusing on states that are near increments of the pay-off of the starting state. It also augments players' ability to learn useful behaviors by making intelligent use of the knowledge implicit in behaviors demonstrated by cooperative mentors (more experienced players). Using reinforcement learning theory we construct a formal framework for security games that allows players to combine prior knowledge and imitative behavior (extracted from other players). This framework uses observations of other players behavior to provide a player with transition probabilities about its capabilities in unexperienced situations. The actor-critic architecture will execute a learning process based on a Stackelberg game theory solution. It will use the best-reply strategies to obtain the estimated model for the occurring actions and states. In order to address the dynamic execution uncertainty in security patrolling, we provide a game-theoretic formulation method able to generate randomized patrol schedules based on Markov decision process.

The formulation of the game is considered as a nonlinear programming problem for finding the strong L_p-Stackelberg/Nash equilibrium point based on cost-functions that are supposed to be (non-obligatory strictly) convex and differentiable on the corresponding sets. This problem is analyzed for a class of ergodic controllable finite Markov chains using the extraproximal method. It is also provided a game-theoretic formulation method able to generate randomized patrol schedules based on the Stackelberg game theory solution.

Moreover, an efficient algorithm for players that accelerate the reinforcement learning process is presented. Computing the best-reply strategies for the game in the actor-critic architecture requires a large computation time compared with the computation time required in the adaptive primary learning architecture. However, both steps of the architecture combined will, in the long run, converge to an estimated transition matrix and estimated utility provided that acting using the best-reply strategy according to the sequence of estimated models leads the players to explore the entire state-action space.

5.2 The Stackelberg security game

A Stackelberg security game [98] includes defenders (the leaders in the game) who aim to protect a set of targets against attackers (the followers in a Stackelberg game). The defenders play first by committing to a randomized strategy. The defenders' commitment is observed by the attackers, who then play a best-reply to the defenders' strategy. The role of the defender is usually played by a security agency, which has the responsibility of protecting critical infrastructure. The strategy set of the defenders can be interpreted as the assignments of (protecting) resources to potential targets. The goal is to minimize the damage. The attacker observes the defenders' randomized strategies (resources deployment), and choose a target to attack in a way that maximizes the damage.

We describe a Stackelberg game as follows. Let us consider a game with $\mathfrak{n} + \mathfrak{m}$ players. Let $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ denote the set of players called defenders and let their strategy set be defined by U. The rest $\mathcal{M} = \{1, ..., \mathfrak{m}\}$ players are called attackers and, similarly, let the set of their strategy profiles be defined by V. Then, $U \times V$ is the set of full strategy profiles. The dynamics of the Stackelberg security game is as follows: the defenders choose a strategy $u \in U$ considering the cost-function $\varphi(u|v)$ for a fixed strategy v of the attackers, the attackers are informed about the strategy u selected by the defenders and choose their strategies considering $\psi(v|u)$ for a fixed u of the defenders. We understand $\psi(v|u)$ as the response of the attackers to the strategy u of the defenders, which is the best-reply in the original game. In the security game framework, we suppose that defenders commit to a randomized strategy while attackers choose their best-reply to this strategy. The solution of the game is a Stackelberg equilibrium point. The formalization of the Stackelberg game was presented in Chapter 4.

Also, it is considered a Stackelberg security game model where each player either staying put or moving along a state to an adjacent state. Adjacency of the states is determined by the probabilities given in the transitions matrices of the Markov chain. The main concern about Stackelberg games is as follows: the highest leader payoff is obtained when the followers always reply in the best possible way for the leader. Then, the defender can *capture* the attacker,

because he implements a strategy that always dominates the current position of the attacker. Once the attacker is caught, the security game is over.

Let us introduce the capture condition at time n (defender and attacker are located at the same state) as follows:

$$\sum_{j=1}^{N} \chi(w : s^l(n) = s_{(j)} \wedge s^m(n) = s_{(j)}) = \sum_{j=1}^{N} \chi(w : s^l(n) = s_{(j)})\, \chi(w : s^m(n) = s_{(j)})$$

where $w \in \Omega$ is a trajectory. The capture event of all the attackers is given by

$$\sum_{l=1}^{\mathfrak{n}} \sum_{m=1}^{\mathfrak{m}} \sum_{j=1}^{N} \chi(w : s^l(n) = s_{(j)}) \chi(w : s^m(n) = s_{(j)}) \tag{5.1}$$

In the dynamics of the game, the defender commits first to a strategy and then, the attacker strategy is played. We consider a *Random Walk* model such that each member either staying put or moving along a state to an adjacent state. The defender can capture the attacker (5.1) if he implements an appropriate strategy such as always moving toward the current position of the attacker. Once the attacker is caught, the game is over. The computational algorithm in Table (5.1) for each player $\iota = 1, ..., \mathfrak{n} + \mathfrak{m}$ is iterative.

5.3 RL security game architecture

The aim of this Section is to introduce the RL architecture for the Stackelberg security game. It is illustrated in Figure 5.1 showing two highest components: the Adaptive Primary Learning architecture and the Actor-critic architecture.

Consider first the Adaptive Primary Learning architecture proposed to increase the learning speed. It is better understood as an attempt to combine prior knowledge with an imitation process for selecting the strategies. In fact, the prior knowledge will be cast and augmented with the imitative learning formalism. The Adaptive Primary Learning architecture of the RL for the game is illustrated in Figure 5.2. It has two main modules: the belief-forming process and the belief-imitating process.

The *belief-forming process* provides the player ι with the ability to seed a learning algorithm about the security problem. It allows the player to use prior knowledge of the problem. This is

Algorithm 1:

1. For the matrix $d^{\iota}_{(k|i)}$ find the action $a^{\iota} = a^{\iota}_{(k)}$ using a random $k \in (1, ..., M)$ distributed according to the stochastic vector $(d^{\iota}_{(1|i)}, ..., d^{\iota}_{(M|i)})$ for a fixed $i \in (1, ..., N)$.

2. Using the matrix $\pi^{\iota}_{(j|i,k)}$ find the next state $s_{(j)}$ selecting randomly $j \in (1, ..., N)$ distributed according to the stochastic vector $(\pi^{\iota}_{(1|i,k)}, ..., \pi^{\iota}_{(N|i,k)})$ for a fixed $i \in (1, ..., N)$ and action $k \in (1, ..., M)$.

3. Add the state $s_{(j)}$ to the patrol schedule and update the initial value of i with j.

4. Repeat steps (1), (2) and (3) until the capture condition (5.1) is satisfied.

Table 5.1 Patrol Schedule

given in terms of the Markov assumptions and a uniform distribution that represent the initial solution of the game. The focus of the belief-forming process is to solve general security situations given a uniform distribution of the game. The uniform distribution of the strategies is useful for balancing exploration and exploitation in a basic reinforcement learning (one drawback is that when it explores it chooses equally among all actions). This process is marred by generic optimization criteria. A stochastic strategy selector is used to generate exploratory random action of player ι from $d^{\iota}_{(k|i)}$ at the beginning of the training process. As well as, a stochastic strategy selector is used to generate exploratory random next step from $\pi^{\iota}_{(j|i,k)}$. We employ two different learning rules $\hat{\pi}^{\iota}_{(\hat{j}|\hat{\imath},\hat{k})}(t)$ and $\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t)$ for estimating the resulting values.

The *belief-imitating process* provides a player ι with a system of rules that idealize the mentor's belief-forming behavior. It augments a player's ability to learn using the knowledge implicit in behaviors demonstrated by more experienced players. An estimated value is considered to be imitated according to a rule which encourages or discourages the current strategy

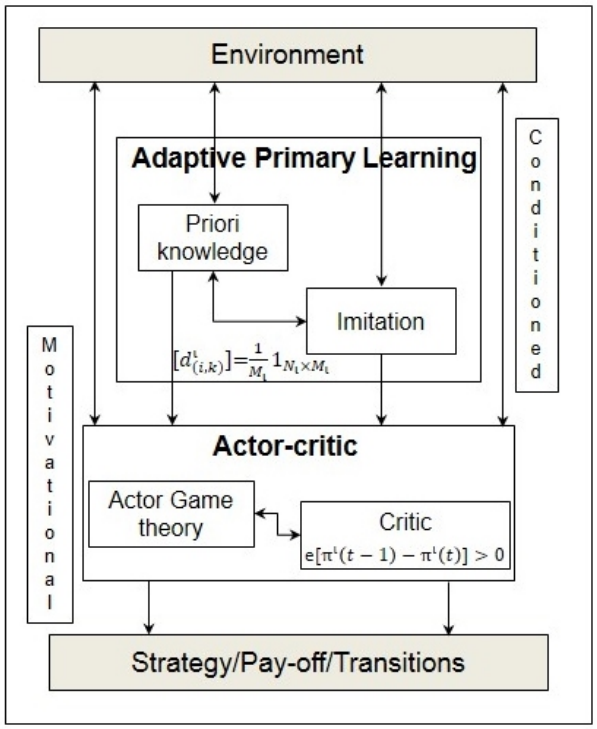


Figure 5.1 Reinforcement learning architecture.

depending on cost. The main idea is that if the estimated value is considered if the cost is smaller than the corresponding estimated value because it decreases the expected cost for the visited state. In security games, a defender tries to minimize the capture time, as well as, the attacker tries to maximize the escape time.

The dynamics of the Adaptive Primary Learning architecture is as follows. The process for player ι begins with an initial state $s^{\iota}(0) = s^{\iota}_{(\hat{\imath})}$ (for the estimated value $\hat{\imath}$) and it is considered a fixed uniform distribution of the strategies as a solution of the security game given by $[d^{\iota}_{(k|\hat{\imath})}] = \frac{1}{M}[1_{(i,k)}]_{i\in[1,N],k\in[1,M]}$, where $[1_{(i,k)}]$ is a matrix of "ones" of size $i \times k$. Then, it is chosen randomly an action $a^{\iota}(t) = a_{(\hat{k})}$ (for the estimated value $\hat{k}$) from the vector $d^{\iota}_{(k|\hat{\imath})}$ (for a fixed $\hat{\imath}$). After that, the transition matrix $\Pi^{\iota} = \left[\pi^{\iota}_{(j|i,k)}\right]$ is used to choose randomly the consecutive state $s^{\iota}(t+1) = s^{\iota}_{(\hat{\jmath})}$ (for the estimated value $\hat{\jmath}$) from the vector $\pi^{\iota}_{(j|\hat{\imath},\hat{k})}$ (for a fixed $\hat{\imath}$ and $\hat{k}$). Once $a^{\iota}(t)$ and $s^{\iota}(t+1)$ are selected the estimating values are updated employing the adaptive module in which the learning rules $\hat{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)$ and $\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t)$ are computed. Then, it is determined a mentor κ. The player ι ($\iota \neq \kappa$) imitate the estimated value $\hat{\pi}^{\kappa}_{(\hat{\imath},\hat{\jmath}|\hat{k})}$ according to a rule which

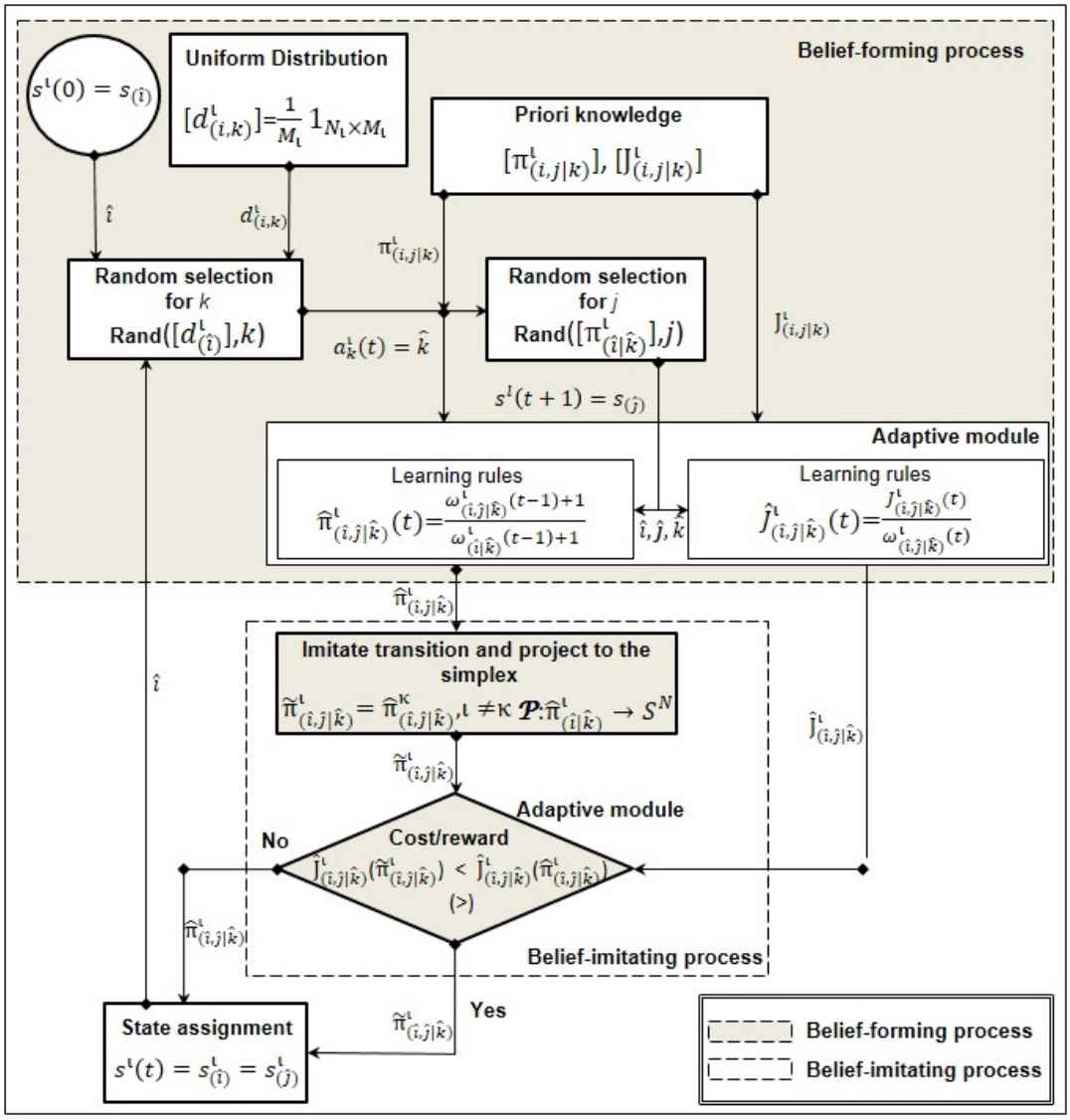


Figure 5.2 Adaptive primary learning architecture.

encourages or discourages the current strategy depending on cost $\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath}|\hat{k})}$, updating $\tilde{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}$ if it is the case, which has to be projected to the simplex. Finally, it is made the assignment of the next state $\hat{\jmath}$ to the currently state $s^{\iota}(t+1) = s^{\iota}_{(\hat{\imath})} = s^{\iota}_{(\hat{\jmath})}$, i.e. $\hat{\imath} = \hat{\jmath}$, and the process begins again until it converges.

Actor-critic methods are temporal-difference learning methods. The process responsible for generating the policy structure ($d^{\iota}_{(k|i)}$) and, selecting an action and next state is known as the actor, while the process in charge of estimating the value function is known as the critic.

The learning process is all the time on the policy $d^{\iota}_{(k|i)}$ of the players. The critic task is to learn about the complete process and analyze if the policy represents the best-reply that must be followed by the actor. To fulfill the task the critic uses an error estimator (e) which manages all the learning decisions for both the actor and the critic.

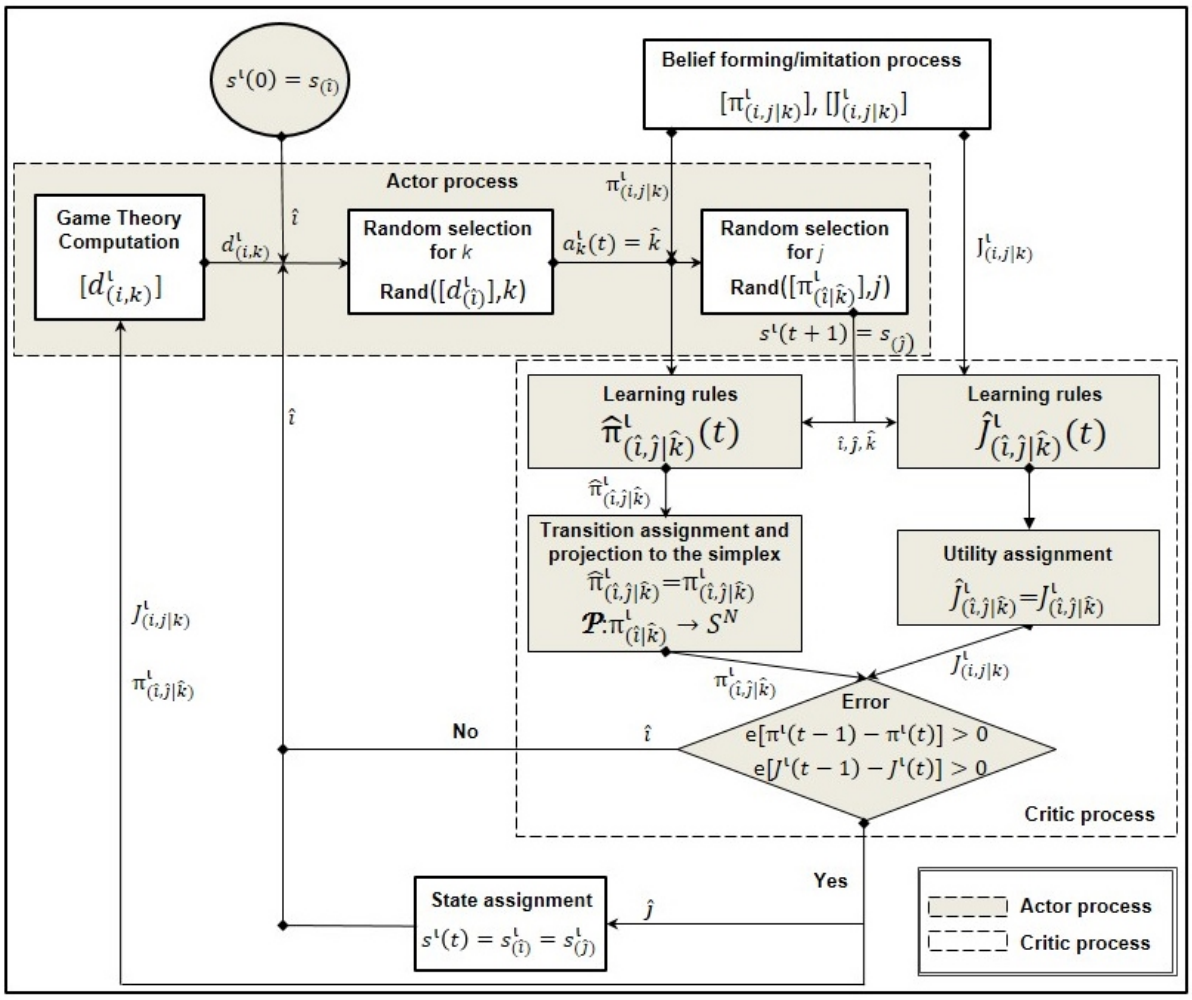


Figure 5.3 Actor critic architecture.

The actor-critic architecture of the RL is illustrated in Figure 5.3. The initial $\pi^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)$ and $J^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t)$ are given as a result of applying the belief-forming process and the belief-imitating process. The selection of $a(t)$ and $s(t+1)$ is equal as in the Adaptive Primary Learning architecture. The added value is in the use of a game theory for computing the distribution $d^{\iota}_{(k|i)}$ of the strategies in order to obtain the policies for RL process. In this case, the trade-off between

exploration and exploitation of the actions in reinforcement learning is given by the solution of the game. Then, the process for player ι begins with an initial state $s^{\iota}(0) = s^{\iota}_{(\hat{\imath})}$ (for the estimated value $\hat{\imath}$) and it is considered a fixed uniform distribution of the strategies as a solution of the security game given by $[d^{\iota}_{(k|i)}]$. Then, it is selected randomly an action $a^{\iota}(t) = a_{(\hat{k})}$ (for the estimated value $\hat{k}$) from the vector $d^{\iota}_{(\hat{k}|\imath)}$ (for a fixed $\hat{\imath}$). After that, the transition matrix $\Pi^{\iota} = \left[\pi^{\iota}_{(j|i,k)}\right]$ is used to choose randomly the consecutive state $s^{\iota}(t+1) = s^{\iota}_{(\hat{\jmath})}$ (for the estimated value $\hat{\jmath}$) from the vector $\pi^{\iota}_{(j|\hat{\imath},\hat{k})}$ (for a fixed $\hat{\imath}$ and $\hat{k}$). Once $a^{\iota}(t)$ and $s^{\iota}(t+1)$ are selected, the estimating values are updated employing the adaptive module in which are computed the learning rules $\hat{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)$ and $\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t)$. The value-maximizing action at each state is taken whether the actor-critic learning rule $\hat{\pi}^{l}_{(\hat{\jmath}|\hat{\imath},\hat{k})}$ ensures convergence $\left(e\left[\hat{\Pi}^{\iota}(t-1) - \hat{\Pi}^{\iota}(t)\right] > 0\right)$. If the condition of estimated error e is not satisfied, then the selection of the random variables $s^{\iota}_{(i)}$, $s^{\iota}_{(j)}$ and $a^{\iota}_{(k)}$ is carried out again. On the other hand, the distribution $d^{\iota}_{(k|i)}$ of the strategies is computed again using the game theory module until it converges. The proposed architecture converges by the ergodicity restriction for Markov chains imposed on the definition of the game.

5.4 Learning model

The aim of this section is to present how the selection of actions from the best-reply strategy affects the RL process and, also how to learn the transition and cost/reward models.

5.4.1 Exploration and exploitation

One of the most critical problems in RL is that the players need to make decisions as they learn. There are two basic motivations for choosing an action: i) exploitation, which selects an action from the best-reply strategy that leads to an "optimal" cost/reward; and ii) exploration, that selects an action from the best-reply of a strategy that provides support information that will benefit future behavior (to act better in the future).

By employing the exploitation approach the players follow only the best-reply strategy action and they use the next state reached, and the reward received, to adapt its behavior in

the future. However, it can happen that players (defenders/attackers) will get trapped in a local minima/maxima during the RL process. On the other hand, exploration is related to choosing actions of any strategy to learn how to behave in general situations. The tradeoff between exploration and exploitation is that the first assists the players (defenders/attackers) in minimizing/maximizing the costs/rewards and, the second helps the players to learn an overall policy in the long term.

With pure exploration, the players will never obtain the benefits of "optimal" best-reply learning. With pure exploitation, the players will get stuck in a local minima/maxima. A combination of both approaches is needed.

In our case, the trade-off between exploration and exploitation is implicit in the proposed architecture for RL. The Adaptive Primary Learning architecture is assumed to have a fixed uniform distribution of strategies $[d^{\iota}_{(k|i)}] = \frac{1}{M}[1_{(i,k)}]_{i\in[1,N],k\in[1,M]}$, and an initial state $s(0) = s_{(\hat{\imath})}$. We also suppose that the players take advantage of external knowledge represented by some current transition matrix $\Pi^{\iota} = (\pi^{\iota}_{(j|i,1)}, ..., \pi^{\iota}_{(j|i,M)})$ useful when learning a new task from reinforcement. Such evidence is available as historical data. During the dynamics of the learning process, the players interact with their environment and at each stage of the process randomly select an action $a(t) = a_{(\hat{k})}$ based on the best-reply strategies $d^{\iota}_{(k|i)}$ and a next state $s(t+1) = s_{(\hat{\jmath})}$ given previously $s(t) = s_{(\hat{\imath})}$ and $a(t) = a_{(\hat{k})}$. We ensure that in the long run every action is taken in every state an unbounded number of times and the learning rate is eventually small enough. This is equivalent to pure exploration.

On the other hand, in the actor-critic architecture, we also suppose that the players take advantage of external knowledge represented by the transition matrices $\Pi^{\iota} = (\pi^{\iota}_{(j|i,1)}, ..., \pi^{\iota}_{(j|i,M)})$ and cost/reward matrices $J^{\iota} = (J^{\iota}_{(i,j,1)}, ..., J^{\iota}_{(i,j,M)})$ available as a result of executing the belief-forming process and the belief-imitating process. It is also assumed that the distribution of strategies $[d^{\iota}_{(k|i)}]$ is computed using a game theory model for solving the game, and an initial state $s(0) = s_{(\hat{\imath})}$. As well, during the dynamics of the learning process, the players interact with their environment and at each stage of the RL process randomly select an action $a(t) = a_{(\hat{k})}$ and a state $s(t+1) = s_{(\hat{\jmath})}$. Because $[d^{\iota}_{(k|i)}]$ is the best-reply strategy resulting of computing the

game, the distribution of $d^{\iota}_{(k|i)}$ give importance to the actions $a(t) = a_{(\hat{k})}$ having a high probability to minimize/maximize the costs/rewards of the players (defenders/attackers). Then, the process naturally makes emphasis over the exploitation approach and reduces the exploration rate (because actions have low probability to be randomly selected).

5.4.2 Adaptive module

The proposed approach to reinforcement learning just learns the transition and costs/reward models. It is important to note that in the original Markov game formulation for Stackelberg security games, we usually assume fixed and statics domains not able to be adapted to the environment: fixing a state and action the costs/reward and transition remains always the same. The reason is that the goal is minimizing/maximizing the players' expected costs/reward that depends on the transitions at each state.

However, in Stackelberg security games it is an unrealistic assumption: the costs/reward and transitions received are commonly non-static. For reasoning about more realistic patrol strategies, we need to specify that there is a probability distribution over the possible costs/rewards for any action in any state. Then, for learning in a cost/reward model, we may obtain different costs/rewards at different times for the same action and state. As well as, for the transition model where we may obtain different transitions at different times.

To do this, we introduce a model from experiences that can simply be done by counting the frequency ω of observed experiences. Towards this goal the players use the following variables defined recursively as:

$$\omega^{\iota}_{(i,k)}(t) = \sum_{n=1}^{t} \chi\left(s^{\iota}(n) = s^{\iota}_{(i)}, a^{\iota}(n) = a^{\iota}_{(k)}\right)$$

$$\omega^{\iota}_{(j|i,k)}(t) = \sum_{n=1}^{t} \chi\left(s^{\iota}(n+1) = s^{\iota}_{(j)} | s^{\iota}(n) = s^{\iota}_{(i)}, a^{\iota}(n) = a^{\iota}_{(k)}\right)$$

such that

$$\chi(\mathcal{E}_t) = \begin{cases} 1 & \text{if the event } \mathcal{E} \text{ occurs at interaction } t \\ 0 & \text{otherwise} \end{cases}$$

where $\omega^{\iota}_{(i,k)}(t)$ is the total number of times that the player ι evolves from state i applying action k in the the RL process and, $\omega^{\iota}_{(j|i,k)}$ is the total number of times that player ι evolves from state

i to state j applying action k in the RL process. As well as, for the cost/reward model, we keep a running average of the rewards observed upon taking each action in each state as follows

$$\mathcal{J}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t) = \sum_{n=1}^{t} \xi^{\iota}_{J}(n) \cdot \chi\left(s(n+1) = s_{(\hat{\jmath})} \mid s(n) = s_{(\hat{\imath})}, a(n) = a_{(\hat{k})}\right)$$

such that

$$\xi^{\iota}_{J} := J^{\iota}_{(j|i,k)} + (\Delta_J)\, r, \quad \text{for} \quad \Delta_J \leq J^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})} \quad \text{and} \quad r = \operatorname{rand}([-1,1])$$

where $\mathcal{J}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)$ is the sum over all immediate costs/rewards received after executing action a in state i and stepping to state j, incremented by Δ_J multiplied by a random value r, $-1 \leq r \leq 1$.

The learning rules of the architecture are computed considering the maximum likelihood model where $\frac{0}{0} := 0$. The designing of the adaptive module for the belief-forming process involves the following learning rules:

a. The learning rule for estimating $\pi^{\iota}_{(j|i,k)}$ is given by

$$\hat{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t) = \frac{\omega^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t-1)+1}{\omega^{\iota}_{(\hat{\imath},\hat{k})}(t-1)+1}$$

such that

$$\omega_{(\hat{\jmath}|\hat{\imath},\hat{k})}(0) = 0, \quad \text{and} \quad \omega_{(\hat{\imath},\hat{k})}(0) = 0.$$

b. The learning rule for estimating $J^{\iota}_{(i,j,k)}$ is as follows

$$\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t) = \frac{\mathcal{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t)}{\omega^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)}$$

The designing of the adaptive module for the actor-critic architecture consists of the following learning rules. The definition involves the variable t_0 which is the time required to compute the matrices by the belief-forming process and the belief-imitating process.

a. The learning rule for estimating $\pi^{\iota}_{(j|i,k)}$ is given by

$$\hat{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t) = \frac{\omega^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)}{\omega^{\iota}_{(\hat{\imath},\hat{k})}(t)} = \frac{\omega^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t_0) + \Xi^{\iota,1}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)}{\omega^{\iota}_{(\hat{\imath},\hat{k})}(t_0) + \Xi^{\iota,2}_{(\hat{\imath},\hat{k})}(t)} = \frac{\hat{\pi}^{\iota}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t_0) + \dfrac{1}{\omega^{\iota}_{(\hat{\imath},\hat{k})}(t_0)}\Xi^{\iota,1}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t)}{1 + \dfrac{1}{\omega^{\iota}_{(\hat{\imath},\hat{k})}(t_0)}\Xi^{\iota,2}_{(\hat{\imath},\hat{k})}(t)}$$

where

$$\Xi^{\iota,1}_{(\hat{\jmath}|\hat{\imath},\hat{k})}(t) = \sum_{n=t_0+1}^{t} \chi\left(s^{\iota}(n+1) = s^{\iota}_{(\hat{\jmath})} | s^{\iota}(n) = s^{\iota}_{(\hat{\imath})}, a^{\iota}(n) = a^{\iota}_{(\hat{k})}\right) \cdot$$

$$\chi\left(s^{\iota}(n) = s^{\iota}_{(\hat{\imath})}, a^{\iota}(n) = a^{\iota}_{(\hat{k})}\right)$$

and

$$\Xi^{\iota,2}_{(\hat{\imath},\hat{k})}(t) = \sum_{n=t_0+1}^{t} \chi\left(s^{\iota}(n) = s^{\iota}_{(\hat{\imath})}, a^{\iota}(n) = a^{\iota}_{(\hat{k})}\right).$$

b. The learning rule for estimating $J^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}$ is as follows

$$A(t) = \sum_{n=1}^{t} \xi^{\iota}_{J}(n) \cdot \chi\left(s^{\iota}(n+1) = s^{\iota}_{(\hat{\jmath})} \mid s^{\iota}(n) = s^{\iota}_{(\hat{\imath})}, a^{\iota}(n) = a^{\iota}_{(\hat{k})}\right)$$

$$B(t) = \sum_{n=1}^{t} \chi\left(s^{\iota}(n+1) = s^{\iota}_{(\hat{\jmath})} \mid s^{\iota}(n) = s^{\iota}_{(\hat{\imath})}, a^{\iota}(n) = a^{\iota}_{(\hat{k})}\right)$$

$$\hat{J}^{\iota}_{(\hat{\imath},\hat{\jmath},\hat{k})}(t) = \frac{A(t)}{B(t)} = \frac{J_{(\hat{\imath},\hat{\jmath},\hat{k})}(t_0) + A(n = t_0 + 1)}{\omega_{(\hat{\imath},\hat{\jmath},\hat{k})}(t_0) + B(n = t_0 + 1)}$$

In the end, we have a RL process where if there exist changes in the system the players are able to learn and adapt to the environment. In addition, there is a natural trade-off between exploration and exploitation: in the actor-critic architecture, we make emphasis over the exploitation approach and drastically decrements the exploration rate in a careful way.

We are able to estimate the aleatory variables corresponding to the entry $(\hat{\imath}, \hat{\jmath}, \hat{k})$ of both, the transition and the cost/reward matrices. An advantage of the Adaptive Primary Learning architecture step is that the distribution is easily generated and does not need any computation. The process continues until its convergence (the process converges because it is ergodic).

5.5 Shopping mall security game

This example is suggested to illustrate how the RL method presented in this chapter can be employed to improve the strategy for patrolling four shopping malls located geographically in different areas.

5.5.1 Game overview

Shopping malls are multi-storied structures housing a large number of stores that sell diverse products and services adjoined by pedestrian areas. Families usually choose to visit shopping malls, for a family outing, it is a more convenient option because a parking service is provided. Attributes of the location of the shopping malls have a strong impact on the retailer's strategy. The geographic areas for a shopping mall are selected in terms of the socio-economic characteristics of the residents of the area. Usually, 3 to 5-mile radius generates 50 to 70% of the customers. We will consider four shopping malls (targets) located in different areas: the mall 1 is located in a lower class neighborhood, usually an urban area with low-quality of civil services, mall 2 and 3 are located in a middle-class area and finally the mall 4 is located in an upper-class area where residents are prosperous.

Taking into consideration these properties shopping malls are considered targets for robbery and burglary, involving the theft of property from an individual or the unlawful entry to a structure with the intent to steal or commit a felony. Victims can be or not have to be present. Robberies usually happen during the day and burglaries during the night. The time that burglaries and robberies occur is different which result in higher levels of coverage at all times. Then, protecting shopping malls of the perpetrators is a complicated task.

For representing the Stackelberg security game we will consider four players, two attackers (followers $m = 3, 4$) that try to reach different goals (to commit a crime in a shopping mall) maximizing the expected damage and two defenders (leaders $l = 1, 2$) that try to stop the attackers minimizing his expected loss. Here defenders work cooperatively and the defender 2 imitates and learns from the defender 1 who has more experience in this type of crimes, the attackers also work cooperating and the attacker 4 imitates and learns from the attacker 3. In the dynamics of the game the players take alternate turns: defenders commit first to a strategy and then, the attackers' strategies are played. Let the number of states of each player (shopping malls) $N = 4$ and $M = 2$ the number of actions of each player (burglary and robbery).

The defenders and attackers already have of prior knowledge about the problem, which is recovered from historical data that provide exact information about the crimes occurred in the

geographical area where each mall is located. In addition to economic information, which is strongly correlated with high-risk areas of crime, distance and traffic is an additional important factor that must be considered in solving the patrolling problem. This information is represented in terms of the Markov transition matrices, utility matrices and a uniform distribution that denotes the initial solution of the game.

Let the initial transition matrices for each player be defined as follows

$$\pi^1_{(i,j,1)} = \begin{bmatrix} 0.1971 & 0.3490 & 0.3119 & 0.1421 \\ 0.1348 & 0.3041 & 0.2942 & 0.2669 \\ 0.2118 & 0.3286 & 0.2628 & 0.1967 \\ 0.1866 & 0.3332 & 0.2946 & 0.1857 \end{bmatrix} \quad \pi^1_{(i,j,2)} = \begin{bmatrix} 0.2972 & 0.2825 & 0.2462 & 0.1742 \\ 0.3616 & 0.2237 & 0.2432 & 0.1715 \\ 0.3505 & 0.2135 & 0.2113 & 0.2246 \\ 0.2063 & 0.2917 & 0.2419 & 0.2602 \end{bmatrix}$$

$$\pi^2_{(i,j,1)} = \begin{bmatrix} 0.3968 & 0.1003 & 0.2403 & 0.2626 \\ 0.2006 & 0.1484 & 0.2140 & 0.4370 \\ 0.2943 & 0.2318 & 0.2103 & 0.2637 \\ 0.1740 & 0.2872 & 0.2293 & 0.3096 \end{bmatrix} \quad \pi^2_{(i,j,2)} = \begin{bmatrix} 0.2033 & 0.2456 & 0.2667 & 0.2844 \\ 0.2277 & 0.2527 & 0.2538 & 0.2658 \\ 0.0928 & 0.3023 & 0.3075 & 0.2974 \\ 0.2332 & 0.5366 & 0.1238 & 0.1064 \end{bmatrix}$$

$$\pi^3_{(i,j,1)} = \begin{bmatrix} 0.2739 & 0.3103 & 0.1945 & 0.2213 \\ 0.2512 & 0.1834 & 0.2524 & 0.3130 \\ 0.2853 & 0.4364 & 0.0994 & 0.1789 \\ 0.1571 & 0.2909 & 0.2764 & 0.2756 \end{bmatrix} \quad \pi^3_{(i,j,2)} = \begin{bmatrix} 0.2538 & 0.2403 & 0.2144 & 0.2915 \\ 0.2058 & 0.2482 & 0.2550 & 0.2910 \\ 0.2312 & 0.1876 & 0.3374 & 0.2438 \\ 0.2416 & 0.2613 & 0.2196 & 0.2775 \end{bmatrix}$$

$$\pi^4_{(i,j,1)} = \begin{bmatrix} 0.2175 & 0.2621 & 0.3406 & 0.1798 \\ 0.2905 & 0.3822 & 0.1211 & 0.2062 \\ 0.1852 & 0.1649 & 0.3652 & 0.2847 \\ 0.2267 & 0.3092 & 0.2133 & 0.2507 \end{bmatrix} \quad \pi^4_{(i,j,2)} = \begin{bmatrix} 0.1930 & 0.3877 & 0.1851 & 0.2343 \\ 0.3020 & 0.2704 & 0.3076 & 0.1200 \\ 0.2463 & 0.4126 & 0.1491 & 0.1920 \\ 0.2113 & 0.3330 & 0.2692 & 0.1864 \end{bmatrix}$$

and let the individual utility matrices for each player be defined by

$$U^1_{(i,j,1)} = \begin{bmatrix} 81 & 246 & 219 & 90 \\ 63 & 258 & 54 & 204 \\ 90 & 288 & 192 & 63 \\ 84 & 270 & 240 & 225 \end{bmatrix} \quad U^1_{(i,j,2)} = \begin{bmatrix} 51 & 81 & 171 & 207 \\ 81 & 255 & 72 & 297 \\ 150 & 75 & 138 & 270 \\ 294 & 138 & 153 & 258 \end{bmatrix}$$

$$U^2_{(i,j,1)} = \begin{bmatrix} 360 & 16 & 12 & 4 \\ 28 & 40 & 252 & 32 \\ 36 & 0 & 4 & 20 \\ 380 & 264 & 196 & 4 \end{bmatrix} \quad U^2_{(i,j,2)} = \begin{bmatrix} 6 & 444 & 36 & 0 \\ 48 & 48 & 234 & 12 \\ 198 & 24 & 90 & 54 \\ 300 & 42 & 30 & 150 \end{bmatrix}$$

$$U^3_{(i,j,1)} = \begin{bmatrix} 44 & 14 & 22 & 12 \\ 20 & 0 & 38 & 16 \\ 46 & 56 & 46 & 18 \\ 180 & 10 & 24 & 2 \end{bmatrix} \quad U^3_{(i,j,2)} = \begin{bmatrix} 33 & 0 & 63 & 21 \\ 9 & 39 & 120 & 3 \\ 48 & 9 & 30 & 78 \\ 33 & 51 & 90 & 24 \end{bmatrix}$$

$$U^4_{(i,j,1)} = \begin{bmatrix} 0 & 120 & 4 & 52 \\ 20 & 52 & 72 & 96 \\ 28 & 112 & 56 & 48 \\ 16 & 24 & 32 & 76 \end{bmatrix} \quad U^4_{(i,j,2)} = \begin{bmatrix} 90 & 36 & 81 & 18 \\ 0 & 60 & 3 & 24 \\ 9 & 12 & 75 & 30 \\ 18 & 45 & 72 & 21 \end{bmatrix}$$

5.5.2 RL process for security games

Once the defenders and attackers have the initial information, they begin an iterative reinforcement learning process for security games proposed in the adaptive primary learning architecture. The purpose of this first step is that making use of the exploration properties, the players choose equally among all actions to learn how to behave in general situations, that is, players explore the shops and the area where they are located in order to learn how to move and act in different situations. In fact, players improve their transition and cost/reward matrices by combining their initial information with learning rules and an imitative behavior extracted from other players that are selected as more experienced, defender 2 learns from the defender 1 and the attacker 4 imitates the attacker 3.

When the defenders and attackers finish the adaptive primary learning process, they begin a new iterative procedure, represented by an actor-critic architecture. In this stage defenders and attackers improve their transition and cost/reward matrices (obtained from the adaptive primary learning architecture) through the application of new learning rules and the calculation of the strategies of the Stackelberg security game employing the extraproximal method.

Remark 5.1 *It is clear that attackers' actions are dependent on their past successes and failures. The proposed adaptive model captures this adaptive nature of the attackers' behavior by modifying probabilities in the transition matrices and the values in the utility matrices. Then, the RL process estimates the new matrices and recalculate the resulting strategies of the security game every time the behavior of the attackers or the environment change.*

The resulting estimations of the transition matrices of the RL process are the following:

$$\hat{\pi}^1_{(i,j,1)} = \begin{bmatrix} 0.1907 & 0.3449 & 0.3447 & 0.1197 \\ 0.1412 & 0.2959 & 0.3114 & 0.2515 \\ 0.2216 & 0.3049 & 0.2705 & 0.2030 \\ 0.1837 & 0.3860 & 0.2656 & 0.1647 \end{bmatrix} \quad \hat{\pi}^1_{(i,j,2)} = \begin{bmatrix} 0.3220 & 0.2813 & 0.2449 & 0.1518 \\ 0.3434 & 0.1820 & 0.2875 & 0.1872 \\ 0.3433 & 0.1924 & 0.2474 & 0.2169 \\ 0.2386 & 0.2973 & 0.2275 & 0.2366 \end{bmatrix}$$

$$\hat{\pi}^2_{(i,j,1)} = \begin{bmatrix} 0.4388 & 0.1052 & 0.2288 & 0.2272 \\ 0.1887 & 0.1939 & 0.2606 & 0.3568 \\ 0.2920 & 0.2402 & 0.1878 & 0.2800 \\ 0.1665 & 0.2927 & 0.2631 & 0.2777 \end{bmatrix} \quad \hat{\pi}^2_{(i,j,2)} = \begin{bmatrix} 0.2023 & 0.2367 & 0.2957 & 0.2652 \\ 0.2186 & 0.2606 & 0.2782 & 0.2427 \\ 0.1271 & 0.2506 & 0.3205 & 0.3018 \\ 0.2142 & 0.4197 & 0.1576 & 0.2085 \end{bmatrix}$$

$$\hat{\pi}^3_{(i,j,1)} = \begin{bmatrix} 0.2811 & 0.3041 & 0.1723 & 0.2426 \\ 0.2649 & 0.1729 & 0.2393 & 0.3229 \\ 0.2675 & 0.4784 & 0.0854 & 0.1686 \\ 0.1930 & 0.2692 & 0.2894 & 0.2484 \end{bmatrix} \quad \hat{\pi}^3_{(i,j,2)} = \begin{bmatrix} 0.2728 & 0.2576 & 0.1841 & 0.2855 \\ 0.1983 & 0.2218 & 0.2728 & 0.3071 \\ 0.2428 & 0.1778 & 0.3429 & 0.2366 \\ 0.2646 & 0.2337 & 0.2330 & 0.2686 \end{bmatrix}$$

$$\hat{\pi}^4_{(i,j,1)} = \begin{bmatrix} 0.1760 & 0.3492 & 0.3093 & 0.1655 \\ 0.2684 & 0.3365 & 0.2072 & 0.1880 \\ 0.1924 & 0.1864 & 0.3307 & 0.2905 \\ 0.1921 & 0.3153 & 0.2232 & 0.2694 \end{bmatrix} \quad \hat{\pi}^4_{(i,j,2)} = \begin{bmatrix} 0.2489 & 0.2926 & 0.2645 & 0.1939 \\ 0.2911 & 0.2607 & 0.2849 & 0.1633 \\ 0.2450 & 0.3278 & 0.1839 & 0.2433 \\ 0.2400 & 0.3325 & 0.2391 & 0.1883 \end{bmatrix}$$

And the resulting utility matrices for each player are as follows

$$\hat{U}^1_{(i,j,1)} = \begin{bmatrix} 47.822 & 67.034 & 22.627 & 8.032 \\ 21.748 & 96.729 & 16.847 & 180.253 \\ 25.835 & 281.720 & 100.239 & 35.541 \\ 125.741 & 118.075 & 179.806 & 128.743 \end{bmatrix} \quad \hat{U}^1_{(i,j,2)} = \begin{bmatrix} 27.715 & 26.283 & 4.514 & 283.899 \\ 20.333 & 43.550 & 54.349 & 157.131 \\ 101.889 & 39.178 & 190.407 & 57.975 \\ 69.107 & 31.595 & 24.990 & 48.384 \end{bmatrix}$$

$$\hat{U}^2_{(i,j,1)} = \begin{bmatrix} 104.944 & 11.107 & 4.259 & 0.259 \\ 21.793 & 14.472 & 168.325 & 36.846 \\ 23.551 & 0 & 2.253 & 16.483 \\ 148.615 & 246.275 & 156.610 & 2.731 \end{bmatrix} \quad \hat{U}^2_{(i,j,2)} = \begin{bmatrix} 3.037 & 57.997 & 21.857 & 0 \\ 10.547 & 17.522 & 47.336 & 1.931 \\ 47.569 & 8.045 & 15.247 & 27.541 \\ 30.802 & 5.786 & 10.441 & 102.284 \end{bmatrix}$$

$$\hat{U}^3_{(i,j,1)} = \begin{bmatrix} 19.881 & 3.507 & 10.956 & 1.826 \\ 4.687 & 0 & 4.515 & 13.516 \\ 30.828 & 46.227 & 4.657 & 20.855 \\ 113.018 & 3.831 & 11.719 & 0.963 \end{bmatrix} \quad \hat{U}^3_{(i,j,2)} = \begin{bmatrix} 10.872 & 0 & 63.043 & 5.121 \\ 1.450 & 24.269 & 153.231 & 4.503 \\ 34.276 & 4.058 & 12.099 & 30.349 \\ 40.081 & 25.870 & 9.222 & 10.912 \end{bmatrix}$$

$$\hat{U}^4_{(i,j,1)} = \begin{bmatrix} 0 & 58.833 & 3.203 & 11.789 \\ 9.585 & 39.761 & 100.258 & 10.602 \\ 20.336 & 29.004 & 33.262 & 31.653 \\ 13.177 & 1.855 & 7.098 & 103.302 \end{bmatrix} \quad \hat{U}^4_{(i,j,2)} = \begin{bmatrix} 14.555 & 27.783 & 56.755 & 15.630 \\ 0 & 3.968 & 6.252 & 16.363 \\ 2.531 & 3.553 & 18.185 & 27.441 \\ 25.482 & 26.219 & 74.518 & 11.881 \end{bmatrix}$$

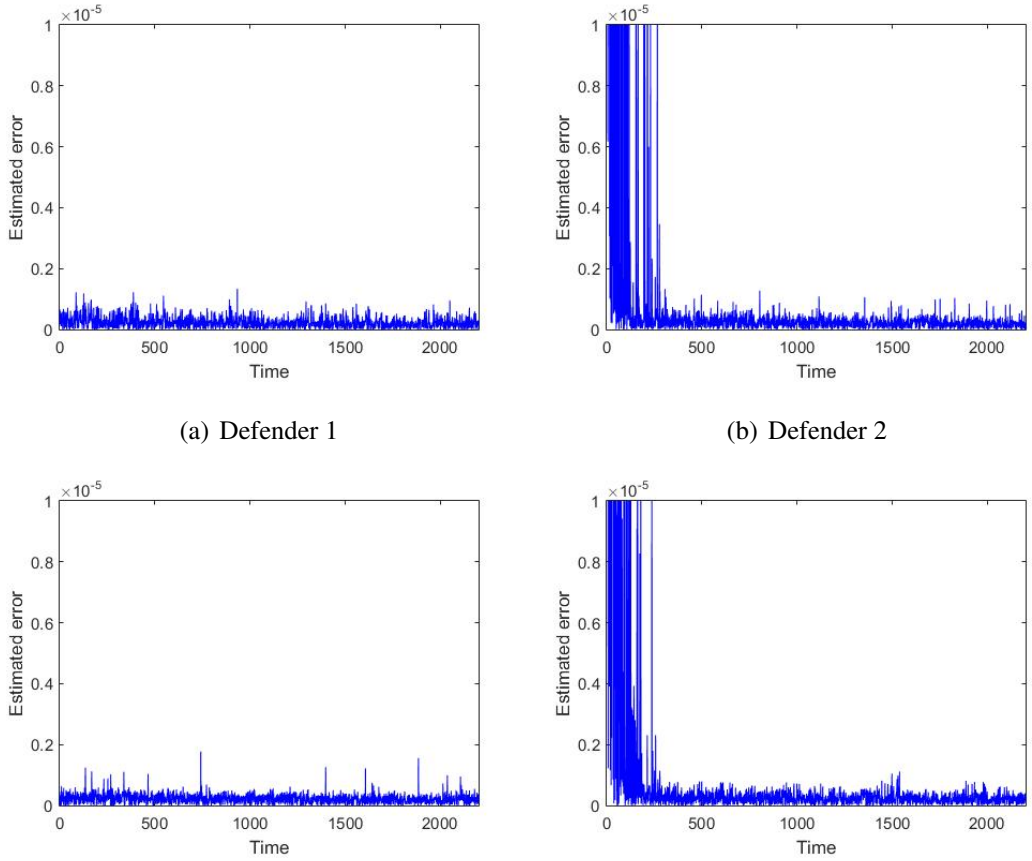
Figure 5.4 shows the estimation function error of the transition matrices which has a decreasing behavior, we can see in these graphs that the estimation error is greater for less experienced players, defender 2 and attacker 2 need more time to improve their transition matrices. Figure 5.5 shows the estimation function error of the utility matrices which have a decreasing behavior.

The solution of the game is obtained employing the Stackelberg game formulation with Markov chains, the game is solved making use of the extraproximal method from which we obtain the convergence of the strategies of defenders and attackers.

$$d^1 = \begin{bmatrix} 0.3523 & 0.6477 \\ 0.5743 & 0.4257 \\ 0.6821 & 0.3179 \\ 0.7388 & 0.2612 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.5463 & 0.4537 \\ 0.6581 & 0.3419 \\ 0.4919 & 0.5081 \\ 0.8213 & 0.1787 \end{bmatrix}$$

$$d^3 = \begin{bmatrix} 0.7258 & 0.2742 \\ 0.7169 & 0.2831 \\ 0.2993 & 0.7007 \\ 0.2656 & 0.7344 \end{bmatrix} \quad d^4 = \begin{bmatrix} 0.7058 & 0.2942 \\ 0.2305 & 0.7695 \\ 0.2732 & 0.7268 \\ 0.3464 & 0.6536 \end{bmatrix}$$

At the end of the adaptive process, combining exploration and exploitation, players select an action from the best-reply strategy that leads to an optimal cost or reward. Figure 5.6 show



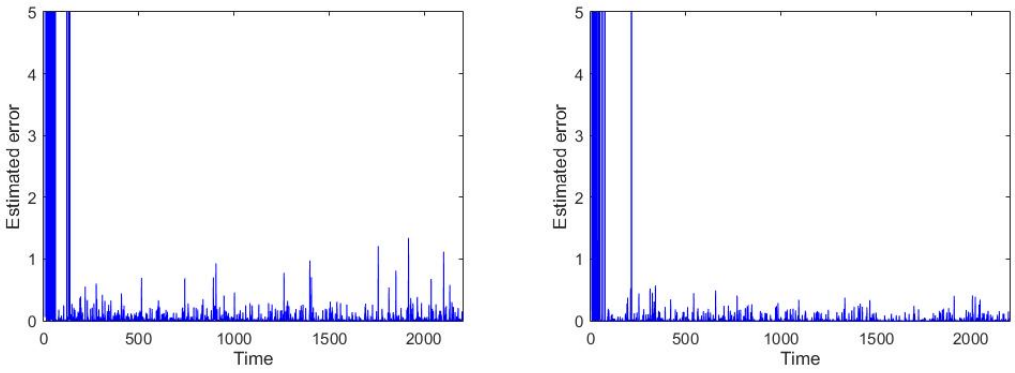
(a) Defender 1 (b) Defender 2

(c) Attacker 1 (d) Attacker 2

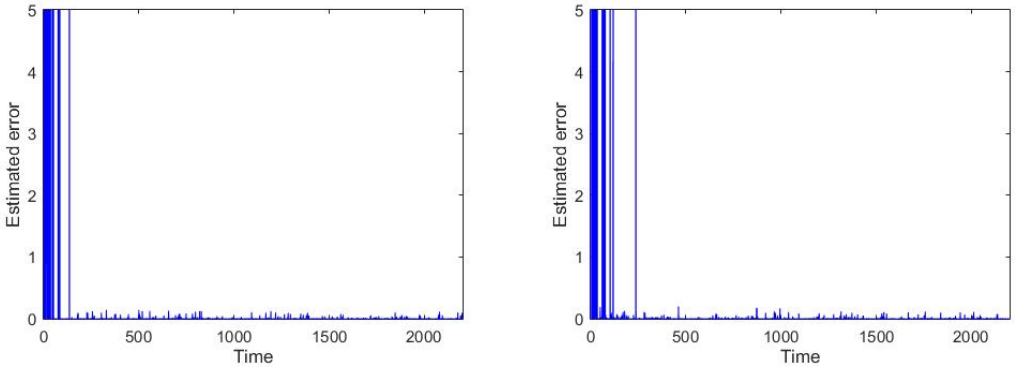
Figure 5.4 Estimation function error of the transition matrices.

the convergence of the strategies of defenders and Figure 5.7 show the convergence of the strategies of attackers.

Remark 5.2 *For modeling a security real-world application, it is necessary to compute the values of the parameters of the environment given by the transition probabilities and the cost functions. Transition probabilities can be computed recovering statistical data related to occurrences of crimes. Nevertheless, it is impossible to recover all the data necessary for computing the exact values for the transition matrices. In addition, the cost functions are typically hand-tuned by experts in the security field until it is acquired a satisfactory value, which can result in an undesired process. Then, the RL process plays a fundamental role for computing the transition and cost matrices very close to the real values in Stackelberg security games.*

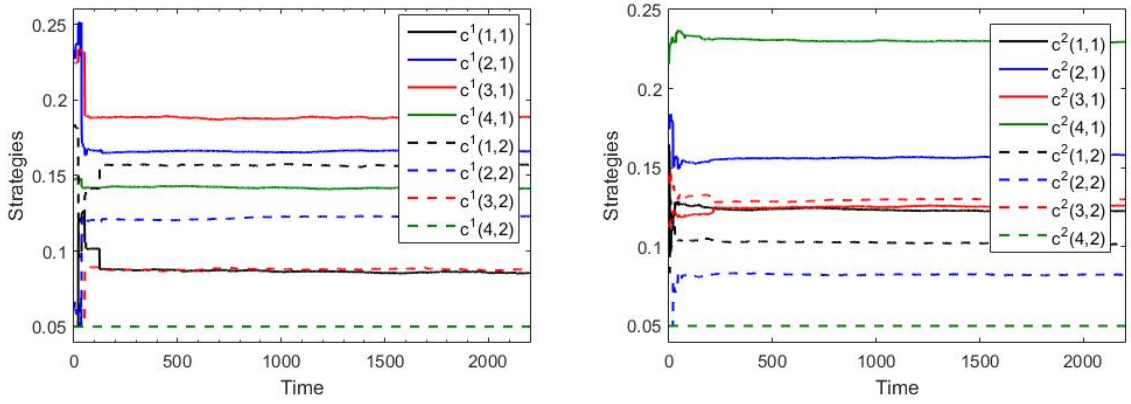


(a) Defender 1 (b) Defender 2



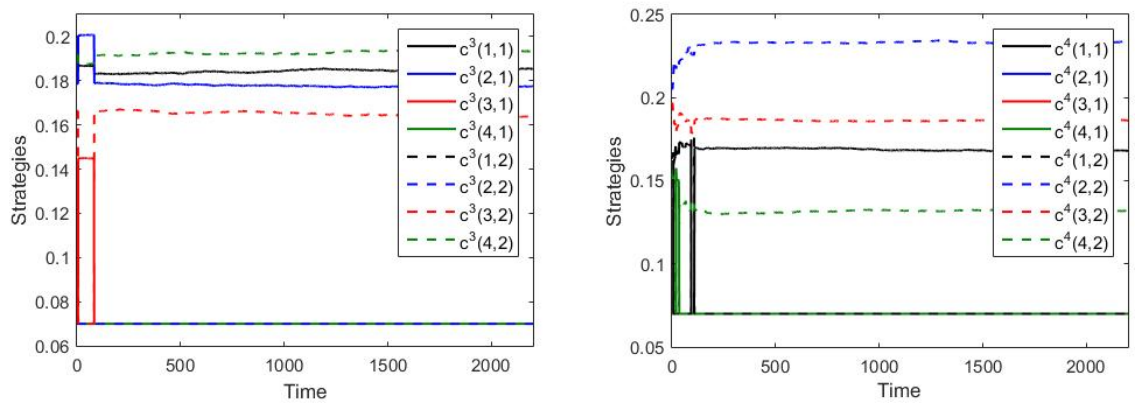
(c) Attacker 1 (d) Attacker 2

Figure 5.5 Estimation function error of the utility matrices.



(a) Defender 1 (b) Defender 2

Figure 5.6 The convergence of the defenders strategies.



(a) Attacker 1 (b) Attacker 2

Figure 5.7 The convergence of the attackers strategies.

5.5.3 Realization of the security game

The main objective of the defenders is to provide protection and security to people who visit the mall as well as to the stores, and deciding to act taking into account the infrastructure, alarms and security systems of each mall according to the location area. Taking into account that the shopping mall 1 is located in a lower-class area that is usually unprotected during the night, defender 1 prefers to protect the mall from a possible burglary ($d^1_{(1,2)} = 0.6477$) while defender 2 has almost the same preference to proceed during the day ($d^2_{(1,1)} = 0.5463$) or at night ($d^2_{(1,2)} = 0.4537$); on the other hand, attackers know that this mall is more protected at night so they decide to commit robberies ($d^3_{(1,1)} = 0.7258$ and $d^4_{(1,1)} = 0.7058$). For shopping malls 2 and 3 located in a middle-class area, defenders have similar preferences about work at day or night, the defender 1 has almost the same preference for protecting mall 2 at day ($d^1_{(2,1)} = 0.5743$) or overnight ($d^1_{(2,2)} = 0.4257$) while chooses to protect mall 3 from a possible robbery ($d^1_{(3,1)} = 0.6821$), as well as, the defender 2 decides to protect the mall 2 from possible robberies ($d^2_{(2,1)} = 0.6581$) and for mall 3 has almost the same preference to protect it at day ($d^2_{(3,1)} = 0.4919$) or during the night ($d^2_{(3,2)} = 0.5081$). For mall 2, the attacker 3 chooses to commit a robbery ($d^3_{(2,1)} = 0.7169$) while the attacker 4 prefers to commit a burglary ($d^4_{(2,2)} = 0.7695$), and for mall 3 both attackers decide to act during the night ($d^3_{(3,2)} = 0.7007$ and $d^4_{(3,2)} = 0.7268$). Because the shopping mall 4 is located in a high-class area, defenders

consider that the mall and the whole area have efficient night-time security systems, so both defenders decide to protect during the day to provide security and tranquility to people who visit the mall ($d^1_{(4,1)} = 0.7388$ and $d^2_{(4,1)} = 0.8213$). Attackers observe that this mall is more protected during the day, and even if committing robberies means high profits because wealthy people visit the mall, they prefer not to take the risk and act overnight ($d^3_{(4,2)} = 0.7344$ and $d^4_{(4,2)} = 0.6536$).

With the strategies calculated from the RL process, we considered a Random walk model, such that each player either staying put or moving to another state, i.e., defenders and attackers decide to remain patrolling the same mall or move to one of the other malls that must protect, defenders can capture attackers if they implement the appropriate strategy (to protect at day or overnight). Here the game is over when the attackers are caught (5.1).

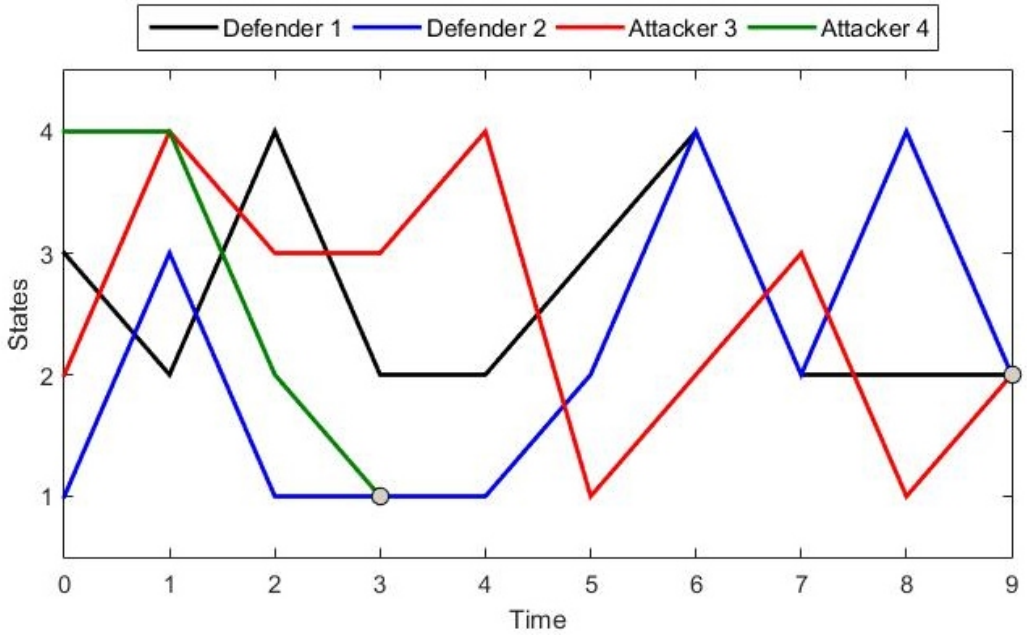


Figure 5.8 Random Walk.

The initial position (state) of each player are as follows: defender 1 is patrolling the shopping mall 3 and defender 2 is patrolling the mall 1, while attacker 3 set the mall 2 as his target and attacker 4 plans to commit a crime in the mall 4, that is, $s^1_{(i)}(0) = 3,\ s^2_{(i)}(0) = 1,\ s^3_{(i)}(0) = 2,\ s^4_{(i)}(0) = 4$. Applying the patrol schedule in Algorithm 1 (Table 5.1) of the security game we obtain that the attacker 4 is caught at shopping mall 1 after 3 steps by the defender 2, to

capture the attacker 3 the defenders work cooperatively and both catch him after 9 steps while he tries to commit a crime in the mall 2, so with both attackers captured the game is over (see Figure 5.8).

Part II

The bargaining game

Chapter 6

The Nash bargaining solution

6.1 Introduction

The starting point of bargaining theory is the Nash formulation [61] who presented this situation as a new treatment of a classical economic problem. A two-player bargaining situation involves two individuals who have the opportunity to collaborate for mutual benefit, each player has to make in turn a proposal, after one player has made an offer, the other must decide either to accept it, or to reject it and continue the bargaining process.

Nash [61] idealized the bargaining problem by assuming that the two individuals are highly rational, that they are equal in bargaining skill, and that each has full knowledge of the tastes and preferences of the other. Nash proved that a solution for all convex bargaining problems always maximizes the product of individuals' utilities under four axioms that describe the behavior of players and provide a unique solution: Symmetry, Pareto optimality, Invariance with respect to positive affine transformations, and Independence of irrelevant alternatives; however, this last axiom came under criticism because empirical evidence shows that it is not often satisfied even in individual decision-making.

Since Nash [61], a bargaining problem is usually defined as a pair (Ψ, ϕ) where Ψ is a compact and convex subset of $\mathbb{R}^2$ containing both ϕ and a point that strictly dominates ϕ. Point $\psi = (\psi^1, \psi^2) \in \Psi$ represents levels of utility for players 1 and 2 that can be reached by an outcome of the game which is feasible for the two players when they do cooperate, and $\phi = (\phi, \phi)$ is the level of utility that players receive if the two players do not cooperate with

each other (status-quo outcome). The goal is to find an outcome in Ψ which will be agreeable to both players.

Applications for bargaining situations beyond economic models, the latest applications take place also in the communications area where some problems are formulated as a two-person bargaining problem. For example, Zhang et al. [112] proposed a cooperation strategy among rational nodes in a wireless cooperative relaying network as an effort to solve two basic problems, when to cooperate and how to cooperate. Another example is proposed by Han et al. [36] where a fair scheme to allocate subcarrier, rate, and power for multiuser orthogonal frequency-division multiple-access systems is proposed. The problem here was to maximize the overall system rate, under each user's maximal power and minimal rate constraints, while considering the fairness among users. This approach considers a new fairness criterion, which is a generalized proportional fairness based on Nash bargaining solutions and coalitions. On the other hand, Birkeland and Tungodden [14] studied the role of fairness motivation in bargaining showing that the bargaining outcome is sensitive to the fairness motivation of the two individuals, unless they both consider an equal division fair. A bargaining between two strongly fairness motivated individuals who have different views about what represents a fair division may end in disagreement.

6.2 The Nash bargaining model

Nash bargaining solution is based on a model in which the players are assumed to negotiate on which point of the set of feasible payoffs $\Psi \subset \mathbb{R}^{\mathfrak{n}}$ will be agreed upon and realized by concerted actions of the members of the grand coalition $l = 1, ..., \mathfrak{n}$. A pivotal element of the model is a fixed disagreement vector $\phi \in \mathbb{R}^{\mathfrak{n}}$ which plays the role of a deterrent. If negotiations break down and no agreement is reached, then the player are committed to the disagreement point. Thus the whole bargaining problem $\mathcal{B}$ will be concisely given by the pair $\mathcal{B} = (\Psi, \phi)$, this is called the condensed form of the bargaining problem (see [29, 61]).

A bargaining problem can be derived from the normal form of an $\mathfrak{n}$-person game $G = \{C^1, ..., C^{\mathfrak{n}}; \psi^1, ..., \psi^{\mathfrak{n}}\}$ in a natural way. The set of all feasible payoffs (outcomes) is defined

as

$$\Phi = \left\{ \psi \mid \psi = \left(\psi^1(c), ..., \psi^{\mathfrak{n}}(c) \right) \right\},$$

where $c \in C$ and $C = C^1 \times ... \times C^{\mathfrak{n}}$

Given a disagreement vector $\phi \in \mathbb{R}^{\mathfrak{n}}$, $\mathcal{B} = (\Phi, \phi)$ is a bargaining problem in condensed form. We can derive another bargaining problem $\mathcal{B} = (\Psi, \phi)$ from G by extending the set of feasible outcomes Φ to its convex hull Ψ. Notice that any element $\varphi \in \Psi$ can be represented as

$$\varphi = \sum_{l=1}^{\mathfrak{n}} \lambda^l \psi^l,$$

where $\psi = (\psi^1(c), ..., \psi^{\mathfrak{n}}(c)),\ c \in)$, $\lambda^l \geq 0$ for all player l and $\sum_{l=1}^{\mathfrak{n}} \lambda^l = 1$.

The payoff vector φ can be realized by playing the strategies c^l with probability λ^l, and so φ is the expected payoff of the players. Thus, when the players face the bargaining problem $\mathcal{B}$ the question is, which point of Ψ should be selected taking into account the different position and strength of the players that is reflected in the set Ψ of extended payoffs and the disagreement point ϕ.

Nash approached this problem by assigning a one-point solution to $\mathcal{B}$ in an axiomatic manner. Let $\mathcal{B}$ denote the set of all pairs (Ψ, ϕ) such that

1. $\Psi \subset \mathbb{R}^{\mathfrak{n}}$ is compact, convex;

2. there exists at least one $\psi \in \Psi$ such that $\psi > \phi$.

A Nash solution to the bargaining problem is a function $f : \mathcal{B} \to \mathbb{R}^{\mathfrak{n}}$ such that $f(\Psi, \phi) \in \Psi$. We shall confine ourselves to functions satisfying the following axioms (see [61, 29, 60]).

1. Feasibility: $f(\Psi, \phi) \in \Psi$.

2. Rationality: $f(\Psi, \phi) \geq \phi$.

3. Pareto Optimality: For every $(\Psi, \phi) \in \mathcal{B}$ there is $\psi \in \Psi$ such that $\psi \geq f(\Psi, \phi)$ and imply $\psi = f(\Psi, \phi)$.

4. Symmetry: If for a bargaining problem $(\Psi, \phi) \in \mathcal{B}$, there exist indices i, j such that $\varphi = (\varphi^1, ..., \varphi^{\mathfrak{n}}) \in \Psi$ if and only if $\bar{\varphi} = (\bar{\varphi}^1, ..., \bar{\varphi}^{\mathfrak{n}}) \in \Psi$, $(\bar{\varphi}^l = \varphi^l, l \neq i, l \neq j, \bar{\varphi}^i = \varphi^j, \bar{\varphi}^j = \varphi^i)$ and $\phi^i = \phi^j$ for $\phi = (\phi^1, ..., \phi^{\mathfrak{n}})$, then $f^i = f^j$ for the solution vector $f(\Psi, \phi) = (f^1, ..., f^{\mathfrak{n}})$.

5. Invariance with respect to affine transformations of utility: Let $\alpha^l > 0, \beta^l, (l = 1, ..., \mathfrak{n})$ be arbitrary constants and let

$$\phi' = (\alpha^1 \phi^1 + \beta^1, ..., \alpha^{\mathfrak{n}} \phi^{\mathfrak{n}} + \beta^{\mathfrak{n}}) \quad \text{with} \quad \phi = (\phi^1, ..., \phi^{\mathfrak{n}})$$

and

$$\Psi' = (\alpha^1 \varphi^1 + \beta^1, ..., \alpha^{\mathfrak{n}} \varphi^{\mathfrak{n}} + \beta^{\mathfrak{n}}) : (\varphi^1, ..., \varphi^{\mathfrak{n}}) \in \Psi.$$

Then $f(\Psi', \phi') = (\alpha^1 f^1 + \beta^1, ..., \alpha^{\mathfrak{n}} f^{\mathfrak{n}} + \beta^{\mathfrak{n}})$, where $f(\Psi, \phi) = (f^1, ..., f^{\mathfrak{n}})$.

6. Independence of irrelevant alternatives: If (Ψ, ϕ) and (Θ, ϕ) are bargaining pairs such that $\Psi \subset \Theta$ and $f(\Theta, \phi) \in \Psi$, then $f(\Theta, \phi) = f(\Psi, \phi)$.

Theorem 6.1 *There is a unique function f satisfying axioms 1-6, furthermore for all $(\Psi, \phi) \in \mathcal{B}$, the vector $f(\Psi, \phi) = (f^1, ..., f^{\mathfrak{n}}) = (\psi^1, ..., \psi^{\mathfrak{n}})$ is the unique solution of the optimization problem*

$$\begin{aligned} &\textit{maximize} && g(\psi) = \prod_{l=1}^{\mathfrak{n}} (\psi^l - \phi^l) \\ &\textit{subject to} && \psi \in \Psi,\ \psi \geq \phi \end{aligned} \tag{6.1}$$

The objective function of problem in eq. (6.1) is usually called the Nash product.

Proof. See [29] ■

Remark 6.2 *There are exactly two solutions satisfying axioms 1, 2, 4, 5, and 6. One is the Nash's solution and the other is the disagreement solution.*

For the next conjectures consider a bargaining problem as a pair (Ψ, ϕ) where $\Psi \subset \mathbb{R}^2$ and $\phi \in \mathbb{R}^2$.

Corollary 6.3 *[60] The Pareto frontier Ω^e of the set Ψ is the graph of a concave function, denoted by h, whose domain is a closed interval $\mathcal{B} \subseteq R$. Furthermore, there exists $\psi^1 \in \mathcal{B}$ such that $\psi^1 > \phi^1$ and $h(\psi^1) > \phi^2$.*

Corollary 6.4 *[60] The set Ω^w of weakly Pareto efficient utility pairs is closed.*

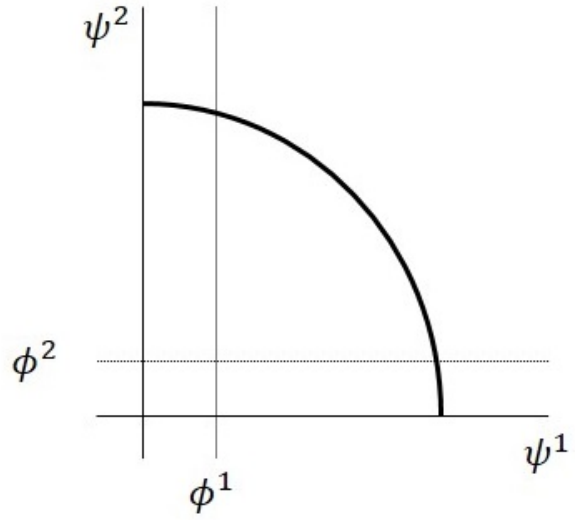


Figure 6.1 Pareto front.

Considering a two-person bargaining problem. Denote the disagreement cost of each player as ϕ^1 and ϕ^2, and the solution for the Nash bargaining problem as the point (ψ^1, ψ^2), therefore the Pareto front is as the Figure 6.1.

6.2.1 Formulation of the problem

Stated in general terms, a $\mathfrak{n}$-person bargaining problem is a situation in which $\mathfrak{n}$ players have a common interest to cooperate, but have conflicting interests over exactly how to cooperate. This process involves the players making offers and counteroffers to each other.

Consider a $\mathfrak{n}$-person bargaining problem [102, 95]. Let us denote the disagreement utility that depends on the strategies $c^l_{(i,k)}$ as $\phi^l(c^1, ..., c^{\mathfrak{n}})$ for each player $(l = 1, ..., \mathfrak{n})$, and the solution for the Nash bargaining problem as the point $(\psi^1, ..., \psi^{\mathfrak{n}})$. Following the eq. (2.5) the

utilities ψ^l, in the same way that the disagreement utilities, are for Markov chains as follows

$$\psi^l = \psi^l\left(c^1, ..., c^{\mathfrak{n}}\right) := \sum_{i}^{N}\sum_{k}^{M} W^l_{(i,k)} \prod_{l=1}^{\mathfrak{n}} c^l_{(i,k)} \tag{6.2}$$

where the matrices $W^l_{(i,k)}$ represent the behavior of each player. This point is better than the disagreement point, therefore must satisfy that $\psi^l > \phi^l$.

The process to solve the bargaining problem consists of two main steps, firstly to find the disagreement point we define it as the Nash equilibrium point of the problem (see [62]), this formulation is detailed in Chapter 2; while for the solution of the bargaining process we follow the model presented by Nash [61]. The function for finding the solution to the Nash Bargaining problem is

$$g(c^1, ..., c^{\mathfrak{n}}) = \prod_{l=1}^{\mathfrak{n}} (\psi^l - \phi^l)^{\alpha^l \chi(\psi^l > \phi^l)} \tag{6.3}$$

where $\alpha^l \geq 0$ and $\sum_{l=1}^{\mathfrak{n}} \alpha^l = 1$, $(l = 1, .., \mathfrak{n})$, which are weighting parameters for each player. We can rewrite (6.3) for purposes of implementation as follows

$$\tilde{g}(c^1, ..., c^{\mathfrak{n}}) = \sum_{l=1}^{\mathfrak{n}} \alpha^l \chi(\psi^l > \phi^l) \ln(\psi^l - \phi^l)$$

Thus, the strategy x^*, which is the vector $x^* = (c^1, ..., c^{\mathfrak{n}}) \in X_{\text{adm}} := \bigotimes_{l=1}^{\mathfrak{n}} C^l_{\text{adm}}$ satisfying the simplex (2.7) and ergodicity (2.8) restrictions, is the solution for the Nash bargaining problem

$$x^* \in \operatorname{Arg}\max_{x \in X_{\text{adm}}} \left\{\tilde{g}(c^1, ..., c^{\mathfrak{n}})\right\}$$

Applying the Lagrange principle,

$$\mathcal{L}(x, \mu, \eta) = \tilde{g}(c^1, ..., c^{\mathfrak{n}}) - \sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(c^l) - \sum_{l=1}^{\mathfrak{n}}\sum_{i,k=1}^{N,M} \eta^l \left(c^l_{(i,k)} - 1\right)$$

The approximative solution obtained by the Tikhonov's regularization is given by

$$x^*, \mu^*, \eta^* = \arg\max_{x \in X_{\text{adm}}} \min_{\mu, \eta \geq 0} \mathcal{L}_\delta(x, \mu, \eta)$$

where

$$\begin{aligned} \mathcal{L}_\delta(x, \mu, \eta) = \tilde{g}(c^1, ..., c^{\mathfrak{n}}) - \sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(c^l) - \\ \sum_{l=1}^{\mathfrak{n}}\sum_{i,k=1}^{N,M} \eta^l \left(c^l_{(i,k)} - 1\right) - \tfrac{\delta}{2}\left(\|x\|^2 - \|\mu\|^2 - \|\eta\|^2\right) \end{aligned} \tag{6.4}$$

Notice that the Lagrange function (6.4) satisfies the saddle-point condition, namely, for all $x \in X_{\text{adm}}$ and $\mu, \eta \geq 0$ we have

$$\mathcal{L}_\delta(x_\delta, \mu_\delta^*, \eta_\delta^*) \leq \mathcal{L}_\delta(x_\delta^*, \mu_\delta^*, \eta_\delta^*) \leq \mathcal{L}_\delta(x_\delta^*, \mu_\delta, \eta_\delta)$$

6.2.2 The proximal format

In the proximal format [5] the relation (6.4) can be expressed as

$$\begin{aligned} \mu_\delta^* &= \arg\min_{\mu \geq 0} \left\{ \tfrac{1}{2}\|\mu - \mu_\delta^*\|^2 + \gamma \mathcal{L}_\delta(x_\delta^*, \mu, \eta_\delta^*) \right\} \\ \eta_\delta^* &= \arg\min_{\eta \geq 0} \left\{ \tfrac{1}{2}\|\eta - \eta_\delta^*\|^2 + \gamma \mathcal{L}_\delta(x_\delta^*, \mu_\delta^*, \eta) \right\} \\ x_\delta^* &= \arg\max_{x \in X} \left\{ -\tfrac{1}{2}\|x - x_\delta^*\|^2 + \gamma \mathcal{L}_\delta(x, \mu_\delta^*, \eta_\delta^*) \right\} \end{aligned} \tag{6.5}$$

where the solutions x_δ^*, μ_δ^* and η_δ^* depend on the parameters $\delta > 0$ and $\gamma > 0$.

6.2.3 The Extraproximal method

We design the method for the static Nash bargaining game in a general format with some fixed admissible initial values ($x_0 \in X_{\text{adm}}$ and $\mu_0, \eta_0 \geq 0$), considering that we want to maximize the function as follows:

1. The *first half-step* (prediction):

$$\begin{aligned} \bar{\mu}_n &= \arg\max_{\mu \geq 0} \left\{ -\tfrac{1}{2}\|\mu - \mu_n\|^2 - \gamma \mathcal{L}_\delta(x_n, \mu, \eta_n) \right\} \\ \bar{\eta}_n &= \arg\max_{\eta \geq 0} \left\{ -\tfrac{1}{2}\|\eta - \eta_n\|^2 - \gamma \mathcal{L}_\delta(x_n, \bar{\mu}_n, \eta) \right\} \\ \bar{x}_n &= \arg\max_{x \in X} \left\{ -\tfrac{1}{2}\|x - x_n\|^2 + \gamma \mathcal{L}_\delta(x, \bar{\mu}_n, \bar{\eta}_n) \right\} \end{aligned} \tag{6.6}$$

2. The *second half-step* (basic)

$$\begin{aligned} \mu_{n+1} &= \arg\max_{\mu \geq 0} \left\{ -\tfrac{1}{2}\|\mu - \mu_n\|^2 - \gamma \mathcal{L}_\delta(\bar{x}_n, \mu, \bar{\eta}_n) \right\} \\ \eta_{n+1} &= \arg\max_{\eta \geq 0} \left\{ -\tfrac{1}{2}\|\eta - \eta_n\|^2 - \gamma \mathcal{L}_\delta(\bar{x}_n, \bar{\mu}_n, \eta) \right\} \\ x_{n+1} &= \arg\max_{x \in X} \left\{ -\tfrac{1}{2}\|x - x_n\|^2 + \gamma \mathcal{L}_\delta(x, \bar{\mu}_n, \bar{\eta}_n) \right\} \end{aligned} \tag{6.7}$$

6.2.4 Convergence Analysis

The following theorem presents the convergence conditions of eqs. (6.6) and (6.7) and gives the estimate of its rate of convergence for the Nash bargaining equilibrium. As well, we prove that the extraproximal method converges to an equilibrium point. Let us define the following extended vectors

$$\tilde{x} = x \in \tilde{X}, \quad \tilde{\mu} = \begin{pmatrix} \mu \\ \eta \end{pmatrix} \in \mathbb{R}^+ \times \mathbb{R}^+$$

Then, the regularized Lagrange function can be expressed as

$$\tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu}) := \mathcal{L}_\delta(x, \mu, \eta)$$

The equilibrium point that satisfies (6.5) can be expressed as

$$\tilde{\mu}^*_\delta = \arg\min_{\tilde{\mu}\geq 0} \left\{ \tfrac{1}{2}\|\tilde{\mu} - \tilde{\mu}^*_\delta\|^2 + \gamma\tilde{\mathcal{L}}_\delta(\tilde{x}^*_\delta, \tilde{\mu}) \right\}$$
$$\tilde{x}^*_\delta = \arg\max_{\tilde{x}\in\tilde{X}} \left\{ -\tfrac{1}{2}\|\tilde{x} - \tilde{x}^*_\delta\|^2 + \gamma\tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu}^*_\delta) \right\}$$

Now, introducing the following variables

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} \in \tilde{X} \times \mathbb{R}^+, \quad \tilde{z} = \begin{pmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{pmatrix} \in \tilde{X} \times \mathbb{R}^+$$

and let define the Lagrange function in term of $\tilde{y}$ and $\tilde{z}$

$$L_\delta(\tilde{y}, \tilde{z}) := \tilde{\mathcal{L}}_\delta(\tilde{y}_1, \tilde{z}_2) - \tilde{\mathcal{L}}_\delta(\tilde{z}_1, \tilde{y}_2)$$

For $\tilde{y}_1 = \tilde{x}$, $\tilde{y}_2 = \tilde{\mu}$, $\tilde{z}_1 = \tilde{z}^*_1 = \tilde{x}^*_\delta$ and $\tilde{z}_2 = \tilde{z}^*_2 = \tilde{\mu}^*_\delta$ we have

$$L_\delta(\tilde{y}, \tilde{z}^*) := \tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu}^*_\delta) - \tilde{\mathcal{L}}_\delta(\tilde{x}^*_\delta, \tilde{\mu})$$

In these variables the relation (6.5) can be represented by

$$\tilde{z}^* = \arg\max_{\tilde{y}\in\tilde{X}\times\mathbb{R}^+} \left\{ -\tfrac{1}{2}\|\tilde{y} - \tilde{z}^*\|^2 + \gamma L_\delta(\tilde{y}, \tilde{z}^*) \right\} \tag{6.8}$$

Finally, we have that the extraproximal method can be expressed by

1. First step

$$\hat{z}_n = \arg \max_{\tilde{y} \in \tilde{X} \times \mathbb{R}^+} \left\{ -\tfrac{1}{2} \|\tilde{y} - \tilde{z}_n\|^2 + \gamma L_\delta(\tilde{y}, \tilde{z}_n) \right\} \tag{6.9}$$

2. Second step

$$\tilde{z}_{n+1} = \arg \max_{\tilde{y} \in \tilde{X} \times \mathbb{R}^+} \left\{ -\tfrac{1}{2} \|\tilde{y} - \tilde{z}_n\|^2 + \gamma L_\delta(\tilde{y}, \hat{z}_n) \right\} \tag{6.10}$$

Lemma 6.5 *Let $\tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu})$ be differentiable in $\tilde{x}$ and $\tilde{\mu}$, whose partial derivative with respect to $\tilde{\mu}$ satisfies the Lipschitz condition with positive constant K_0. Then,*

$$\|\tilde{z}_{n+1} - \hat{z}_n\| \leq \gamma K_0 \|\tilde{z}_n - \hat{z}_n\|$$

Lemma 6.6 *Consider the set of regularized solutions of a non-empty game. The behavior of the regularized function is described by the following inequality:*

$$L_\delta(\tilde{y}, \tilde{y}) - L_\delta(\tilde{z}_\delta^*, \tilde{y}) \geq \delta \|\tilde{y} - \tilde{z}_\delta^*\|$$

for all $\tilde{y} \in \{\tilde{y} \mid \tilde{y} \in X \times \mathbb{R}^+\}$ and $\delta > 0$.

Theorem 6.7 (Convergence and rate of convergence) *Let $\tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu})$ be differentiable in $\tilde{x}$ and $\tilde{\mu}$, whose partial derivative with respect to $\tilde{\mu}$ satisfies the Lipschitz condition with positive constant K. Then, for any $\delta > 0$ there exists a small-enough*

$$\gamma_0 = \gamma_0(\delta) < K := \min \left\{ \tfrac{1}{\sqrt{2} K_0}, \tfrac{1 + \sqrt{1 + 2(K_0)^2}}{2(K_0)^2} \right\}$$

where such that, for any $0 < \gamma \leq \gamma_0$, sequence $\{\tilde{z}_n\}$, which generated by the equivalent extraproximal procedure in eqs. (6.9) and (6.10), monotonically converges with exponential rate $r \in (0, 1)$ to a unique equilibrium point $\tilde{z}^*$, i.e.,

$$\|\tilde{z}_n - \tilde{z}^*\|^2 \leq e^{n \ln r} \|\tilde{z}_0 - \tilde{z}^*\|^2$$

where

$$r = 1 + \tfrac{4(\delta\gamma)^2}{1 + 2\delta\gamma - 2\gamma^2 K^2} - 2\delta\gamma < 1$$

and $r_{\min}$ is given by

$$r_{\min} = 1 - \tfrac{2\delta\gamma}{1 + 2\delta\gamma} = \tfrac{1}{1 + 2\delta\gamma}.$$

Please refer to Appendix C for the proofs of these results.

6.3 The disagreement point model

A pivotal element of the model is a fixed disagreement vector (sometimes also called as *status quo* or *threat point*). The player are committed to the disagreement point in the case of failing to reach a consensus on which feasible payoff to realize. introduce the variables

$$x := col\ c^l, \quad \hat{x} := col\ c^{\hat{l}},$$

The strategies of the players $l = \overline{1, \mathfrak{n}}$ are denoted by the vector x, and $\hat{x}$ is a strategy of the rest of the players adjoint to x. For reaching the goal of the game, players try to find a join strategy $x^* = (c^1, c^2, c^3)$ satisfying

$$f(x, \hat{x}) := \sum_{l=1}^{\mathfrak{n}} \left[\phi^l \left(c^l, c^{\hat{l}} \right) - \phi^l \left(\bar{c}^l, c^{\hat{l}} \right) \right]$$

where $\phi^l \left(c^l, c^{\hat{l}} \right)$ is the utility-function of the player l which plays the strategy $c^l \in C^l$ and the rest of the players play the strategy $c^{\hat{l}} \in C^{\hat{l}}$, and $\bar{c}^l$ is the utopia point defined as follows

$$\bar{c}^l := \arg\max_{c^l \in C^l} \phi^l \left(c^l, c^{\hat{l}} \right)$$

The functions $\phi^l \left(c^l, c^{\hat{l}} \right)$ $\left(l = \overline{1, \mathfrak{n}} \right)$ are assumed to be concave in all their arguments.

Property 6.8 *The function $f(x, \hat{x})$ satisfies the Nash condition*

$$\phi^l \left(c^l, c^{\hat{l}} \right) - \phi^l \left(\bar{c}^l, c^{\hat{l}} \right) \leq 0$$

for any $c^l \in C^l$ and all $l = \overline{1, \mathfrak{n}}$

Definition 6.9 *A strategy $x^* \in X_{adm} := \bigotimes_{l=1}^{\mathfrak{n}} C^l_{adm}$ (Restrictions 2.7 and 2.8) is said to be a Nash equilibrium if*

$$x^* \in \ Arg \max_{x \in X_{adm}} \ \{f(x, \hat{x})\}$$

Applying the regularized Lagrange principle we have the solution for the Nash equilibrium

$$x^*, \hat{x}^*, \mu^*, \eta^* = \arg \max_{x \in X, \hat{x} \in \hat{X}} \min_{\mu, \eta \geq 0} \ \mathcal{L}_{\theta, \delta}(x, \hat{x}, \mu, \eta)$$

where

$$\begin{aligned}\mathcal{L}_{\theta,\delta}(x,\hat{x},\mu,\eta) := (1-\theta)f(x,\hat{x}) - \sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N}\mu_{(j)}^{l}h_{(j)}^{l}(c^{l}) - \\ \sum_{l=1}^{\mathfrak{n}}\sum_{i,k=1}^{N,M}\eta^{l}\left(c_{(i,k)}^{l}-1\right) - \tfrac{\delta}{2}(\|x\|^{2}+\|\hat{x}\|^{2}-\|\mu\|^{2}-\|\eta\|^{2})\end{aligned} \tag{6.11}$$

Notice also that the Lagrange function (6.11) satisfies the saddle-point condition, namely, for all $x \in X$, $\hat{x} \in \hat{X}$, and $\mu, \eta \geq 0$ we have

$$\mathcal{L}_{\theta,\delta}(x_\delta,\hat{x}_\delta,\mu_\delta^*,\eta_\delta^*) \leq \mathcal{L}_{\theta,\delta}(x_\delta^*,\hat{x}_\delta^*,\mu_\delta^*,\eta_\delta^*) \leq \mathcal{L}_{\theta,\delta}(x_\delta^*,\hat{x}_\delta^*,\mu_\delta,\eta_\delta)$$

6.3.1 The proximal format

In the proximal format the relation (6.11) can be expressed as

$$\begin{aligned}\mu_\delta^* &= \arg\min_{\mu\geq 0}\left\{\tfrac{1}{2}\|\mu-\mu_\delta^*\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x_\delta^*,\hat{x}_\delta^*,\mu,\eta_\delta^*)\right\}\\ \eta_\delta^* &= \arg\min_{\eta\geq 0}\left\{\tfrac{1}{2}\|\eta-\eta_\delta^*\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x_\delta^*,\hat{x}_\delta^*,\mu_\delta^*,\eta)\right\}\\ x_\delta^* &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_\delta^*\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x,\hat{x}_\delta^*,\mu_\delta^*,\eta_\delta^*)\right\}\\ \hat{x}_\delta^* &= \arg\max_{\hat{x}\in\hat{X}}\left\{-\tfrac{1}{2}\|\hat{x}-\hat{x}_\delta^*\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x_\delta^*,\hat{x},\mu_\delta^*,\eta_\delta^*)\right\}\end{aligned}$$

where the solutions x_δ^*, $\hat{x}_\delta^*(u)$, μ_δ^* and η_δ^* depend on the parameters $\delta > 0$ and $\gamma > 0$.

6.3.2 The Extraproximal method

We design the method for the static Nash game in a general format with some fixed admissible initial values ($x_0 \in X$, $\hat{x}_0 \in \hat{X}$, and $\mu_0, \eta_0 \geq 0$), considering that we want to maximize the function, as follows:

1. The *first half-step*:

$$\begin{aligned}\bar{\mu}_n &= \arg\max_{\mu\geq 0}\left\{-\tfrac{1}{2}\|\mu-\mu_n\|^2 - \gamma\mathcal{L}_{\theta,\delta}(x_n,\hat{x}_n,\mu,\eta_n)\right\}\\ \bar{\eta}_n &= \arg\max_{\eta\geq 0}\left\{-\tfrac{1}{2}\|\eta-\eta_n\|^2 - \gamma\mathcal{L}_{\theta,\delta}(x_n,\hat{x}_n,\mu_n,\eta)\right\}\\ \bar{x}_n &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_n\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x,\hat{x}_n,\bar{\mu}_n,\bar{\eta}_n)\right\}\\ \bar{\hat{x}}_n &= \arg\max_{\hat{x}\in\hat{X}}\left\{-\tfrac{1}{2}\|\hat{x}-\hat{x}_n\|^2 + \gamma\mathcal{L}_{\theta,\delta}(x_n,\hat{x},\bar{\mu}_n,\bar{\eta}_n)\right\}\end{aligned} \tag{6.12}$$

2. The *second half-step*

$$\begin{aligned}
\mu_{n+1} &= \arg\max_{\mu\geq 0}\left\{-\tfrac{1}{2}\|\mu-\mu_n\|^2-\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\overline{\hat{x}}_n,\mu,\bar{\eta}_n)\right\}\\
\eta_{n+1} &= \arg\max_{\eta\geq 0}\left\{-\tfrac{1}{2}\|\eta-\eta_n\|^2-\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\overline{\hat{x}}_n,\bar{\mu}_n,\eta)\right\}\\
x_{n+1} &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_n\|^2+\gamma\mathcal{L}_{\theta,\delta}(x,\overline{\hat{x}}_n,\bar{\mu}_n,\bar{\eta}_n)\right\}\\
\hat{x}_{n+1} &= \arg\max_{\hat{x}\in \hat{X}}\left\{-\tfrac{1}{2}\|\hat{x}-\hat{x}_n\|^2+\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\hat{x},\bar{\mu}_n,\bar{\eta}_n)\right\}
\end{aligned} \tag{6.13}$$

6.4 Numerical Examples

Example 1

Our goal is to analyze a 2-player Nash Bargaining situation in a class of ergodic controllable finite Markov chains. Denote the disagreement utility that depends on the strategies $c^l_{(i,k)}$ $(l = 1, 2)$ for players 1 and 2 as $\phi^1(c^1, c^2)$ and $\phi^2(c^1, c^2)$ respectively, and the solution for the Nash bargaining problem as the point (ψ_1, ψ_2). Let the states $N = 3$ and the number of actions $M = 2$ for each player. The individual utility for each player are defined by

$$J^1_{(i,j,1)} = \begin{bmatrix} 7 & 17 & 13 \\ 0 & 1 & 18 \\ 13 & 7 & 10 \end{bmatrix} \quad J^1_{(i,j,2)} = \begin{bmatrix} 18 & 3 & 10 \\ 9 & 0 & 7 \\ 15 & 6 & 16 \end{bmatrix}$$

$$J^2_{(i,j,1)} = \begin{bmatrix} 9 & 11 & 6 \\ 9 & 17 & 3 \\ 11 & 1 & 4 \end{bmatrix} \quad J^2_{(i,j,2)} = \begin{bmatrix} 10 & 18 & 0 \\ 12 & 7 & 18 \\ 17 & 6 & 10 \end{bmatrix}$$

The transition matrices for each player are as follows

$$\pi^1_{(i,j,1)} = \begin{bmatrix} 0.5144 & 0.2877 & 0.1978 \\ 0.3775 & 0.0893 & 0.5332 \\ 0.3305 & 0.2703 & 0.3992 \end{bmatrix} \quad \pi^1_{(i,j,2)} = \begin{bmatrix} 0.3438 & 0.3846 & 0.2717 \\ 0.2484 & 0.0756 & 0.6759 \\ 0.1378 & 0.4655 & 0.3968 \end{bmatrix}$$

$$\pi^2_{(i,j,1)} = \begin{bmatrix} 0.3541 & 0.1945 & 0.4514 \\ 0.5929 & 0.2559 & 0.1512 \\ 0.4288 & 0.2434 & 0.3278 \end{bmatrix} \quad \pi^2_{(i,j,2)} = \begin{bmatrix} 0.6435 & 0.0216 & 0.3349 \\ 0.2990 & 0.3905 & 0.3105 \\ 0.5575 & 0.2203 & 0.2221 \end{bmatrix}$$

Computing the disagreement point. Given δ, γ and applying the extraproximal method we obtain the convergence of the strategies for each player in the disagreement point in terms of the variable $c^l_{(i,k)}$ (Figure 6.2 and Figure 6.3).

$$c^1 = \begin{bmatrix} 0.1683 & 0.1551 \\ 0.1829 & 0.0973 \\ 0.1853 & 0.2111 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.2618 & 0.2122 \\ 0.0673 & 0.1320 \\ 0.1305 & 0.1962 \end{bmatrix}$$

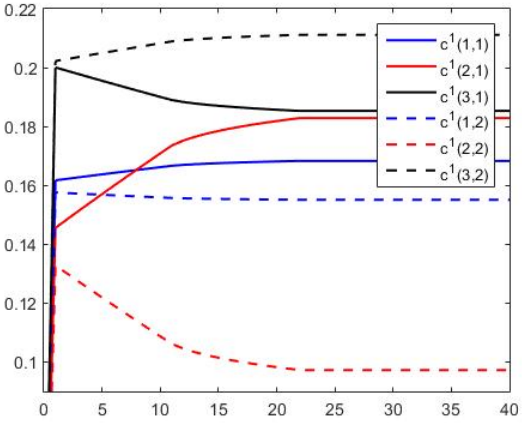


Figure 6.2 Strategies for player 1

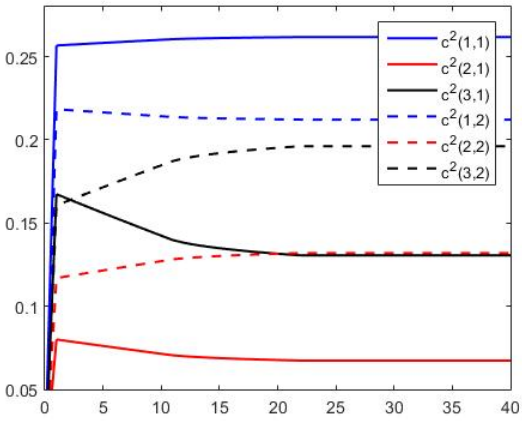


Figure 6.3 Strategies for player 2.

With the strategies calculated, the resulting utilities in the disagreement point for each player are as follows:

$$\phi^1(c^1, c^2) = 120.3001 \qquad \phi^2(c^1, c^2) = 97.0832$$

Computing the Nash Bargaining solution. The Nash's solution has a simple geometric interpretation in a two-person game: given a bargaining pair, for every point (ψ_1, ψ_2), consider the product (area of a rectangle) $(\psi_1 - \phi^1)(\psi_2 - \phi^2)$. Then (ψ_1, ψ_2) is the unique point in the Pareto front that maximizes this product [60].

Following the method presented and applying the extraproximal method for the Nash bargaining problem (6.6 - 6.7), we obtain the convergence of the strategies for the bargaining solution in terms of the variable $c^l_{(i,k)}$ for each player (see Figure 6.4 and Figure 6.5).

$$c^1 = \begin{bmatrix} 0.1890 & 0.1178 \\ 0.3057 & 0.0010 \\ 0.0010 & 0.3854 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.3463 & 0.0881 \\ 0.0010 & 0.2325 \\ 0.0010 & 0.3310 \end{bmatrix}$$

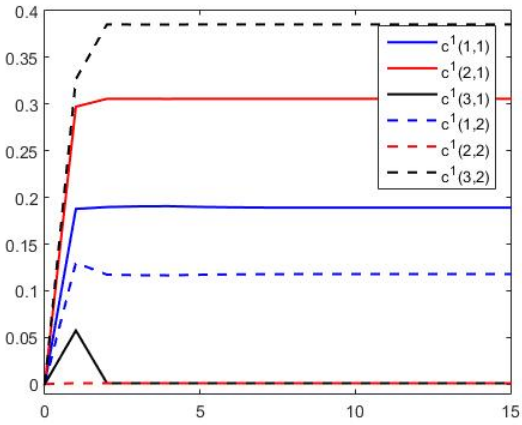


Figure 6.4 Strategies of player 1

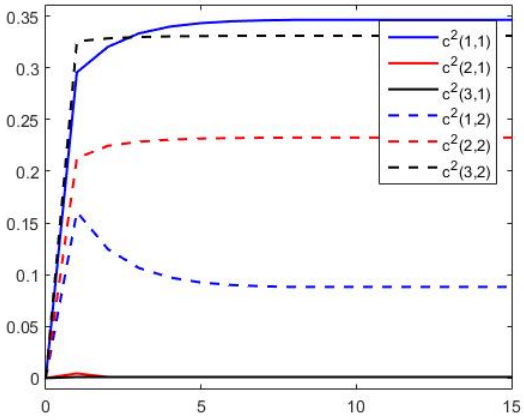


Figure 6.5 Strategies of player 2.

With the strategies calculated, the resulting utilities in the bargaining solution are as follows:

$$\psi^1(c^1, c^2) = 139.6854 \qquad \psi^2(c^1, c^2) = 119.4296$$

We can see that the profits obtained at the point of Nash bargaining solution are greater than those obtained at the disagreement point.

Example 2

Our goal is to analyze a three-player Nash bargaining situation in a class of ergodic controllable finite Markov chains, we have $\mathfrak{n} = 3$. Denote the disagreement cost that depends on the strategies $c^l_{(i,k)}$ $(l = 1, 2, 3)$ for players 1, 2 and 3 as $\phi^1(c^1, c^2, c^3)$, $\phi^2(c^1, c^2, c^3)$ and $\phi^3(c^1, c^2, c^3)$ respectively, and the solution for the Nash bargaining problem as (ψ^1, ψ^2, ψ^3).

Let the states $N = 5$ and the number of actions $M = 2$ for each player. The individual utility for each player are defined by

$$U^1_{(i,j,1)}=\begin{bmatrix}10 & 9 & 3 & 7 & 6\\ 11 & 19 & 6 & 8 & 10\\ 9 & 7 & 13 & 19 & 5\\ 14 & 9 & 15 & 2 & 16\\ 12 & 4 & 9 & 3 & 10\end{bmatrix} \quad U^2_{(i,j,1)}=\begin{bmatrix}14 & 9 & 12 & 6 & 1\\ 18 & 12 & 7 & 9 & 10\\ 5 & 14 & 8 & 11 & 6\\ 19 & 13 & 8 & 4 & 10\\ 6 & 9 & 12 & 10 & 8\end{bmatrix} \quad U^3_{(i,j,1)}=\begin{bmatrix}5 & 8 & 7 & 8 & 10\\ 10 & 4 & 13 & 11 & 5\\ 14 & 11 & 6 & 17 & 1\\ 12 & 0 & 9 & 7 & 3\\ 16 & 12 & 8 & 4 & 10\end{bmatrix}$$

$$U^1_{(i,j,2)}=\begin{bmatrix}12 & 6 & 7 & 10 & 5\\ 16 & 0 & 9 & 14 & 5\\ 18 & 10 & 16 & 9 & 4\\ 2 & 16 & 9 & 7 & 13\\ 11 & 9 & 3 & 17 & 10\end{bmatrix} \quad U^2_{(i,j,2)}=\begin{bmatrix}10 & 17 & 6 & 9 & 11\\ 16 & 9 & 4 & 12 & 8\\ 10 & 13 & 9 & 1 & 18\\ 12 & 18 & 15 & 9 & 4\\ 17 & 3 & 9 & 10 & 6\end{bmatrix} \quad U^3_{(i,j,2)}=\begin{bmatrix}9 & 13 & 7 & 10 & 11\\ 19 & 5 & 0 & 7 & 20\\ 11 & 2 & 19 & 6 & 9\\ 3 & 10 & 14 & 5 & 18\\ 9 & 10 & 17 & 6 & 11\end{bmatrix}$$

The transition matrices for each player are defined as follows

$$\pi^1_{(i,j,1)}=\begin{bmatrix}0.233 & 0.130 & 0.090 & 0.278 & 0.266\\ 0.123 & 0.029 & 0.174 & 0.394 & 0.279\\ 0.229 & 0.187 & 0.276 & 0.182 & 0.124\\ 0.129 & 0.265 & 0.336 & 0.190 & 0.078\\ 0.331 & 0.189 & 0.108 & 0.172 & 0.198\end{bmatrix} \quad \pi^1_{(i,j,2)}=\begin{bmatrix}0.210 & 0.235 & 0.166 & 0.299 & 0.088\\ 0.177 & 0.054 & 0.483 & 0.143 & 0.141\\ 0.102 & 0.345 & 0.294 & 0.167 & 0.089\\ 0.187 & 0.100 & 0.364 & 0.249 & 0.099\\ 0.087 & 0.197 & 0.193 & 0.388 & 0.132\end{bmatrix}$$

$$\pi^2_{(i,j,1)}=\begin{bmatrix}0.240 & 0.132 & 0.307 & 0.160 & 0.159\\ 0.353 & 0.152 & 0.090 & 0.170 & 0.232\\ 0.243 & 0.138 & 0.185 & 0.249 & 0.183\\ 0.134 & 0.214 & 0.244 & 0.290 & 0.116\\ 0.170 & 0.267 & 0.215 & 0.167 & 0.179\end{bmatrix} \quad \pi^2_{(i,j,2)}=\begin{bmatrix}0.389 & 0.013 & 0.202 & 0.117 & 0.278\\ 0.171 & 0.224 & 0.178 & 0.323 & 0.101\\ 0.315 & 0.124 & 0.125 & 0.217 & 0.217\\ 0.185 & 0.122 & 0.330 & 0.171 & 0.189\\ 0.111 & 0.285 & 0.208 & 0.205 & 0.190\end{bmatrix}$$

$$\pi^3_{(i,j,1)}=\begin{bmatrix}0.070 & 0.334 & 0.261 & 0.143 & 0.189\\ 0.053 & 0.085 & 0.446 & 0.126 & 0.288\\ 0.127 & 0.325 & 0.140 & 0.180 & 0.227\\ 0.317 & 0.265 & 0.031 & 0.227 & 0.158\\ 0.101 & 0.291 & 0.039 & 0.311 & 0.256\end{bmatrix} \quad \pi^3_{(i,j,2)}=\begin{bmatrix}0.466 & 0.108 & 0.084 & 0.124 & 0.215\\ 0.205 & 0.241 & 0.107 & 0.143 & 0.301\\ 0.044 & 0.216 & 0.305 & 0.313 & 0.120\\ 0.287 & 0.171 & 0.205 & 0.098 & 0.235\\ 0.145 & 0.214 & 0.166 & 0.245 & 0.227\end{bmatrix}$$

Computing the disagreement point. Given δ and γ and applying the extraproximal method (eqs. 6.12 and 6.13) we obtain the convergence of the strategies for the disagreement point in terms of the variable $c^l_{(i,k)}$ for each player (see Figure 6.6).

$$c^1 = \begin{bmatrix} 0.1729 & 0.0399 \\ 0.1445 & 0.0010 \\ 0.1549 & 0.0641 \\ 0.0010 & 0.2475 \\ 0.1732 & 0.0010 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.0499 & 0.1642 \\ 0.0010 & 0.1682 \\ 0.2139 & 0.0010 \\ 0.1856 & 0.0440 \\ 0.0734 & 0.0989 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.1004 & 0.0533 \\ 0.2270 & 0.0010 \\ 0.1296 & 0.0645 \\ 0.2027 & 0.0010 \\ 0.1505 & 0.0699 \end{bmatrix}$$

Following eq. (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.8125 & 0.1875 \\ 0.9931 & 0.0069 \\ 0.7073 & 0.2927 \\ 0.0040 & 0.9960 \\ 0.9943 & 0.0057 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.2330 & 0.7670 \\ 0.0059 & 0.9941 \\ 0.9953 & 0.0047 \\ 0.8085 & 0.1915 \\ 0.4261 & 0.5739 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.6530 & 0.3470 \\ 0.9956 & 0.0044 \\ 0.6676 & 0.3324 \\ 0.9951 & 0.0049 \\ 0.6827 & 0.3173 \end{bmatrix}$$

With the strategies calculated, the resulting utilities, following eq. (6.2), in the disagreement point for each player $\phi^l(c^1, c^2, c^3)$ are as follows:

$$\phi^1(c^1, c^2, c^3) = 93.1288 \qquad \phi^2(c^1, c^2, c^3) = 100.9968 \qquad \phi^3(c^1, c^2, c^3) = 96.0779$$

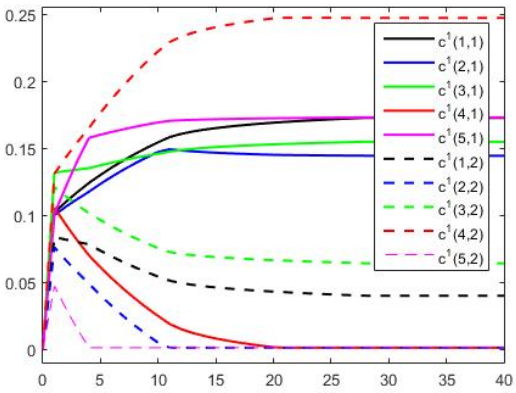
Computing the Nash bargaining solution. The Nash's unique solution has a very simple geometric interpretation in a three-person game: given a bargaining pair, for every point (ψ^1, ψ^2, ψ^3), consider the product (volume of a rectangular prism) $(\psi^1 - \phi^1)(\psi^2 - \phi^2)(\psi^3 - \phi^3)$. Then (ψ^1, ψ^2, ψ^3) is the unique point in the Pareto front that maximizes this product (see [60]). The function for finding the solution to the Nash Bargaining problem in a three-person game is

$$g(c^1, c^2, c^3) = (\psi^1 - \phi^1)^{\alpha^1 \chi(\psi^1 > \phi^1)} \cdot (\psi^2 - \phi^2)^{\alpha^2 \chi(\psi^2 > \phi^2)} \cdot (\psi^3 - \phi^3)^{\alpha^3 \chi(\psi^3 > \phi^3)} \qquad (6.14)$$

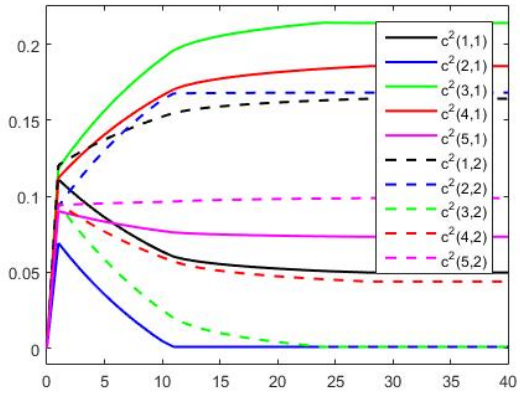
where the parameters $\alpha_1, \alpha_2, \alpha_3 \geq 0$ and $\alpha^1 + \alpha^2 + \alpha^3 = 1$. For the implementation we can rewrite (6.14) as follows

$$\tilde{g}(c^1, c^2, c^3) = \alpha^1 \chi(\psi^1 > \phi^1) \ln(\psi^1 - \phi^1)$$

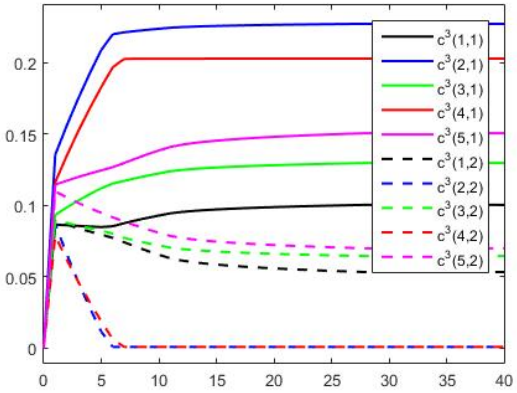
$$+\alpha^2 \chi(\psi^2 > \phi^2) \ln(\psi^2 - \phi^2) + \alpha^3 \chi(\psi^3 > \phi^3) \ln(\psi^3 - \phi^3)$$



(a) Player 1.



(b) Player 2.

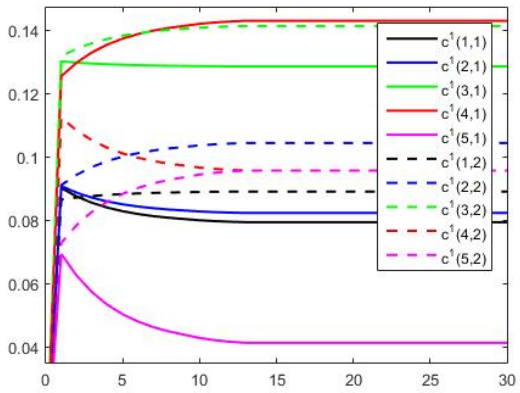


(c) Player 3.

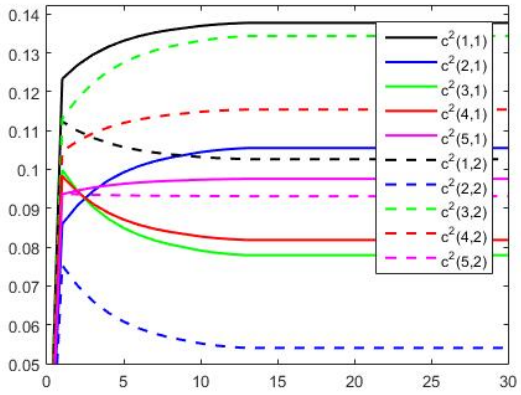
Figure 6.6 Convergence of players' strategies.

Then, given δ, γ, $\alpha^l = 1/3$ and applying the extraproximal method (eqs. 6.6 and 6.7) for the Nash bargaining problem we obtain the convergence of the strategies for the bargaining solution in terms of the variable $c^l_{(i,k)}$ for each player (see Figure 6.7).

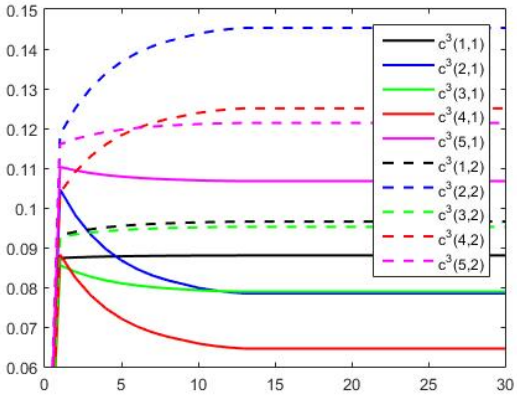
$$c^1 = \begin{bmatrix} 0.0793 & 0.0890 \\ 0.0822 & 0.1043 \\ 0.1285 & 0.1413 \\ 0.1429 & 0.0956 \\ 0.0412 & 0.0956 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.1377 & 0.1026 \\ 0.1055 & 0.0540 \\ 0.0779 & 0.1343 \\ 0.0818 & 0.1154 \\ 0.0976 & 0.0931 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.0880 & 0.0965 \\ 0.0784 & 0.1453 \\ 0.0789 & 0.0952 \\ 0.0645 & 0.1250 \\ 0.1067 & 0.1214 \end{bmatrix}$$



(a) Player 1.



(b) Player 2.



(c) Player 3.

Figure 6.7 Convergence of players' strategies.

Following eq. (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.4712 & 0.5288 \\ 0.4408 & 0.5592 \\ 0.4764 & 0.5236 \\ 0.5991 & 0.4009 \\ 0.3011 & 0.6989 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.5730 & 0.4270 \\ 0.6613 & 0.3387 \\ 0.3671 & 0.6329 \\ 0.4149 & 0.5851 \\ 0.5117 & 0.4883 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.4769 & 0.5231 \\ 0.3506 & 0.6494 \\ 0.4530 & 0.5470 \\ 0.3404 & 0.6596 \\ 0.4679 & 0.5321 \end{bmatrix}$$

Then, we can see that the resulting utilities obtained in the Nash bargaining solution are greater than those obtained at the disagreement point.

$$\psi^1(c^1, c^2.c^3) = 118.0408 \qquad \psi^2(c^1, c^2, c^3) = 117.3255 \qquad \psi^3(c^1, c^2, c^3) = 122.5102$$

Chapter 7

Solving Bargaining by Manipulation

7.1 Introduction

The Machiavellianism is defined as a political strategy of social conduct that involves manipulating others for personal gain, often against the other's self-interest [16]. This concept coincides with the central insight that gave rise to modern economics where the common good is well served by the free actions of self-interested agents in a market. The profit maximization of agents assigns essentially no role to generosity and social conscience because actions in many domains of application commonly conform to standards of manipulation. In this sense, the manipulation model presents an advantage for expanding the classical economic models as a more realistic behavioral assumption. In the classical economic theory, agents are assumed to be rationally law-abiding but not fair. This non-fairness assumption can be explained by the Machiavellianism in terms of the immorality (considered by Christie and Geis [16] to be among the three key elements of Machiavellianism) which has deep roots in the history of the economy. It is important to note that a manipulation strategy is a political strategy and it is not an economical strategy as is presented in compensation contracts which solves a particular case of manipulation [12].

The concept of Machiavellianism was first studied by Christie and Geis [16] as the ability to manipulate others as an important personality trait. They analyze whether the principles associated with three of Machiavelli's greatest works (The Prince [54], The Discourses [54]

and The Art of War [55]) were practiced by individuals in today's society. The fundamental idea throughout Machiavelli's discourses is the degree to which people can be manipulated.

Christie and Geis [16] defined the Machiavellian personality type as someone who seeks to manipulate others to achieve his or her own ends. Machiavellianism structure is composed by three key elements: 1) the belief in manipulative tactics, 2) a cynical world view, and 3) a pragmatic morality (immorality). For individuals who manipulate, others are viewed entirely as objects or as means to personal ends (views) having an utilitarian, rather than a moral view of their interactions with others (immorality) and focused on applying manipulation strategies for accomplishing their goals related to power situations (tactics).

The interest in the subject of Machiavellianism was almost among social psychologists [107, 23, 8, 85]. These works only made an interpretation of the concept of Machiavellianism related to well-known games of game theory and in game-theoretic experiments focusing on explaining the rationality of the players. These works has not converged on the framework of game theory and a small number of articles are inspired by traditional game theory. In addition, the effects of repeated interactions with the intention to exploit others have not been addressed.

The manipulation game is conceptualized under the Machiavellianism psychological theory which determines a Stackelberg game model consisting of manipulating and manipulated players that employ manipulation strategies to achieve power situations with the disposition to not become attached to a conventional moral. The Stackelberg game focuses on computing the strong Stackelberg equilibrium. In this paper we are considering manipulating and manipulated players engaged in a cooperative Nash game. Power situations suggest that the advantage is for the manipulating players. The equilibrium may be imposed on the manipulated players without their approval but considering that every player is able of manipulative behavior to some degree (manipulated players try to minimize their lost). The resulting manipulation strategy is an outcome which is optimally better for the manipulating players with the manipulated players necessarily worse off. The rationality of the players follows this two basics principles: a) no manipulating player will agree to accept a payoff lower than the one guaranteed to him under disagreement, and b) the agreement will represent a situation that could not be improved by the manipulated players.

The main results of this chapter are summarized as follows.

- The manipulation game is conceptualized under the Machiavellianism psychological theory as a Stackelberg game model involving manipulating (leaders) and manipulated (followers) players.
- We consider a game model involving manipulating and manipulated players engaged cooperatively in a Nash game restricted by a Stackelberg game.
- The cooperation is represented by the Nash bargaining solution.
- It is proposed an analytical method for finding the manipulation equilibrium point. There is a manipulating strategy solution (which arises as the maximum of the quotient of two Nash products) which under a feasibility condition is a manipulation equilibrium point.
- We represent the Stackelberg game model as Nash game for relaxing the interpretation of the game and the equilibrium selection problem
- The solution concept applied to the manipulation game focuses on computing the manipulation equilibrium which is a political strategy.
- Under conditions of unequal relative power among players, the player with high power tends to behave exploitative, while the less powerful player tends to behave submissively.
- The weights of the players for the Nash solution are determined by their role in the Stackelberg game.
- The manipulated players break ties optimally for the manipulating players finding a new strong Stackelberg equilibrium point solution where manipulating maximize the gain and the manipulated minimize the lost. There is an equilibrium selection problem forcing the manipulating players to manipulate on which equilibrium to converge.
- The manipulation equilibrium point is a political strategy with an outcome which is optimally better for the manipulating players.

- The computation of the problem is fitted into a class of homogeneous, ergodic, controllable and finite Markov chains games.

7.2 The Manipulation Game

7.2.1 Machiavellian structure

The *Machiavellianism structure* that encodes the set of characteristics of a Machiavellian individual is represented by three fundamental concepts [16]:

- Views: The belief that the world can be manipulated - the world consists of manipulating and manipulated
- Tactics: The use of a manipulation strategies needed to achieve specific power situations (goals).
- Immorality: The disposition to not become attached to a conventional moral.

Remark 7.1 *Strategies are based on the Machiavelli's The Prince [54], The Discourses [54], The Art of War [55] and the psychological behavior patterns [107].*

A *manipulation game* is a Stackelberg game model consisting of manipulating and manipulated players (views) that employ manipulation strategies to achieve power situations (tactics) with the disposition to not become attached to a conventional moral (immorality) [18].

The solution concept applied to the manipulation game is the strong Stackelberg equilibrium. In the manipulation game the manipulating players consider the best-reply of the manipulated players selecting the strategy that maximizes the payoff anticipating the predicted best-reply of the manipulated players. The manipulated players break ties optimally for the manipulating players and in equilibrium select the expected strategy as a best-reply. We are considering manipulating and manipulated players engaged in a cooperative Nash game restricted by a Stackelberg game.

The formal definition and rationality of the solution for the bargaining problem based on the manipulation game is as follows.

7.2.2 The bargaining manipulation solution

Consider a manipulation game, for a finite set of players $\mathcal{I} = \{\mathcal{N} \cup \mathcal{M}\}$ with $\mathfrak{n} + \mathfrak{m}$ elements, let $\mathbb{R}^{\mathcal{I}}$ denote the $(\mathfrak{n} + \mathfrak{m})$-dimensional Euclidean space with coordinates indexed by the elements of $\mathcal{I}$. For representing a Stackelberg game as a Nash game a strategy profile $x = (u, v) \in X \subseteq R^{\mathfrak{n}+\mathfrak{m}}$ is constructed, where $X = U \otimes V$ is the concatenation of U and V, such that $x = (u^1, \ldots, u^{\mathfrak{n}}, v^1, \ldots, v^{\mathfrak{m}})$. The strategy profile u represents the proposal of the manipulating players and the strategy profile v represents the proposal of the manipulated players. The manipulation solution is based on a model in which the players are assumed to manipulate on which point of the feasible payoff vector $\Phi(x) = (\varphi^1(u), ..., \varphi^{\mathfrak{n}}(u), \psi^1(v), ..., \psi^{\mathfrak{m}}(v))$ where $\varphi(u) = (\varphi^1(u), ..., \varphi^{\mathfrak{n}}(u))$ and $\psi(v) = (\psi^1(v), ..., \psi^{\mathfrak{m}}(v))$ are the payoff vectors corresponding to the manipulating and manipulated players, respectively, and the vector $x = (u^1, \ldots, u^{\mathfrak{n}}, v^1, \ldots, v^{\mathfrak{m}}) \in X$. Let us denote $\Psi := \{\Phi(x) \in \mathbb{R}^{\mathfrak{n}+\mathfrak{m}} \mid x \in X\}$ as the adjunct set of payoff vectors $\Phi(x)$.

Players have strictly opposed preferences and each one is concerned only with the share of benefits it obtains from manipulation. A fundamental point of the model is a fixed disagreement vector $\phi = (\tilde{\varphi}(u), \tilde{\psi}(v)) \in \Psi$ which plays the role of a deterrent where $\tilde{\varphi}(u) = (\tilde{\varphi}^1(u), ..., \tilde{\varphi}^{\mathfrak{n}}(u))$ is the disagreement payoff vector corresponding to the manipulating players and $\tilde{\psi}(v) = (\tilde{\psi}^1(v), ..., \tilde{\psi}^{\mathfrak{m}}(v))$ is the disagreement payoff vector corresponding to the manipulated players.

The manipulating players would like to increase their components in ϕ and to achieve a $\varrho \in \Psi(x)$ for which $\varrho \geq \phi$ (where $\varrho_\iota \geq \phi_\iota$, $\iota = 1, ..., \mathfrak{n} + \mathfrak{m}$). We will suppose the conflict of interest involves all external factors, depends only on the agreement being considered and on the objections, and therefore the manipulation process is independent of time, history, and experience.

Remark 7.2 *We are considering manipulating and manipulated players engaged in a cooperative Nash game restricted by a Stackelberg game.*

Remark 7.3 *In a Stackelberg game leaders and followers move asynchronous. For instance, if we first fix the followers then, we will have a Nash product for the leaders given by* $\prod_{l=1}^{\mathfrak{n}}(\varphi^l(u|v)-$

$\tilde{\varphi}^l$). *On the other hand, if we fix the leaders we will have a Nash product for the followers given by* $\prod_{m=1}^{\mathfrak{m}}(\tilde{\psi}^m - \psi^m(v|u))$. *In a Stackelberg game we look for*

$$\max_{u\in U} \prod_{l=1}^{\mathfrak{n}} (\varphi^l(u|v) - \tilde{\varphi}^l),\ \varphi^l > \tilde{\varphi}^l$$

while

$$\min_{v\in V} \prod_{m=1}^{\mathfrak{m}} (\tilde{\psi}^m - \psi^m(v|u)),\ \tilde{\psi}^m > \psi^m$$

Then, for fitting the manipulation game to more real situations we consider that manipulating and manipulated players can move simultaneously. In addition, we consider that every player is capable of manipulative behavior to some degree, but making emphasis in the fact that some are more willing and more able than others. We represent the Stackelberg game model as Nash game for relaxing the interpretation of the game and the equilibrium selection problem.

Remark 7.4 *The transformation of a Stackelberg game into a Nash game is already described by [90] where the authors suggest that a leadership game is a two-stage game played as follows: the leaders choose and commit to their strategies which are announced to the followers, who then simultaneously choose their strategies, which are played together with the strategies of the leaders. The players' payoffs for the strategy profile are as in the original game. Then, there is no need for understanding leadership as an asynchronous game.*

Following the Remark 7.2, Remark 7.3 and Remark 7.4 we approach the solution of the bargaining by manipulation problem as the maximum of the quotient of two Nash products as follows.

Definition 7.5 *A strategy* $x^* = (u^*, v^*) \in X$ *is called a manipulation strategy solution of the game if it is an optimal solution of the maximization problem*

$$\max_{x\in X} \qquad \zeta(\varrho(x)) = \frac{\prod_{l=1}^{\mathfrak{n}} (\varphi^l(u) - \tilde{\varphi}^l)}{\prod_{m=1}^{\mathfrak{m}} (\tilde{\psi}^m - \psi^m(v))} \tag{7.1}$$

$$\textit{subject to} \qquad \varrho(x) \in \Psi(x)$$
$$\varphi^l > \tilde{\varphi}^l \textit{ and } \tilde{\psi}^m > \psi^m$$

where $\phi(x\prime) = (\tilde{\varphi}(u), \tilde{\psi}(v)) \in \Psi(x\prime)$, $x\prime \in X$, $\varrho(x^*) = (\varphi(u^*), \psi(v^*))$ *such that* $\varphi(u^*) = \left(\varphi^l(u^*)\right)_{l=\overline{1,\mathfrak{n}}}$, *and* $\psi(v^*) = (\psi^m(v^*))_{m=\overline{1,\mathfrak{m}}}$.

Remark 7.6 *The pay-off vector given by* $\varrho(x^*) = (\varphi(u^*), \psi(v^*))$ *generated by manipulation solution* $x^* = (u^*, v^*) \in X$ *is called the manipulation solution payoff.*

For making emphasis in representing the Stackelberg game model as a Nash game we consider that the problem (7.1) for finding the solution to the manipulation problem is given by

$$\zeta(\varrho(x)) = \frac{\prod\limits_{l=1}^{\mathfrak{n}} (\varphi^l - \tilde{\varphi}^l)^{\alpha^l \chi(\varphi^l > \tilde{\varphi}^l)}}{\prod\limits_{m=1}^{\mathfrak{m}} (\tilde{\psi}^m - \psi^m)^{\beta^m \chi(\tilde{\psi}^m > \psi^m)}} \quad \rightarrow \max_{x \in X} \tag{7.2}$$

where $\varrho(x) = (\varphi(u), \psi(v))$ and $\alpha^l \geq \beta^m > 0$ $(l = 1, ..., \mathfrak{n},\ m = 1, ..., \mathfrak{m})$ are the weighting parameters for manipulating and manipulated players, respectively. Then, we rewrite (7.2) as follows

$$\begin{aligned} \zeta(\varrho(x)) = \sum_{l=1}^{\mathfrak{n}} \alpha^l \chi(\varphi^l > \tilde{\varphi}^l) \ln(\varphi^l - \tilde{\varphi}^l) - \\ \sum_{m=1}^{\mathfrak{m}} \beta^m \chi(\tilde{\psi}^m > \psi^m) \ln(\tilde{\psi}^m - \psi^m) \quad \rightarrow \max_{x \in X} \end{aligned} \tag{7.3}$$

where $\varrho(x) = (\varphi(u), \psi(v))$ where $\varphi(u^*) = \varphi_l(u^*)_{l=\overline{1,\mathfrak{n}}}$ and $\psi(v^*) = \psi_m(v^*)_{m=\overline{1,\mathfrak{m}}}$.

Given $\zeta(\varrho(x))$ and considering the disagreement vector $\phi(x\prime) = (\tilde{\varphi}(u), \tilde{\psi}(v)) \in \Psi(x\prime)$, a payoff vector solution to the manipulation problem is a function $\varrho(x) \in \Psi(x)$ such that $x \in X$. The manipulation process result in a particular strategy solution $x^* \in X$ which can be considered the equilibrium point of the manipulation game when it results a particular point satisfying $\varrho(x^*) \in \Psi(x^*)$.

Definition 7.7 *The strategy solution* $x^* = (u^*, v^*) \in X$ *of the manipulation game is called the manipulation equilibrium point.*

In the following statement we present the characterization of the manipulation equilibrium point $x^* = (u^*, v^*) \in X$ of the manipulation game.

Theorem 7.8 *Let* Γ *be a manipulation game. Then, the manipulation strategy solution* $x^* = (u^*, v^*) \in X$ *of the game* Γ *is a manipulation equilibrium point if and only if* $\varrho(x^*) \in \Psi(x^*)$.

Proof. $\Rightarrow$) Suppose that $x^* = (u^*, v^*) \in X$ is an equilibrium point. In addition, let us suppose that there exists a $x \in X$, $x \neq x^*$, such that $\varrho(x) = (\varphi(u), \psi(v)) \in \Psi(x)$ such that $\varphi(u) > \varphi(u^*)$ and $\psi(v) > \psi(v^*)$. It is impossible, because $x^* \in X$ is a solution of the manipulation game Γ.

$\Leftarrow$) By contradiction. Suppose that $\varrho(x^*) \notin \Psi(x^*)$. Then, it is possible for the manipulating player to increase their pay-off. Consistently, it is possible for the manipulated players to reduce their pay-off. Then, it is not a manipulation equilibrium point. ■

Remark 7.9 *The bargaining conditions under manipulation will produce that manipulating players prefers to increase the profit while the manipulated players prefers to decrease it. Under these circumstances, it may be necessary for all players to adjust the profit in order to find a Pareto solution. The change of the profit is also a Pareto solution because the renegotiated profit for the manipulated players is below the efficient profit.*

7.3 Numerical example

For this example, consider the results presented in the previous chapters. Let us analyze a two-player manipulation problem, where player 1 is the manipulating and player 2 is the manipulated, in a class of ergodic controllable finite Markov chains. Let the states $N = 3$, and the number of actions $M = 2$. The individual utility for each player are defined by

$$U^1_{(i,j|1)} = \begin{bmatrix} 7 & 17 & 3 \\ 10 & 6 & 7 \\ 16 & 17 & 4 \end{bmatrix} \qquad U^1_{(i,j|2)} = \begin{bmatrix} 6 & 18 & 13 \\ 10 & 18 & 6 \\ 16 & 8 & 10 \end{bmatrix}$$

$$U^2_{(i,j|1)} = \begin{bmatrix} 19 & 11 & 7 \\ 2 & 7 & 13 \\ 1 & 10 & 7 \end{bmatrix} \qquad U^2_{(i,j|2)} = \begin{bmatrix} 1 & 8 & 10 \\ 5 & 17 & 8 \\ 4 & 16 & 1 \end{bmatrix}$$

The transition matrices for each player are defined as follows

$$\pi^1_{(i,j|1)} = \begin{bmatrix} 0.4554 & 0.2548 & 0.2898 \\ 0.2195 & 0.4718 & 0.3086 \\ 0.2460 & 0.3044 & 0.4496 \end{bmatrix} \qquad \pi^1_{(i,j|2)} = \begin{bmatrix} 0.3088 & 0.3445 & 0.3467 \\ 0.0888 & 0.2358 & 0.6754 \\ 0.2336 & 0.4656 & 0.3008 \end{bmatrix}$$

$$\pi^2_{(i,j|1)}=\begin{bmatrix} 0.2906 & 0.3389 & 0.3705 \\ 0.4773 & 0.2058 & 0.3168 \\ 0.4783 & 0.1561 & 0.3656 \end{bmatrix} \qquad \pi^2_{(i,j|2)}=\begin{bmatrix} 0.5628 & 0.1440 & 0.2932 \\ 0.3416 & 0.4461 & 0.2123 \\ 0.4114 & 0.1624 & 0.4262 \end{bmatrix}$$

Given the parameter δ and γ and applying the extraproximal method for finding the Nash equilibrium point of the manipulation situation we obtain the convergence of the strategies for the disagreement point in terms of the variable $c^1_{(i,k)}$ for the manipulating player (see Figure 7.1) and the convergence of the strategies $c^2_{(i,k)}$ for the manipulated player (see Figure 7.2).

$$c^1 = \begin{bmatrix} 0.2403 & 0.0329 \\ 0.2786 & 0.0944 \\ 0.1120 & 0.2419 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.0998 & 0.3811 \\ 0.1789 & 0.0010 \\ 0.1801 & 0.1591 \end{bmatrix}$$

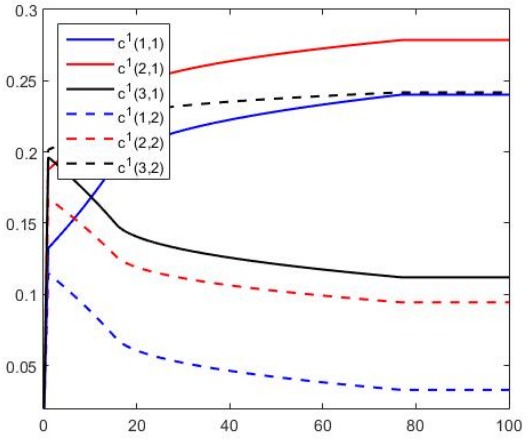


Figure 7.1 Convergence of the strategies for the manipulating player.

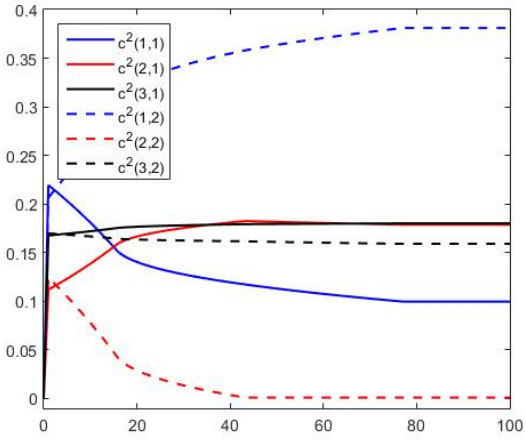


Figure 7.2 Convergence of the strategies for the manipulated player.

Following eq. (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.8795 & 0.1205 \\ 0.7470 & 0.2530 \\ 0.3164 & 0.6836 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.2075 & 0.7925 \\ 0.9944 & 0.0056 \\ 0.5310 & 0.4690 \end{bmatrix}$$

With the strategies calculated, the resulting utilities in the disagreement point for each player are as follows:

$$\tilde{\varphi}_1(c^1, c^2) = 27.1564 \qquad \tilde{\psi}_2(c^1, c^2) = 17.1008$$

Once the disagreement point is fixed, the manipulation process begins. With δ, γ and the weighting parameters $\alpha^1 = 40$ for the manipulating player and $\beta^1 = 25$ for the manipulated player, and applying the extraproximal method we obtain the convergence of the strategies for the manipulation problem in terms of the variable $c^1_{(i,k)}$ for the manipulating player (see Figure 7.3) and the convergence of the strategies $c^2_{(i,k)}$ for the manipulated player (see Figure 7.4).

$$c^1 = \begin{bmatrix} 0.1313 & 0.1157 \\ 0.1873 & 0.1674 \\ 0.1984 & 0.2000 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.0010 & 0.5207 \\ 0.1571 & 0.0010 \\ 0.3192 & 0.0010 \end{bmatrix}$$

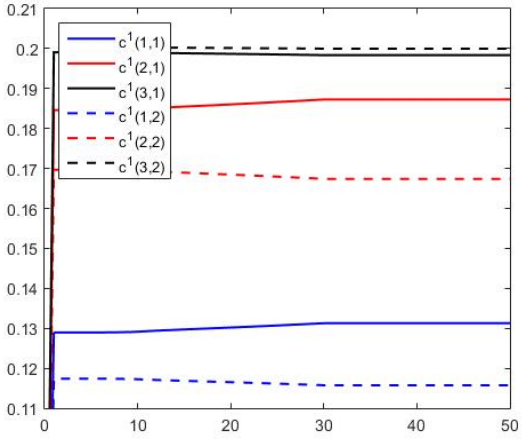


Figure 7.3 Convergence of the strategies for the manipulating player.

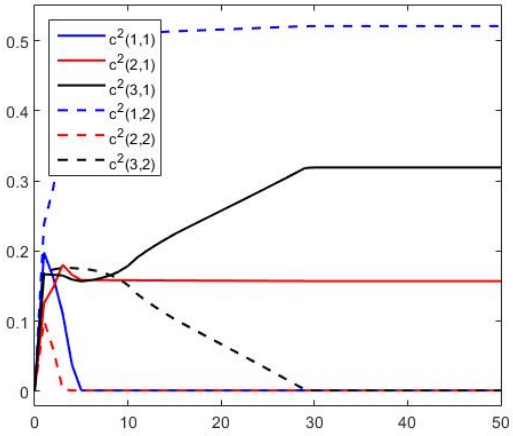


Figure 7.4 Convergence of the strategies for the manipulated player.

Following eq. (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.5314 & 0.4686 \\ 0.5280 & 0.4720 \\ 0.4980 & 0.5020 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.0019 & 0.9981 \\ 0.9937 & 0.0063 \\ 0.9969 & 0.0031 \end{bmatrix}$$

With the strategies calculated, the resulting utilities in the manipulation solution are as follows:

$$\varphi^1(c^1, c^2) = 29.0885 \qquad \psi^1(c^1, c^2) = 14.8154$$

We can see that the profits obtained after the manipulation process are for the manipulating player greater than the disagreement point while for the manipulated player are smaller than the obtained in the disagreement solution, see Figure 7.5.

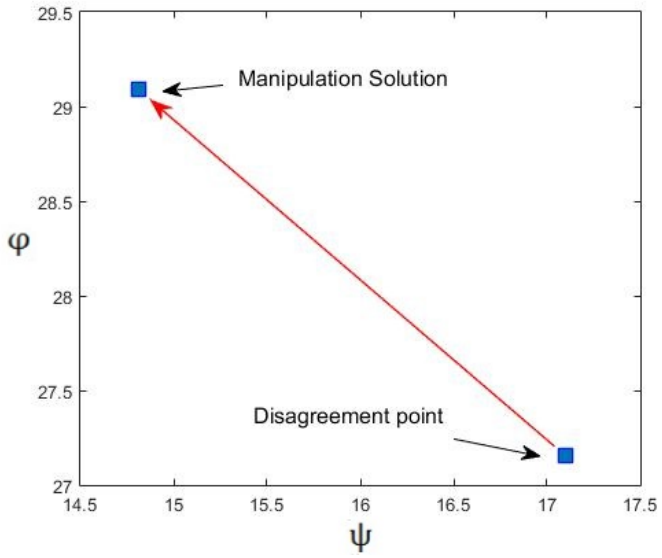


Figure 7.5 Manipulation Solution.

Chapter 8

The Kalai-Smorodinsky bargaining solution

8.1 Introduction

One of the most prominent alternatives to the Nash bargaining solution is the solution presented by Raiffa [77] for two-person bargaining games, which has been axiomatized by Kalai and Smorodinsky [46] suggesting an alternative axiom, the axiom of Monotonicity, which leads to another unique solution. Considering a two-person bargaining situation, this axiom states that if, for every utility level that player one may demand, the maximum feasible utility level that player two can simultaneously reach is increased; then the utility level assigned to player two according to the solution should also be increased.

Anant et al. [3] showed that the Kalai-Smorodinsky's result is true and the unique one satisfying the four axioms even if it is generalized the domain of bargaining games to allow for non-convex utility feasibility sets. Roth [80] showed that the Kalai-Smorodinsky solution for two-person bargaining games does not generalize in a straightforward manner to general n-person bargaining games. Specifically, the solution is not Pareto Optimal on the class of all n-person bargaining games, and no solution which is can possess the other properties which characterize in the two-person case.

Following this problem, Peters and Tijs [71] introduced a rather large subclass of n-person bargaining games. They described all bargaining solutions on this subclass having the four axioms of the Kalai-Smorodinsky solution, and exactly one of these solutions is symmetric; also, they proved that all these solutions are risk sensitive.

8.2 The Bargaining Model

With the property of independence of irrelevant alternatives, Nash's solution is not sensitive to the range of outcomes contained in the feasible set, for instance, by the utopia point $\psi^*(\Psi,\phi) = (\psi^{1*}(\Psi,\phi), ..., \psi^{\mathcal{N}*}(\Psi,\phi))$ defined by

$$\psi^{l*}(\Psi,\phi) = \max\left\{\psi^l : \psi^l \in \Psi, \psi^l \geq \phi^l\right\}$$

this point is the highest possible utility payoff that player l can attain in the bargaining problem (Ψ,ϕ). Raiffa [77] proposed a solution for for two-players bargaining problems which is sensitive to changes in $\psi^*(\Psi,\phi)$. He proposed the solution ψ for two-player games such that $\psi = f(\Psi,\phi)$ is the Pareto-optimal point at which $(\psi^1-\phi^1)/(\psi^{1*}-\phi^1) = (\psi^2-\phi^2)/(\psi^{2*}-\phi^2)$. The solution ψ selects the maximal point on the line joining ϕ to ψ^*, yielding each player the largest reward consistent with the constraint that the players' actual gains should be in proportion to their maximum gains, as measured by the ideal point $\psi^*(\Psi,\phi)$.

The Kalai-Smorodinsky solution of the bargaining problem amounts to normalizing the utility function of each agent in such a way that it is worth zero at the status-quo and one at this agent's best outcome, given that all others get al least their status quo utility level; and to share equally the benefit from cooperation. This solution has been proposed by Raiffa [77] and axiomatically characterized by Kalai and Smorodinsky [46] when society $\mathcal{N}$ contains only two agents, i.e., $l = 1, 2$. Consider the pair (Ψ,ϕ), where the point in the plane $\phi = (\phi^1,\phi^2)$ is the level of utility that player $l = 1, 2$ receives if the two players do not cooperate with each other, this point is called the status quo; and Ψ is a subset of the plane, every point $\psi = (\psi^1,\psi^2) \in \Psi$ represents levels of utility for players 1 and 2 that can be reached by an outcome of the game which is feasible for the two players when they do cooperate.

Let $\mathcal{B}$ denote the set of all pairs (Ψ,ϕ) such that

1. $\Psi \subset \mathbb{R}^2$ is compact, convex;

2. There exists at least one point $\psi \in L$ such that $\psi^l > \phi^l$, for $l = 1, 2$.

A Kalai-Smorodinsky solution to the bargaining problem is a function $f : \mathcal{B} \to \mathbb{R}^2$ such that $f(\Psi,\phi) \in \Psi$ and satisfies the following axioms [46]

1. Pareto Optimality: For every $(\Psi, \phi) \in \mathcal{B}$ there is no $\psi \in \Psi$ such that $\psi \geq f(\Psi, \phi)$ and imply $\psi \neq f(\Psi, \phi)$.

2. Symmetry: We let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(\psi^1, \psi^2) = (\psi^2, \psi^1)$ and we require that for every $(\Psi, \phi) \in \mathcal{B}$, $f(T(\Psi), T(\phi)) = T(f(\Psi, \phi))$.

3. Invariance with respect to affine transformations of utility: A is an affine transformation of utility if $A = (A^1, A^2) : \mathbb{R}^2 \to \mathbb{R}^2$, $A((\psi^l, \psi^2)) = (A^1(\psi^l), A^2(\psi^2))$, and the maps $A^l(\psi)$ are of the form $c^l\psi + d^l$ for some positive constant c^l and some constant d^l. We require that for A, $f(A(L), A(\phi)) = A(f(\Psi, \phi))$.

4. Monotonicity: For a pair $(\Psi, \phi) \in \mathcal{B}$, let $\psi^*(\Psi) = (\psi^{1*}(\Psi), \psi^{2*}(\Psi))$ and $g_\Psi(\psi^1)$ be a function defined for $\psi^1 \leq \psi^{1*}(\Psi)$ in the following way

$$g_L(\psi^1) = \begin{cases} \psi^2, & \text{if} \quad (\psi^1, \psi^2) \text{ is the Pareto of} (\Psi, \phi) \\ \psi^{2*}(\Psi), & \text{if} \quad \text{there is no such } \psi^2 \end{cases}$$

   If (Ψ^2, ϕ) and (Ψ^1, ϕ) are bargaining pairs such that $\psi^{1*}(\Psi^1) = \psi^{1*}(\Psi^2)$ and $g_{\Psi^1} \leq g_{\Psi^2}$, then $f^2(\Psi^1, \phi) \leq f^2(\Psi^2, \phi)$, where $f(\Psi, \phi) = (f^1(\Psi, \phi), f^2(\Psi, \phi))$.

Consider a Pareto optimal outcome and the line segments connecting that outcome to the disagreement point and to the utopia point. For any pair of players we may then project these line segments into the plane (see Figure 8.1).

The axiom of monotonicity states that if, for every utility level that player 1 may demand, the maximum feasible utility level that player 2 can simultaneously reach is increased, then the utility level assigned to player 2 according to the solution should also be increased.

8.2.1 Generalization of the Kalai-Smorodinsky solution for $\mathcal{N}$-player

We consider the set of all $\mathfrak{n}$-player bargaining problems defined by Peters and Tijs [71], and on this set we define a class of asymmetric $\mathfrak{n}$-person Kalai-Smorodinsky solutions. The set of players $\mathfrak{n} = \{1, ..., \mathfrak{n}\}$ is indexed by $l = (1, ..., \mathfrak{n})$, with $\mathfrak{n} \geq 2$. A set $\Psi \subseteq \mathbb{R}^{\mathfrak{n}}$ is comprehensive if $x \in \Psi$ and $x \geq y$ imply $y \in \Psi$, for all $x, y \in \mathbb{R}^{\mathfrak{n}}$.

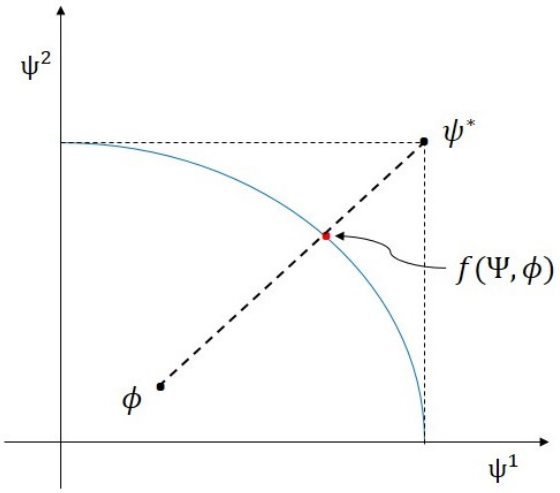


Figure 8.1 The Kalai-Smorodinsky solution.

We talk about comprehensiveness in the sense that any player can choose a lower utility without this leading to an infeasible outcome. A bargaining problem for $\mathcal{N}$ is a pair (Ψ, ϕ) where: $\Psi \subseteq \mathbb{R}^{\mathfrak{n}}$ is compact, convex, and comprehensive; and there exists a $\psi \in \Psi$ such that $\psi > \phi$ and $\phi \in \Psi$. For all bargaining problem $(\Psi, \phi) \in B^{\mathfrak{n}}$ we define the Pareto set of Ψ as

$$P(\Psi) = \{\psi \in \Psi : \text{ for all } x \in \mathbb{R}^{\mathfrak{n}}, \text{ if } x \geq \psi \text{ and } x \neq \psi, \text{ then } x \notin \Psi\}$$

A bargaining solution is a map $f : B^{\mathfrak{n}} \to \mathbb{R}^{\mathfrak{n}}$ that assigns to each bargaining problem $(\Psi, \phi) \in B^{\mathfrak{n}}$ a single point $f(\Psi, \phi) \in \Psi$. Roth [80] observed that the $\mathfrak{n}$-player extension of the Kalai-Smorodinsky solution is not Pareto optimal on all bargaining problems in $B^{\mathfrak{n}}$, i.e., does not assign an element of $P(\Psi)$ to each $(\Psi, \phi) \in B^{\mathfrak{n}}$. Therefore, Peters and Tijs [71] introduced a subclass of bargaining problems in $B^{\mathfrak{n}}$ for which this problem does not occur.

Property 8.1 *For all $\psi \in \Psi$, $\psi \geq \phi$, $l = (1, ..., \mathfrak{n})$: $\psi \notin P(\Psi)$ and $\psi^l < \psi^{l*}(\Psi, \phi) \Rightarrow \exists\, \varepsilon > 0$ with $\psi + \varepsilon e^l \in \Psi$, where the vector $e^l \in \mathbb{R}^{\mathfrak{n}}$ has the l-th coordinate equal to 1 and all other coordinates equal to 0.*

If a feasible outcome ψ is not Pareto optimal, then for any player l who receives less than his utopia payoff it is possible to increase his utility while all other players still receive ψ. Let $\mathcal{I}^{\mathfrak{n}} \subseteq B^{\mathfrak{n}}$ consist of all bargaining problems satisfying Property 8.1. The class of bargaining

problems $(\Psi, 0) \in \mathcal{I}^{\mathfrak{n}}$ is denoted by $\mathcal{I}_0^{\mathfrak{n}}$. Peters and Tijs [71] defined the $\mathfrak{n}$-player extension of the solution by making use of monotonic curves. A monotonic curve for $\mathfrak{n}$ is a map

$$\vartheta : [1, \mathfrak{n}] \to \left\{ \psi \in \mathbb{R}_+^{\mathfrak{n}} \mid \psi^l \leq 1 \text{ for all player } l, \text{ and } 1 \leq \sum_{l=1}^{\mathfrak{n}} \psi^l \right\}$$

such that for all $1 \leq s \leq t \leq \mathfrak{n}$ we have $\vartheta(s) \leq \vartheta(t)$ and $\sum_{l=1}^{\mathfrak{n}} \vartheta^l(s) = s$. The set of all monotonic curves for $\mathfrak{n}$ is denoted by $\Theta^{\mathfrak{n}}$.

Lemma 8.2 *Peters and Tijs [71]. For each $\vartheta \in \Theta^{\mathfrak{n}}$ and $(\Psi, 0) \in \mathcal{I}_0^{\mathfrak{n}}$ with $f(\Psi, 0) = e^{\mathfrak{n}}$, the set $P(\Psi) \cap \{\vartheta(t) \mid t \in [1, \mathfrak{n}]\}$ contains exactly one point.*

Let ϑ be some monotonic curve in $\Theta^{\mathfrak{n}}$. Following Lemma 8.2, the solution associated with ϑ is defined as $\rho^{\vartheta} : \mathcal{I}^{\mathfrak{n}} \to \mathbb{R}^{\mathfrak{n}}$. Let $(\Psi, 0) \in \mathcal{I}_0^{\mathfrak{n}}$, if $\psi^*(\Psi, 0) = e^{\mathfrak{n}}$, then

$$\left\{\rho^{\vartheta}(\Psi, 0)\right\} := P(\Psi) \cap \{\vartheta(t) \mid t \in [1, \mathfrak{n}]\}$$

and if $\psi^*(\Psi, 0) = \psi^*$, then $\rho^{\vartheta}(\Psi, 0) := \psi^* \rho^{\vartheta}((\psi^*)^{-1}\Psi)$. For $(\Psi, \phi) \in \mathcal{I}^{\mathfrak{n}}$, we define $\rho^{\vartheta}(\Psi, \phi) = \phi + \rho^{\vartheta}(\Psi - \phi)$. The class of all solutions associated with a monotonic curve in $\Theta^{\mathfrak{n}}$ is referred to as the class of individually monotonic bargaining solutions, the Kalai-Smorodinsky solution is an element of this class. Observe that $\hat{\vartheta}$, the monotonic curve of the Kalai-Smorodinsky solution, defines a straight line in $\mathbb{R}^{\mathfrak{n}}$, which for bargaining games $\Psi \in \mathcal{I}_0^{\mathfrak{n}}$ with $\psi^*(\Psi, 0) = e^{\mathfrak{n}}$, coincides with the line connecting the disagreement point 0 and the utopia point $e^{\mathfrak{n}}$. For general bargaining problems $(\Psi, \phi) \in \mathcal{I}^{\mathfrak{n}}$, the solution is the intersection of the Pareto set $P(\Psi)$ and the straight line that connects the disagreement point ϕ and the utopia point ψ^*.

8.3 Formulation of the problem for Markov chains games

Consider a $\mathfrak{n}$-person bargaining problem [104]. Denote the disagreement utility for each player $(l = 1, ..., \mathfrak{n})$ that depends on the strategies $c_{(i,k)}^l$ as $\phi^l(c^1, ..., c^{\mathfrak{n}})$, and the solution for the bargaining problem as the point $(\psi^1, ..., \psi^{\mathfrak{n}})$. Following (2.5), the utilities ψ^l are for Markov chains as follows

$$\psi^l = \psi^l\left(c^1, ..., c^{\mathfrak{n}}\right) := \sum_{i=1}^{N} \sum_{k=1}^{M} W_{(i,k)}^l \prod_{l=1}^{N} c_{(i,k)}^l$$

where the matrices $W^l_{(i,k)}$ represent the behavior of each player. This point is better than the disagreement point, therefore must satisfy that $\psi^l > \phi^l$.

The process to solve the bargaining problem consists of two main steps: firstly, to find the disagreement point we define it as the Nash equilibrium point of the problem (see [62]), while for the solution of the bargaining process we follow the model presented by Kalai and Smorodinsky [46]. The Kalai-Smorodinsky solution chooses the maximum individually rational payoff profile at which each player's payoff has the same proportion from disagreement point to the utopia point. For solving the bargaining problem we consider there exists an optimal solution that is a strong Pareto optimal point and it is the closest solution to the utopia point. We formulate the problem as the L_p-norm to find the Pareto optimal solution, this formulation reduces the distance to the utopian point in the Euclidean space. Following the model presented in Chapter 2, the function for finding the solution to the bargaining problem is

$$g(c^1, ..., c^{\mathfrak{n}}) = \left[\sum_{l=1}^{\mathfrak{n}} \left| \lambda^l \frac{(\psi^l - \phi^l)^{\alpha^l \chi(\psi^l > \phi^l)}}{(\psi^{l*} - \phi^l)^{\alpha^l \chi(\psi^{l*} > \phi^l)}} \right|^p \right]^{1/p} \tag{8.1}$$

where ψ^{l*} is the utopia point, $\alpha^l \geq 0$ are weighting parameters for each player, and $\lambda \in \Lambda^{\mathfrak{n}}$ such that

$$\Lambda^{\mathfrak{n}} := \left\{ \lambda \in \mathbb{R}^{\mathfrak{n}} : \lambda \in [0,1]\,,\ \sum_{l=1}^{\mathfrak{n}} \lambda^l = 1 \right\}$$

We can rewrite (8.1) for purposes of implementation as follows

$$\tilde{g}(c^1, ..., c^{\mathfrak{n}}) = \left[\sum_{l=1}^{\mathfrak{n}} \lambda^l \left| \alpha^l \chi(\psi^l > \phi^l) \ln(\psi^l - \phi^l) - \alpha^l \chi(\psi^{l*} > \phi^l) \ln(\psi^{l*} - \phi^l) \right|^p \right]^{1/p}$$

Thus, the strategy x^*, which is the vector $x^* = (c^1, ..., c^{\mathfrak{n}}) \in X_{\text{adm}} := \bigotimes_{l=1}^{\mathfrak{n}} C^l_{\text{adm}}$, is the solution for the bargaining problem

$$x^* \in \ \text{Arg} \max_{x \in X_{\text{adm}}, \lambda \in \Lambda^{\mathfrak{n}}} \ \left\{ \tilde{g}(c^1, ..., c^{\mathfrak{n}}) \right\}$$

the strategies $c^l_{(i,k)}$ satisfy the restrictions (2.7), (2.8) and (2.9). Applying the Lagrange principle it follows that

$$\mathcal{L}(x, \lambda, \mu, \xi, \eta) = \tilde{g}(c^1, ..., c^{\mathfrak{n}}) - \sum_{l=1}^{\mathfrak{n}} \sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(c^l) -$$

$$\sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \xi^l_{(j)} q^l_{(j|i,k)} c^l_{(i,k)} - \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{k=1}^{M} \eta^l \left(c^l_{(i,k)} - 1 \right)$$

The solution obtained by the Tikhonov's regularization with $\delta > 0$ is given by

$$x^*, \lambda^*, \mu^*, \xi^*, \eta^* = \arg \max_{x \in X_{\text{adm}}, \lambda \in \Lambda^{\mathfrak{n}}} \min_{\mu, \xi, \eta \geq 0} \mathcal{L}_\delta(x, \lambda, \mu, \xi, \eta)$$

where

$$\begin{aligned} \mathcal{L}_\delta(x, \lambda, \mu, \xi, \eta) = \tilde{g}(c^1, ..., c^{\mathfrak{n}}) - \sum_{l=1}^{\mathfrak{n}} \sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(c^l) - \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{k=1}^{M} \eta^l \left(c^l_{(i,k)} - 1 \right) - \\ \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \xi^l_{(j)} q^l_{(j|i,k)} c^l_{(i,k)} - \frac{\delta}{2} \left(\|x\|^2 + \|\lambda\|^2 - \|\mu\|^2 - \|\xi\|^2 - \|\eta\|^2 \right) \end{aligned} \tag{8.2}$$

Notice that the Lagrange function (8.2) satisfies the saddle-point condition, namely, for all $x \in X_{\text{adm}}$, $\lambda \in \Lambda^{\mathfrak{n}}$ and $\mu, \xi, \eta \geq 0$ we have

$$\mathcal{L}_\delta(x_\delta, \lambda_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \leq \mathcal{L}_\delta(x^*_\delta, \lambda^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \leq \mathcal{L}_\delta(x^*_\delta, \lambda^*_\delta, \mu_\delta, \xi_\delta, \eta_\delta)$$

8.3.1 The proximal format

In the proximal format (see [5]) the relation (8.2) can be expressed as

$$\begin{aligned} \mu^*_\delta &= \arg \min_{\mu \geq 0} \left\{ \tfrac{1}{2} \|\mu - \mu^*_\delta\|^2 + \gamma \mathcal{L}_\delta(x^*_\delta, \lambda^*_\delta, \mu, \xi^*_\delta, \eta^*_\delta) \right\} \\ \xi^*_\delta &= \arg \min_{\xi \geq 0} \left\{ \tfrac{1}{2} \|\xi - \xi^*_\delta\|^2 + \gamma \mathcal{L}_\delta(x^*_\delta, \lambda^*_\delta, \mu^*_\delta, \xi, \eta^*_\delta) \right\} \\ \eta^*_\delta &= \arg \min_{\eta \geq 0} \left\{ \tfrac{1}{2} \|\eta - \eta^*_\delta\|^2 + \gamma \mathcal{L}_\delta(x^*_\delta, \lambda^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta) \right\} \\ x^*_\delta &= \arg \max_{x \in X} \left\{ -\tfrac{1}{2} \|x - x^*_\delta\|^2 + \gamma \mathcal{L}_\delta(x, \lambda^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \right\} \\ \lambda^*_\delta &= \arg \max_{\lambda \in \Lambda^{\mathfrak{n}}} \left\{ -\tfrac{1}{2} \|\lambda - \lambda^*_\delta\|^2 + \gamma \mathcal{L}_\delta(x^*_\delta, \lambda, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \right\} \end{aligned} \tag{8.3}$$

where the solutions x^*_δ, λ^*_δ, μ^*_δ, ξ^*_δ and η^*_δ depend on the parameters $\delta, \gamma > 0$.

8.3.2 The Extraproximal method

We design the method for the static bargaining game in a general format with some fixed admissible initial values ($x_0 \in X$, $\lambda_0 \in \Lambda^{\mathfrak{n}}$ and $\mu_0, \xi_0, \eta_0 \geq 0$), considering that we want to maximize the function as follows:

1. The *first half-step* (prediction):

$$\begin{aligned}
\bar{\mu}_n &= \arg\max_{\mu\geq 0}\left\{-\tfrac{1}{2}\|\mu-\mu_n\|^2-\gamma\mathcal{L}_\delta(x_n,\lambda_n,\mu,\xi_n,\eta_n)\right\}\\
\bar{\xi}_n &= \arg\max_{\xi\geq 0}\left\{-\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_\delta(x_n,\lambda_n,\bar{\mu}_n,\xi,\eta_n)\right\}\\
\bar{\eta}_n &= \arg\max_{\eta\geq 0}\left\{-\tfrac{1}{2}\|\eta-\eta_n\|^2-\gamma\mathcal{L}_\delta(x_n,\lambda_n,\bar{\mu}_n,\bar{\xi}_n,\eta)\right\}\\
\bar{x}_n &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_n\|^2+\gamma\mathcal{L}_\delta(x,\lambda_n,\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}\\
\bar{\lambda}_n &= \arg\max_{\lambda\in\Lambda^{\mathrm{n}}}\left\{-\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(x_n,\lambda,\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}
\end{aligned} \tag{8.4}$$

2. The *second half-step* (basic)

$$\begin{aligned}
\mu_{n+1} &= \arg\max_{\mu\geq 0}\left\{-\tfrac{1}{2}\|\mu-\mu_n\|^2-\gamma\mathcal{L}_\delta(\bar{x}_n,\bar{\lambda}_n,\mu,\bar{\xi}_n,\bar{\eta}_n)\right\}\\
\xi_{n+1} &= \arg\max_{\xi\geq 0}\left\{-\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_\delta(\bar{x}_n,\bar{\lambda}_n,\bar{\mu}_n,\xi,\bar{\eta}_n)\right\}\\
\eta_{n+1} &= \arg\max_{\eta\geq 0}\left\{-\tfrac{1}{2}\|\eta-\eta_n\|^2-\gamma\mathcal{L}_\delta(\bar{x}_n,\bar{\lambda}_n,\bar{\mu}_n,\bar{\xi}_n,\eta)\right\}\\
x_{n+1} &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_n\|^2+\gamma\mathcal{L}_\delta(x,\bar{\lambda}_n,\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}\\
\lambda_{n+1} &= \arg\max_{\lambda\in\Lambda^{\mathrm{n}}}\left\{-\tfrac{1}{2}\|\lambda-\lambda_n\|^2+\gamma\mathcal{L}_\delta(\bar{x}_n,\lambda,\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}
\end{aligned} \tag{8.5}$$

8.3.3 Convergence Analysis

Define the following extended vectors

$$\tilde{x}=\begin{pmatrix} x\\ \lambda\end{pmatrix}\in\tilde{X}:=X\times\mathbb{R}^+,\qquad \tilde{\mu}=\begin{pmatrix}\mu\\ \xi\\ \eta\end{pmatrix}\in\mathbb{R}^+\times\mathbb{R}^+\times\mathbb{R}^+$$

The regularized Lagrange function can be expressed as $\tilde{\mathcal{L}}_\delta(\tilde{x},\tilde{\mu}):=\mathcal{L}_\delta(x,\lambda,\mu,\xi,\eta)$. And the equilibrium point that satisfies (8.3) can be expressed as

$$\begin{aligned}
\tilde{\mu}_\delta^* &= \arg\min_{\tilde{\mu}\geq 0}\left\{\tfrac{1}{2}\|\tilde{\mu}-\tilde{\mu}_\delta^*\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{x}_\delta^*,\tilde{\mu})\right\}\\
\tilde{x}_\delta^* &= \arg\max_{\tilde{x}\in\tilde{X}}\left\{-\tfrac{1}{2}\|\tilde{x}-\tilde{x}_\delta^*\|^2+\gamma\tilde{\mathcal{L}}_\delta(\tilde{x},\tilde{\mu}_\delta^*)\right\}
\end{aligned}$$

Now, introducing the following variables

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix} \in \tilde{X} \times \mathbb{R}^+, \quad \tilde{z} = \begin{pmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{pmatrix} \in \tilde{X} \times \mathbb{R}^+$$

and then, the Lagrange function in terms of $\tilde{y}$ and $\tilde{z}$ can be defined as $L_\delta(\tilde{y}, \tilde{z}) := \mathcal{L}_\delta(\tilde{y}_1, \tilde{z}_2) - \mathcal{L}_\delta(\tilde{z}_1, \tilde{y}_2)$. For $\tilde{y}_1 = \tilde{x}$, $\tilde{y}_2 = \tilde{\mu}$, $\tilde{z}_1 = \tilde{z}_1^* = \tilde{x}_\delta^*$ and $\tilde{z}_2 = \tilde{z}_2^* = \tilde{\mu}_\delta^*$ we have that $L_\delta(\tilde{y}, \tilde{z}^*) := \tilde{\mathcal{L}}_\delta(\tilde{x}, \tilde{\mu}_\delta^*) - \tilde{\mathcal{L}}_\delta(\tilde{x}_\delta^*, \tilde{\mu})$. In these variables the relation (8.3) can be represented by

$$\tilde{z}^* = \arg \max_{\tilde{y} \in \tilde{X} \times \mathbb{R}^+} \left\{ -\tfrac{1}{2} \|\tilde{y} - \tilde{z}^*\|^2 + \gamma L_\delta(\tilde{y}, \tilde{z}^*) \right\} \tag{8.6}$$

Finally, we have that the extraproximal method can be expressed by

1. First step

$$\hat{z}_n = \arg \max_{\tilde{y} \in \tilde{X} \times \mathbb{R}^+} \left\{ -\tfrac{1}{2} \|\tilde{y} - \tilde{z}_n\|^2 + \gamma L_\delta(\tilde{y}, \tilde{z}_n) \right\} \tag{8.7}$$

2. Second step

$$\tilde{z}_{n+1} = \arg \max_{\tilde{y} \in \tilde{X} \times \mathbb{R}^+} \left\{ -\tfrac{1}{2} \|\tilde{y} - \tilde{z}_n\|^2 + \gamma L_\delta(\tilde{y}, \hat{z}_n) \right\} \tag{8.8}$$

Please refer to Appendix C for the results and proofs of convergence analysis.

8.4 The disagreement point model

If negotiations break down and no agreement is reached, then inevitably the disagreement point (also called as *status quo* or *threat point*) will take effect. Following the model presented in section 6.3 but with the consideration that this model is related to continuous time Markov chains, i.e., the strategies $c^l_{(i,k)}$ satisfy the restrictions (2.7), (2.8) and (2.9).

Applying the regularized Lagrange principle we have the solution for the disagreement point

$$x^*, \hat{x}^*, \mu^*, \xi^*, \eta^* = \arg \max_{x \in X, \hat{x} \in \hat{X}} \min_{\mu, \xi, \eta \geq 0} \mathcal{L}_{\theta,\delta}(x, \hat{x}, \mu, \xi, \eta)$$

where

$$\mathcal{L}_{\theta,\delta}(x, \hat{x}, \mu, \xi, \eta) := (1-\theta) f(x, \hat{x}) - \sum_{l=1}^{\mathfrak{n}} \sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(c^l) - \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{k=1}^{M} \eta^l \left(c^l_{(i,k)} - 1 \right) -$$

$$\sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \xi^l_{(j)} q^l_{(j|i,k)} c^l_{(i,k)} - \tfrac{\delta}{2} \left(\|x\|^2 + \|\hat{x}\|^2 - \|\mu\|^2 - \|\xi\|^2 - \|\eta\|^2 \right) \tag{8.9}$$

Notice that the Lagrange function (8.9) satisfies the saddle-point condition, namely, for all $x \in X$, $\hat{x} \in \hat{X}$, and $\mu, \xi, \eta \geq 0$ we have

$$\mathcal{L}_{\theta,\delta}(x_\delta, \hat{x}_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \leq \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \leq \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}^*_\delta, \mu_\delta, \xi_\delta, \eta_\delta)$$

8.4.1 The proximal format

In the proximal format the relation (8.9) can be expressed as

$$\mu^*_\delta = \arg\min_{\mu \geq 0} \left\{ \tfrac{1}{2}\|\mu - \mu^*_\delta\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}^*_\delta, \mu, \xi^*_\delta, \eta^*_\delta) \right\}$$

$$\xi^*_\delta = \arg\min_{\xi \geq 0} \left\{ \tfrac{1}{2}\|\xi - \xi^*_\delta\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}^*_\delta, \mu^*_\delta, \xi, \eta^*_\delta) \right\}$$

$$\eta^*_\delta = \arg\min_{\eta \geq 0} \left\{ \tfrac{1}{2}\|\eta - \eta^*_\delta\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta) \right\}$$

$$x^*_\delta = \arg\max_{x \in X} \left\{ -\tfrac{1}{2}\|x - x^*_\delta\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x, \hat{x}^*_\delta, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \right\}$$

$$\hat{x}^*_\delta = \arg\max_{\hat{x} \in \hat{X}} \left\{ -\tfrac{1}{2}\|\hat{x} - \hat{x}^*_\delta\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x^*_\delta, \hat{x}, \mu^*_\delta, \xi^*_\delta, \eta^*_\delta) \right\}$$

where the solutions x^*_δ, $\hat{x}^*_\delta$, μ^*_δ, ξ^*_δ and η^*_δ depend on the parameters $\delta, \gamma > 0$.

8.4.2 The Extraproximal method

We design the method for the static Nash game in a general format with some fixed admissible initial values ($x_0 \in X$, $\hat{x}_0 \in \hat{X}$, and $\mu_0, \xi_0, \eta_0 \geq 0$), considering that we want to maximize the function, as follows:

1. The *first half-step*:

$$\bar{\mu}_n = \arg\max_{\mu \geq 0} \left\{ -\tfrac{1}{2}\|\mu - \mu_n\|^2 - \gamma \mathcal{L}_{\theta,\delta}(x_n, \hat{x}_n, \mu, \xi_n, \eta_n) \right\}$$

$$\bar{\xi}_n = \arg\max_{\xi \geq 0} \left\{ -\tfrac{1}{2}\|\xi - \xi_n\|^2 - \gamma \mathcal{L}_{\theta,\delta}(x_n, \hat{x}_n, \bar{\mu}_n, \xi, \eta_n) \right\}$$

$$\bar{\eta}_n = \arg\max_{\eta \geq 0} \left\{ -\tfrac{1}{2}\|\eta - \eta_n\|^2 - \gamma \mathcal{L}_{\theta,\delta}(x_n, \hat{x}_n, \bar{\mu}_n, \bar{\xi}_n, \eta) \right\} \tag{8.10}$$

$$\bar{x}_n = \arg\max_{x \in X} \left\{ -\tfrac{1}{2}\|x - x_n\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x, \hat{x}_n, \bar{\mu}_n, \bar{\xi}_n, \bar{\eta}_n) \right\}$$

$$\bar{\hat{x}}_n = \arg\max_{\hat{x} \in \hat{X}} \left\{ -\tfrac{1}{2}\|\hat{x} - \hat{x}_n\|^2 + \gamma \mathcal{L}_{\theta,\delta}(x_n, \hat{x}, \bar{\mu}_n, \bar{\xi}_n, \bar{\eta}_n) \right\}$$

2. The *second half-step*

$$
\begin{aligned}
\mu_{n+1} &= \arg\max_{\mu\geq 0}\left\{-\tfrac{1}{2}\|\mu-\mu_n\|^2-\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\overline{\hat{x}}_n,\mu,\bar{\xi}_n,\bar{\eta}_n)\right\}\\
\xi_{n+1} &= \arg\max_{\xi\geq 0}\left\{-\tfrac{1}{2}\|\xi-\xi_n\|^2-\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\overline{\hat{x}}_n,\bar{\mu}_n,\xi,\bar{\eta}_n)\right\}\\
\eta_{n+1} &= \arg\max_{\eta\geq 0}\left\{-\tfrac{1}{2}\|\eta-\eta_n\|^2-\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\overline{\hat{x}}_n,\bar{\mu}_n,\bar{\xi}_n,\eta)\right\}\\
x_{n+1} &= \arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x_n\|^2+\gamma\mathcal{L}_{\theta,\delta}(x,\overline{\hat{x}}_n,\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}\\
\hat{x}_{n+1} &= \arg\max_{\hat{x}\in \hat{X}}\left\{-\tfrac{1}{2}\|\hat{x}-\hat{x}_n\|^2+\gamma\mathcal{L}_{\theta,\delta}(\bar{x}_n,\hat{x},\bar{\mu}_n,\bar{\xi}_n,\bar{\eta}_n)\right\}
\end{aligned}
\tag{8.11}
$$

8.5 Numerical Example

In this example, we are considering a bargaining on the labor market between three contracting parties corresponding to the government, an employers' federation and a labor union which are often characterized by reciprocal incremental concessions. They are aimed at agreements to regulate working salaries, conditions, benefits, and other aspects of workers' compensation and rights for workers. We expect that the contracting parties carry out a negotiation process that reach a Pareto efficient outcome when solving their differences. The government plays a fundamental role, because can change the equilibrium on the labor market by modifying the reserves related to the wage. This enables us to analyze the convergence of the equilibrium of the labor market in terms of a continuous-time approach considering the changes in the reservation wage along the time, or more generally with respect to public policy. In such scenario, the Kalai-Smorodinsky solution can be applied to labor-market negotiations. The Kalai-Smorodinsky approach is distinguished by the equal proportional concessions of for the three parties in conflict. In this sense, this negotiation process gives the impression to be more intuitive than the Nash bargaining model in representing a solution for the labor market problem, because each party makes concessions with respect to its initial demands. In this bargaining process, it is expected that the parties progressively moderate their demands until an agreement is reached, sooner or later.

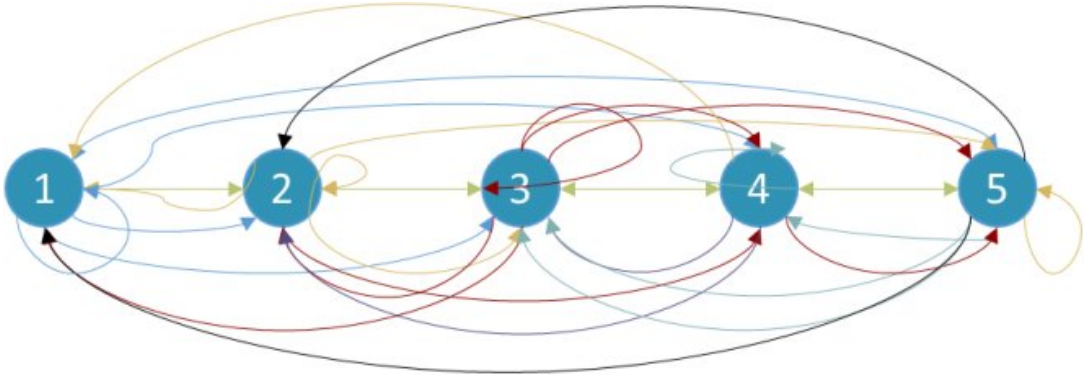


Figure 8.2 Markov chain of the labor market problem.

Our goal is to analyze a three-player bargaining situation on the labor market in a class of continuous time Markov chains using the Kalai-Smorodinsky approach. We assume a continuous-time Markov process defined over a discrete state-space S where the labor force is divided into five possible states, namely, employed (E), unemployed (U), out of the labor force (O), inactivity (I), retired (R). This involves each individual in a labor path. The labor dynamics information is contained in the transition matrices, which represents the individual's characteristics in the long-run time. The set of actions A is a finite space determined by two different actions $A = \{agree, disagree\}$ (see Figure 8.2).

Denote the disagreement cost that depends on the strategies $c^l_{(i,k)}$ for players $l = 1, 2, 3$ (the government, an employers' federation and a labor union) as $\phi^l(c^1, c^2, c^3)$ and the solution for the bargaining problem as the point (ψ^1, ψ^2, ψ^3). Let the number of states $N = 5$ ($S = \{E, U, O, I, R\}$), and the actions $M = 2$ ($A = \{agree, disagree\}$). The individual utility for each player on the labor market are defined by

$$U^1_{(i,j,1)} = \begin{bmatrix} 10 & 8 & 13 & 7 & 6 \\ 11 & 19 & 6 & 8 & 10 \\ 9 & 7 & 13 & 19 & 5 \\ 14 & 9 & 15 & 12 & 16 \\ 12 & 14 & 9 & 8 & 10 \end{bmatrix} \quad U^1_{(i,j,2)} = \begin{bmatrix} 12 & 9 & 7 & 10 & 15 \\ 16 & 0 & 9 & 14 & 6 \\ 18 & 10 & 16 & 9 & 4 \\ 12 & 16 & 9 & 8 & 13 \\ 11 & 9 & 13 & 17 & 10 \end{bmatrix}$$

$$U^2_{(i,j,1)} = \begin{bmatrix} 7 & 9 & 12 & 6 & 10 \\ 18 & 12 & 7 & 9 & 10 \\ 5 & 14 & 8 & 11 & 16 \\ 10 & 13 & 8 & 14 & 10 \\ 16 & 9 & 12 & 10 & 8 \end{bmatrix} \quad U^2_{(i,j,2)} = \begin{bmatrix} 5 & 17 & 6 & 9 & 11 \\ 8 & 9 & 14 & 12 & 8 \\ 10 & 13 & 9 & 1 & 18 \\ 12 & 18 & 15 & 9 & 4 \\ 17 & 13 & 9 & 5 & 6 \end{bmatrix}$$

$$U^3_{(i,j,1)} = \begin{bmatrix} 15 & 9 & 10 & 4 & 10 \\ 12 & 5 & 4 & 0 & 11 \\ 19 & 7 & 6 & 13 & 10 \\ 2 & 16 & 10 & 9 & 7 \\ 10 & 6 & 9 & 14 & 10 \end{bmatrix} \quad U^3_{(i,j,2)} = \begin{bmatrix} 10 & 7 & 12 & 19 & 1 \\ 16 & 9 & 0 & 10 & 12 \\ 19 & 3 & 4 & 13 & 8 \\ 2 & 8 & 1 & 19 & 14 \\ 1 & 13 & 9 & 15 & 16 \end{bmatrix}$$

The transition rate matrices for each player are defined as follows

$$q^1_{(i,j,1)} = \begin{bmatrix} -0.8402 & 0.1555 & 0.1069 & 0.3312 & 0.2466 \\ 0.1090 & -0.7197 & 0.2965 & 0.0878 & 0.2264 \\ 0.4127 & 0.3376 & -1.4437 & 0.3286 & 0.3648 \\ 0.8641 & 0.3679 & 0.3194 & -1.7086 & 0.1571 \\ 0.2886 & 0.1649 & 0.2348 & 0.1507 & -0.8390 \end{bmatrix}$$

$$q^1_{(i,j,2)} = \begin{bmatrix} -0.3601 & 0.0666 & 0.0458 & 0.1420 & 0.1057 \\ 0.0467 & -0.3085 & 0.1271 & 0.0376 & 0.0970 \\ 0.1769 & 0.1447 & -0.6188 & 0.1408 & 0.1563 \\ 0.3703 & 0.1577 & 0.1369 & -0.7323 & 0.0673 \\ 0.1237 & 0.0707 & 0.1006 & 0.0646 & -0.3596 \end{bmatrix}$$

$$q^2_{(i,j,1)} = \begin{bmatrix} -0.1017 & 0.0234 & 0.0236 & 0.0424 & 0.0125 \\ 0.0104 & -0.0819 & 0.0147 & 0.0333 & 0.0235 \\ 0.0144 & 0.0285 & -0.0790 & 0.0235 & 0.0125 \\ 0.0234 & 0.0526 & 0.0456 & -0.1441 & 0.0225 \\ 0.0243 & 0.0322 & 0.0315 & 0.0433 & -0.1311 \end{bmatrix}$$

$$q^2_{(i,j,2)}=\begin{bmatrix} -0.9155 & 0.2102 & 0.2120 & 0.3812 & 0.1121 \\ 0.0936 & -0.7371 & 0.1322 & 0.2994 & 0.2119 \\ 0.1292 & 0.2567 & -0.7106 & 0.2119 & 0.1128 \\ 0.2110 & 0.4730 & 0.4107 & -1.2968 & 0.2021 \\ 0.2183 & 0.2894 & 0.2831 & 0.3892 & -1.1800 \end{bmatrix}$$

$$q^3_{(i,j,1)}=\begin{bmatrix} -1.2086 & 0.1168 & 0.6178 & 0.2118 & 0.2622 \\ 0.0779 & -1.0141 & 0.2118 & 0.5627 & 0.1618 \\ 0.4218 & 0.6426 & -1.4448 & 0.1177 & 0.2626 \\ 0.1178 & 0.2412 & 0.1555 & -0.5857 & 0.0712 \\ 0.3212 & 0.1608 & 0.1573 & 0.2163 & -0.8556 \end{bmatrix}$$

$$q^3_{(i,j,2)}=\begin{bmatrix} -1.2086 & 0.1168 & 0.6178 & 0.2118 & 0.2622 \\ 0.0779 & -1.0141 & 0.2118 & 0.5627 & 0.1618 \\ 0.4218 & 0.6426 & -1.4448 & 0.1177 & 0.2626 \\ 0.1178 & 0.2412 & 0.1555 & -0.5857 & 0.0712 \\ 0.3212 & 0.1608 & 0.1573 & 0.2163 & -0.8556 \end{bmatrix}$$

Given δ and γ and applying the extraproximal method (8.10 - 8.11) to calculate the strategies for the Nash equilibrium, we obtain the resulting utilities at the disagreement point for each player $\phi^l(c^1, c^2, c^3)$ as follows

$$\phi^1(c^1, c^2, c^3) = 127.0052, \quad \phi^2(c^1, c^2, c^3) = 110.9296, \quad \phi^3(c^1, c^2, c^3) = 129.5264$$

The utilities at the utopia point of the bargaining problem are as follows:

$$\psi^{1*}(c^1, c^2, c^3) = 140.7620,\ \psi^{2*}(c^1, c^2, c^3) = 111.9948,\ \psi^{3*}(c^1, c^2, c^3) = 168.6736$$

Then, given δ, γ, $\alpha^1 = 0.35$, $\alpha^1 = 0.2$, $\alpha^3 = 0.45$, and applying the extraproximal method (8.4 - 8.5) for the Kalai-Smorodinsky bargaining solution, we obtain the convergence of the strategies in terms of the variable $c^l_{(i,k)}$ for each player (see Figures 8.3, 8.4 and 8.5).

$$c^1 = \begin{bmatrix} 0.1778 & 0.1131 \\ 0.1438 & 0.0915 \\ 0.0807 & 0.0514 \\ 0.0697 & 0.0443 \\ 0.1392 & 0.0886 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.0603 & 0.0785 \\ 0.1289 & 0.1678 \\ 0.1104 & 0.1437 \\ 0.0815 & 0.1061 \\ 0.0534 & 0.0695 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.0702 & 0.0695 \\ 0.0917 & 0.1230 \\ 0.0779 & 0.0674 \\ 0.1349 & 0.2088 \\ 0.0584 & 0.0982 \end{bmatrix}$$

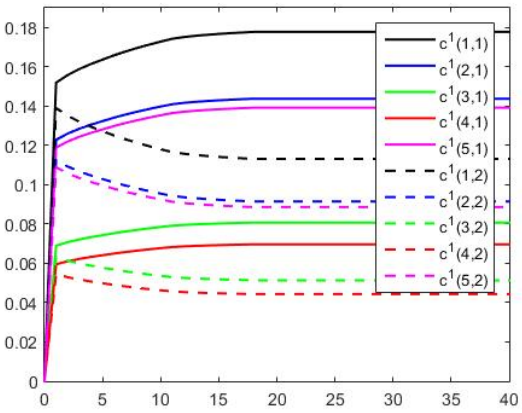


Figure 8.3 Strategies of player 1.

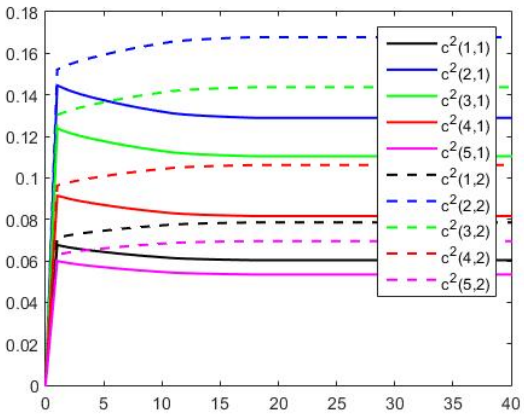


Figure 8.4 Strategies of player 2.

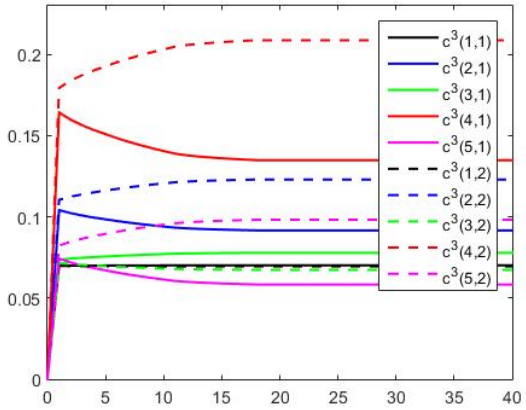


Figure 8.5 Strategies of player 3.

Following (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.6111 & 0.3889 \\ 0.6111 & 0.3889 \\ 0.6111 & 0.3889 \\ 0.6111 & 0.3889 \\ 0.6111 & 0.3889 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.4344 & 0.5656 \\ 0.4344 & 0.5656 \\ 0.4344 & 0.5656 \\ 0.4344 & 0.5656 \\ 0.4344 & 0.5656 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.5024 & 0.4976 \\ 0.4270 & 0.5730 \\ 0.5363 & 0.4637 \\ 0.3925 & 0.6075 \\ 0.3727 & 0.6273 \end{bmatrix}$$

With the strategies calculated, the resulting utilities at the Kalai-Smorodinky bargaining solution, are as follows:

$$\psi^1(c^1, c^2, c^3) = 130.0756 \quad \psi^2(c^1, c^2, c^3) = 111.0906 \quad \psi^3(c^1, c^2, c^3) = 137.4903$$

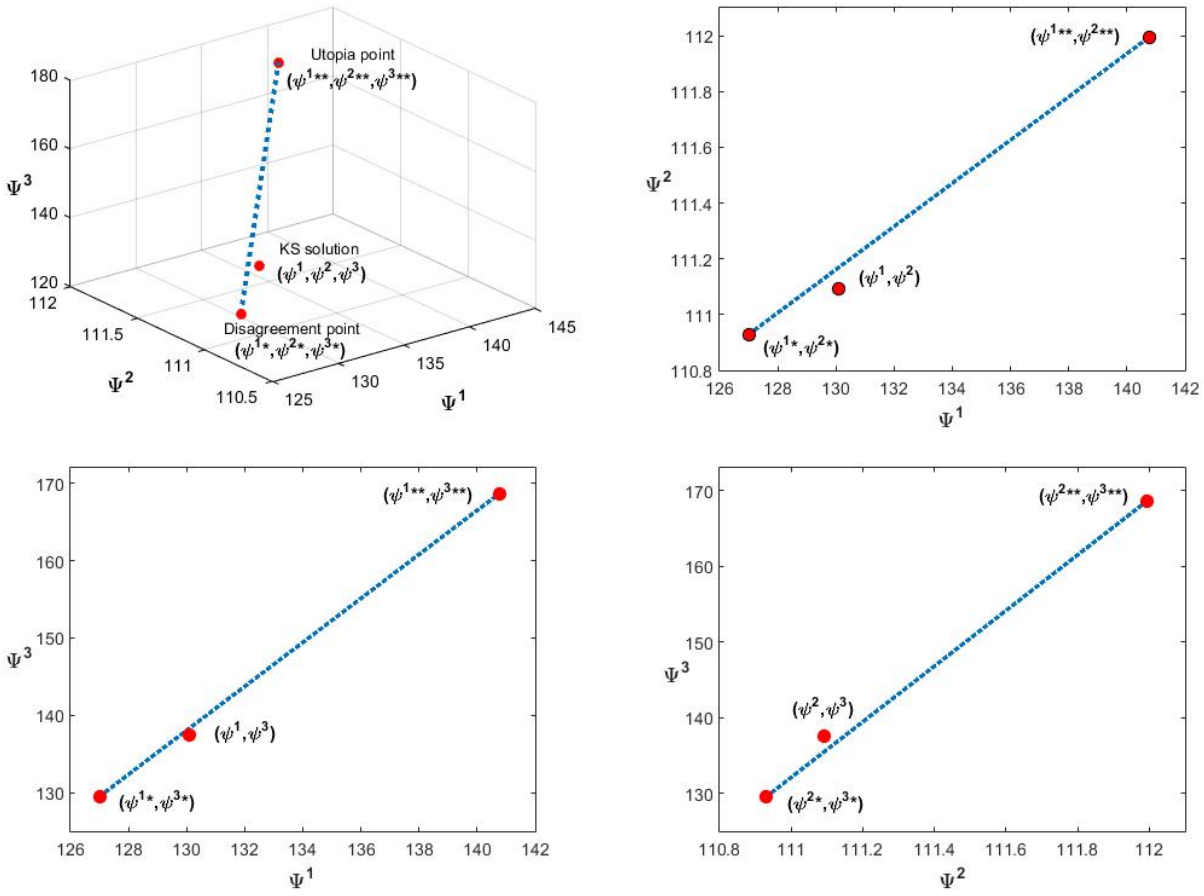


Figure 8.6 The Kalai-Smorodinsky solution.

Finally, we can see that the profits obtained with the Kalai-Smorodinsky solution are greater than those obtained at the disagreement point. Figure 8.6 shows the straight line linking the

utilities obtained at the disagreement point and those obtained at the utopia point. We can also observe that the Kalai-Smorodinsky solution approaches this line but is not exactly on it due to the convergence of the strategies. The significant input that can be illustrated by this example is the choice of the Kalai-Smorodinsky approach for the bargaining solution for the labor market problem. We expected that the parties progressively moderate their demands in equal proportional concessions.

Chapter 9

The non-cooperative bargaining game

9.1 Introduction

There has been a large and growing literature in non-cooperative bargaining. Rubinstein [83] presented a bilateral non-cooperative bargaining process as an alternating offers game with a bargaining cost for each period. Such a model has been studied and extended for three or more players in a variety of papers and situations. The non-cooperative bargaining model and its game-theoretic solution have also been applied in many important contexts like market games, networks, apex games, union formation, and water management.

Despite its wide applicability, crucial assumptions of the traditional bargaining model include that players have complete information about the characteristics of other agents (e.g., their discount factor or their utility) and that players are sophisticated in their behavior (e.g., they are forward-looking). The traditional equilibrium concept has been shown to fail when agents are not sophisticated, for instance when they are not forward-looking ([69, 56, 67, 68, 88, 39]). As such, there is a need to develop a general theory of bargaining that is robust to work in the absence of sophisticated players or incomplete information about other players.

As an aid to the implementation of bargaining solutions in the presence of unsophisticated agents, we propose an alternative approach to the traditional bargaining literature, where a planner has the ability to set up a game to aid the agents to reach an equilibrium. Thus, this chapter presents a novel approach that complements the traditional bargaining literature and

enlarges the class of processes and functions where non-cooperative bargaining solutions might be defined and applied.

To understand the characteristics of our game, consider the simple bargaining game where the planner is able to penalize the agents based on two factors: first players are penalized for their deviation from the previous best response strategy and second, they are penalized over the time taken for the decision-making at each step of the game.

This chapter presents a solution method of the non-cooperative bargaining problem for three different games:

- We solve the game where players are individual-rational, and make offers and counteroffers alternately thinking only of their own interests, i.e., they compute the strategies that maximize only their own utility.
- We present a solution for a game where at each step of the negotiation process players calculate the Nash equilibrium at the same time considering the utility function of all players, but with the particularity that internally each player reaches this equilibrium point in a different time.
- We analyze a game where players make coalitions and alternately each group of players makes an offer to the others until they reach an equilibrium point.

Finally, we illustrate the results of the three methods by a numerical example with continuous-time Markov chains.

9.2 The Rubinstein's alternating-offers model

In the simplest case, Rubinstein [83] considered a bargaining situation where two players ($\mathfrak{n} = 2$) have to reach an agreement on the partition of a pie of size 1; each player has to make in turn an offer (a proposal as to how it should be divided, i.e., an offer is the share of the pie to the proposer and the complete pie minus the offer is the share to the responder). After player 1 has made such an offer, player 2 must decided either to accept it, in this case the bargaining game ends and the players divide the cake according to the accepted offer, or to

reject it and continue with the bargaining process. If player 2 rejected, then this player has to make a counteroffer which player 1 would accept or reject it and continue with the negotiation process. The bargaining game continues until an offer is accepted. Offers are made at discrete points in time: at times $0, \Delta, 2\Delta, ..., t\Delta, ...$, where $\Delta > 0$.

In real situations it is important to consider that there are losses during the time that the players are negotiating, for example, the devaluation or deterioration of assets (in the example presented above, over time the cake can be spoiled). In order to deal with this problem, Rubinstein considered that there exists a cost associated with the time taken by the player $l = \overline{1, \mathfrak{n}}$ to reach an agreement x^l (a share $0 \leq x^l \leq 1$ of the cake) and proposed two class of models: a fixed bargaining cost where players have a fix cost for each period of time, therefore the agreement would produce payoffs of $x^l - c^l t\Delta$ for each player; and a fixed discounting factor for every player given by $x^l \cdot e^{(-r^l t\Delta)}$ that depends of a discount rate r^l (in this example, r^l can be interpreted as the rate at which the cake shrinks) and the function $\beta^l = e^{(-r^l \Delta)}$ is the discount factor of each player. In this way, it is clear that if the players rejects any offer made, then each player's payoff is zero.

Rubinstein showed that there exists a subgame perfect equilibrium in the bargaining problem: in the fixed bargaining cost model, if $c^1 > c^2$ player 1 receives c^2, if $c^1 < c^2$ player 1 receives all and if $c^1 = c^2$ player 1 receives at least c^1; in the fixed discounting factor model (the most used model in a bargaining process) the following offers are a subgame perfect equilibrium:

$$x^{1*} = \frac{1 - \beta^2}{1 - \beta^1 \beta^2} \qquad x^{2*} = \frac{1 - \beta^1}{1 - \beta^1 \beta^2}$$

where player 1 always offers x^{1*} and always accepts an offer x^2 if and only if $x^2 \leq x^{2*}$; and player 2 always offers x^{2*} and always accepts an offer x^1 if and only if $x^1 \leq x^{1*}$.

The alternating offers game with a discount rate $r^l > 0$ has a unique subgame perfect equilibrium, agreement is reached at time 0 and the equilibrium is Pareto efficient, if player 1 makes the offer at time 0, the shares of the cake obtained by players 1 and 2 in the unique subgame perfect equilibrium are x^{1*} and $1 - x^{1*}$ respectively. On the other hand, when $r^l = 0$ there exist many subgame perfect equilibria, including equilibria which are Pareto inefficient, in this case we have a frictionless bargaining game where players do not care how long it takes

to reach an agreement. Then, for any $r^l \geq 0$, the pair of strategies x^l in the bargaining game is a Nash equilibrium; even if $r^l > 0$, the alternating offers game has many Nash equilibria (see [60]).

Now, let us define the model in a general way. Let X be the set of possible agreements. Consider two players ($l = \overline{1, \mathfrak{n}}\,, \mathfrak{n} = 2$) bargaining according to the alternating-offers procedure in which an offer is an element of the set X. If players reach agreement at time $t\Delta$ on $x \in X$, then player l payoff considering the fixed discounting model is $\psi^l(x)e^{(-r^l t\Delta)}$, where $\beta^l = e^{(-r^l\Delta)}$ is the discount factor with a discount rate r^l, and $\psi^l(x) : X \to \mathbb{R}$ is the utility function from agreement x of each player.

Let define the set of possible utility pairs as $\Phi = \{(\psi^1, \psi^2)\}$. Thus, the set of possible utility pairs obtainable through agreement at time $t\Delta$ is

$$\Phi^t = \left\{(\psi^1\beta_1^t, \psi^2\beta_2^t) : (\psi^1, \psi^2) \in \Phi\right\}$$

It should be noted that $\Phi^0 = \Phi$ and let Φ^e denote the Pareto frontier of the set Φ. A utility pair $(\psi^1, \psi^2) \in \Phi^e$ if and only if $(\psi^1, \psi^2) \in \Phi$ and there does not exist another utility pair $(\varphi^1, \varphi^2) \in \Phi$ such that $\varphi^1 \geq \psi^1$, $\varphi^2 \geq \psi^2$. The Pareto frontier Φ^e of the set Φ is the graph function of a strictly decreasing and concave function, denoted by ϕ, whose domain is an interval $I^1 \subseteq \Re$ and range an interval $I^2 \subseteq \mathbb{R}$, with $0 \in I^1$, $0 \in I^2$ and $\phi(0) > 0$. Then,

$$\Phi^e = \left\{(\psi^1, \psi^2) : \psi^1 \in I^1, \psi^2 \in \phi(\psi^1)\right\}$$

Consider ϕ^{-1} the inverse of ϕ, a strictly decreasing and concave function from I^2 to I^1, with $\phi^{-1}(0) > 0$. Then, for any $\psi^1 \in I^1$, $\phi(\psi^1)$ is the maximum utility that player 2 receives subject to player 1 receiving a utility ψ^1; in the same way, for any $\psi^2 \in I^2$, $\phi^{-1}(\psi^2)$ is the maximum utility that player 1 receives subject to player 2 receiving a utility ψ^2.

Let x^{l*} be the equilibrium offer that player l makes during the bargaining process. Also, consider Z^l, a non-empty subset de X, defined as follows

$$Z^l = \left\{x^l := \arg\max_{x \in X} \psi^l(x^l) : \psi^m(x^l) = \beta^m\psi^m(x^m)\,, (m \neq l)\right\}$$

Proposition 9.1 *For any $x^{l*} \in Z^l$, $l = 1, 2$, the following pair of strategies is a subgame perfect equilibrium of the general Rubinstein model (see [60]):*

- *Player 1 always offers x^{1*} and always accepts an offer x^2 if and only if $\psi^1(x^2) \geq \beta^1\psi^{1*}$*
- *Player 2 always offers x^{2*} and always accepts an offer x^1 if and only if $\psi^2(x^1) \geq \beta^2\psi^{2*}$*

where

$$\psi^{1*} = \phi^{-1}(\beta^2\psi^{2*}) \qquad \psi^{2*} = \phi(\beta^1\psi^{1*})$$

If Z^l contains more than one element, then there exist more than one subgame perfect equilibrium in the general Rubinstein model. In any subgame perfect equilibrium, if agreement is reached at time 0 and it is player 1 who makes the offer, then the equilibrium payoff for player 1 is ψ^{1*} and for player 2 is $\phi(\psi^{1*})$; similarly, if it is player 2 who makes the offer at time 0, then the equilibrium payoff for player 1 is $\phi^{-1}(\psi^{2*})$ and for player 2 is ψ^{2*}. This equilibrium pair is Pareto efficient (See Figure 9.1).

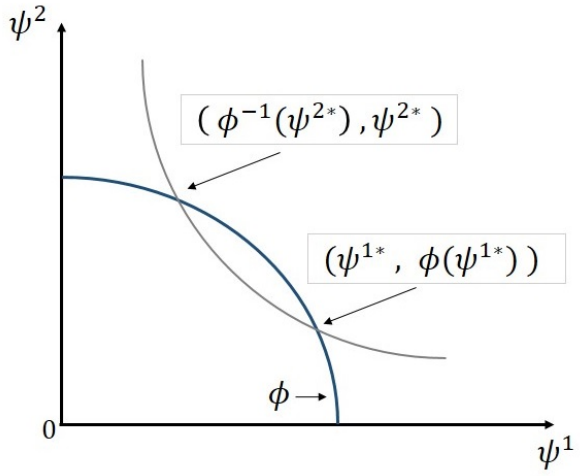


Figure 9.1 The Pareto solution of the bargaining problem at time 0.

Remark 9.2 *In the limit, as $\Delta \to 0$, the unique subgame perfect equilibrium payoff pair in the Rubinstein model converges to the asymmetric Nash bargaining solution of the appropriately defined bargaining problem; if and only if the players' discount factor are identical (i.e., $r^1 = r^2$) the symmetry axiom of the Nash bargaining solution would be satisfied.*

9.3 The non-cooperative bargaining game

Consider the game theory problem of a concave twice-differentiable real-valued function ψ defined on X, which is a compact and convex subset of $\mathbb{R}^{\mathbb{N}}$

$$\max_{x \in X} \psi(x)$$

Following the proximal point algorithm for solving game theory problems presented by Antipin [4], the unique solution is a sequence (x_n) $n \in \mathbb{N}$ with a initial value $x_0 \in X$,

$$\max_{x \in X} \left[\psi(x) - \delta_n \left\| x - x_n \right\|^2 \right] \tag{9.1}$$

where $\delta_n > 0$, $\delta_n \downarrow 0$ and the term $\|x - x_n\|^2$ ensures that the objective function (9.1) is strictly positive definite and that some iterative method presents convergence [93, 94]. The result obtained is not affected by the quadratic term for $\delta_n > 0$ and $\delta_n \downarrow 0$.

The bargaining game model considered in this chapter involves game theory problems with an additional penalization, a time cost related with the time spent for each player to move from one position to another one [6, 59, 9], i.e., to decide either to accept an offer or to reject it and choose another.

In this section we will discuss three different ways to formulate the non-cooperative bargaining game with alternating-offers and time cost. In the models presented below, it is considered that the players start from a point that is Pareto optimal, players could obtain the best utilities if they finished the bargaining process at time 0.

Bargaining model 1

In this first approach, we consider the model presented by Rubinstein [83], and we provide a solution to a bargaining situation where players are individual-rational and alternately make offers and counteroffers thinking only of their own interests, i.e., they compute independently the strategies that maximize only their own utility.

In general terms, the dynamic of the multilateral non-cooperative bargaining game is as follows. The game consists of a set $\mathcal{N} = \{1, ..., \mathfrak{n}\}$ of players bargaining a certain transaction

according to the alternating-offers procedure. Define the behavior of each player $l = \overline{1, \mathfrak{n}}$ as a sequence $x_n^l \in X^l$, $n \in \mathbb{N}$, where X^l is the decision space (strategies) of each player. Then, we can define the strategies set of all players as $x_n = (x_n^1, ..., x_n^{\mathfrak{n}}) \in X$ where X is a convex and compact set. Players take turns to analyze and present their position in the negotiation process, i.e., at each step n the player l in turn must decide between to stay in the same strategy $x_{n+1} = x_n$, that is that player l accepts the offer, or to choose a new strategy $x_{n+1} \neq x_n$, that means that player rejects the offer and makes a new one. The function $\psi^l(x)$ represents the utility function of each player which determines the decision of to accept or to reject the offer.

At turn $n = 0$, the first player to make an offer chooses a strategy set x_n considering the utility function $\psi^l(x)$, then, the rest of the players must decided either to accept the offer and finish the game or to reject it and continue with the process, in this case, at step $n = 1$ the next player makes a counteroffer by choosing a strategy set x_n that benefits him more or in equal measure than the offer proposed by the first player according to his utility function, if this counteroffer is accepted then agreement is struck, otherwise, the player in turn makes a new offer at step $n = 2$, and the process continues.

The time cost between offers is defined for each player as a function $\Lambda^l : X \times X \to \mathbb{R}$ which can be interpreted as a distance function of each player where $\Lambda^l(x_n, x_{n+1}) = \kappa^l(x_n, x_{n+1})$, we have that $\kappa^l(x_n, x_{n+1}) = 0$ if $x_{n+1} = x_n$ (accepts the offer) or $\kappa^l(x_n, x_{n+1}) > 0$ if $x_n \neq x_{n+1}$ (rejects and makes a new one). In general, the time cost function can be reexpressed as $\Lambda^l(x_n, x_{n+1}) := t^l(x_n, x_{n+1})\kappa^l(x_n, x_{n+1})$ where $t^l(x_n, x_{n+1}) \geq 0$ is the time spent for each player to reject an offer x_n and to make a new one x_{n+1} and $\kappa^l(x_n, x_{n+1})$ is the offer cost function associated to each player.

In the simplest case, each player makes a new offer trying to obtain the highest possible payoff according to the utility function, $\psi^l(x_{n+1}) - \psi^l(x_n) \geq 0$ given the time spent $t^l(x_{n+1}) > 0$ to analyze the advantage of to reject the offer x_n and make a new offer x_{n+1}, and $\alpha^l(x_n)$ be the weight that players put on their advantages of to reject the offer x_n. Thus, the advantages of to reject the offer x_n and to propose a new offer x_{n+1} are given by $A^l(x_n, x_{n+1}) = \alpha^l(x_n)t^l(x_{n+1})\left[\psi^l(x_{n+1}) - \psi^l(x_n)\right]$.

The dynamics of the bargaining game with alternating-offers considering the time cost is as follows. At each step $n \in \mathbb{N}$, the player in turn considers to reject the offer x_n and propose a new offer x_{n+1}. For each player, to make a new proposal is acceptable if the advantages $A^l(x_n, x_{n+1})$ are determined by $\delta^l(x_n) \in [0,1]$ (degree of acceptability) of the time cost $\Lambda^l(x_n, x_{n+1})$. Then, the set of strategies that maximizes the utility of each player is defined by

$$F^l(x_n) =$$

$$\left\{x_{n+1} \in X : \alpha^l(x_n) t^l(x_{n+1}) \left[\psi^l(x_{n+1}) - \psi^l(x_n)\right] \geq \delta^l(x_n) t^l(x_{n+1}) \kappa^l(x_n, x_{n+1})\right\}$$

We define a utility function $\psi^l : X \to \mathbb{R}$ such that the impact of experience on cost is constant and limited to the most recent element x_n on the trajectory (x_n). In addition, the advantages to change $A^l(x_n, x_{n+1})$ are determined by the degree of acceptability $\delta_n^l(x_n) \in [0,1]$ of the costs to move $\Lambda^l(x_n, x_{n+1})$.

Thus, the acceptance criterion to propose a new offer satisfies the condition

$$\alpha_n^l(x_n) t^l(x_{n+1}) \psi^l(x_{n+1}) \geq \delta_n^l(x_n) t^l(x_{n+1}) \kappa^l(x_n, x_{n+1})$$

This algorithms are naturally linked with several classical proximal algorithms given in eq. (9.1). That is, by taking the functions $\delta_n^l t^l(x) \kappa^l(x^*, x) = \delta_n^l t^l(x) \left\|(x - x^*)\right\|^2$ and $A^l(x, x^*) := \alpha_n^l t^l(x) \left[\psi^l(x) - \psi^l(x^*)\right]$, the point x^* solves the maximization problem if remains a fixed point of the proximal mapping, that is,

$$x^* = \arg\max_{x \in X} \left\{-\delta_n^l t^l(x) \left\|(x - x^*)\right\|^2 + \alpha_n^l t^l(x) f(x, x^*)\right\} \tag{9.2}$$

where

$$f(x, x^*) := \psi^l(x) - \psi^l(x^*)$$

Once the player in turn makes a new offer according to equation (9.2), the next player must decide either to accept or to reject the offer. If the player rejects the offer, then now it is his turn to calculate the strategies that benefit his utility and to make a new offer. This process continues until an agreement is reached, i.e. the proposals (strategies) of the players do not change (convergence).

Bargaining model 2

In this approach we present a solution where at each step of the negotiation process players calculate the Nash equilibrium with the particularity that internally each player reaches this equilibrium point in a different time. Following the description of the model presented previously, we redefine the advantage of propose a new offer that depends on the utility function

$$f(x_n, x_{n+1}) := \sum_{l=1}^{\mathrm{n}} \left[\psi^l(x_{n+1}) - \psi^l(x_n)\right] \geq 0$$

for all players to reject the offer x_n and making a new offer x_{n+1} given the time spent to benefit of this advantage $t(x_{n+1}) > 0$, and $\alpha^l(x_n)$ be the weight that players put on their advantages to reject the offer x_n. Thus, the advantages to reject the offer x_n and to propose a new offer x_{n+1} are given by $A(x_n, x_{n+1}) = \alpha(x_n)t(x_{n+1})f(x_n, x_{n+1})$.

Remark 9.3 *The function $f(x_n, x_{n+1})$ satisfies the Nash condition*

$$\psi^l(x_{n+1}) - \psi^l(x_n) \geq 0$$

for any $x \in X$ and all players.

Definition 9.4 *A strategy $x^* \in X$ is said to be a Nash equilibrium if*

$$x^* \ \in Arg \max_{x \in X} \ \{f(x_n, x_{n+1})\}$$

Then, at each step of the bargaining game we have in proximal format that the players must select their strategies according to

$$x^* = \arg\max_{x \in X} \left\{-\delta_n t(x) \left\|(x - x^*)\right\|^2 + \alpha_n t(x) f(x, x^*)\right\} \tag{9.3}$$

where

$$f(x, x^*) := \sum_{l=1}^{\mathrm{n}} \left[\psi^l(x) - \psi^l(x^*)\right]$$

At each step of the bargaining process, players calculate simultaneously the Nash equilibrium but considering that each player reach the equilibrium in a different time.

Bargaining model 3

In this approach we analyze a bargaining situation where players make groups and alternately each group makes an offer to the others until they reach an equilibrium point (agreement). We describe a bargaining model with two teams of players as follows. Let us consider a bargaining game with $\mathcal{N}+\mathcal{M}$ players. Let $\mathcal{N}=\{1,...,\mathfrak{n}\}$ denote the set of players called team *A* and let us define the behavior of all players $l=\overline{1,\mathfrak{n}}$ as $x_n=(x_n^1,...,x_n^{\mathfrak{n}})\in X$ where X is a convex and compact set. In the same way, the rest $\mathcal{M}=\{1,...,\mathfrak{m}\}$ players are the team *B* and let the set of the strategy profiles of all player $m=\overline{1,\mathfrak{m}}$ be defined by $y_n=(y_n^1,...,y_n^{\mathfrak{m}})\in Y$ where Y is a convex and compact set. Then, $X\times Y$ in the set of full strategy profiles. In this model the function $\psi(x,y)$ represents the utility function of team *A* which determines the decision of accept or reject the offer; similarly, team *B* makes the decision according to its utility function $\varphi(x,y)$.

Following the description of the model presented above, we redefine the advantage of propose a new offer considering the utility function for team A as follows

$$f(x_n,y_n,x_{n+1},y_{n+1}):=\sum_{l=1}^{\mathfrak{n}}\left[\psi^l(x_{n+1},y_n)-\psi^l(x_n,y_n)\right]\geq 0$$

and, similarly the utility function for team B is as follows

$$g(x_n,y_n,x_{n+1},y_{n+1}):=\sum_{m=1}^{\mathfrak{m}}\left[\varphi^l(x_n,y_{n+1})-\varphi^l(x_n,y_n)\right]\geq 0$$

Thus, the advantages for team A to reject the offer x_n and to propose a new offer x_{n+1} are given by $A(x_n,y_n,x_{n+1},y_{n+1})=\alpha(x_n)t(x_{n+1})f(x_n,y_n,x_{n+1},y_{n+1})$; in the same way, the advantages for team B to reject the offer y_n and to propose a new offer y_{n+1} are given by $A(x_n,y_n,x_{n+1},y_{n+1})=\alpha(y_n)t(y_{n+1})g(x_n,y_n,x_{n+1},y_{n+1})$.

Remark 9.5 *The function $f(x_n,y_n,x_{n+1},y_{n+1})$ satisfies the Nash condition*

$$\psi^l(x_{n+1},y_n)-\psi^l(x_n,y_n)\geq 0$$

for any $x\in X$, $y\in Y$ and $l=\overline{1,\mathfrak{n}}$ players.

Remark 9.6 *The function $g(x_n, y_n, x_{n+1}, y_{n+1})$ satisfies the Nash condition*

$$\varphi^l(x_n, y_{n+1}) - \varphi^l(x_n, y_n) \geq 0$$

for any $x \in X$, $y \in Y$ and $m = \overline{1, \mathfrak{m}}$ players.

The dynamics of the bargaining game is as follows: at each step of the negotiation process the team *A* chooses a strategy $x \in X$ considering the utility function $f(x_n, y_n, x_{n+1}, y_{n+1})$, then team *B* must decide between to accept or reject the offer calculating a new offer (strategies) $y \in Y$ considering the utility function of the group $g(x_n, y_n, x_{n+1}, y_{n+1})$. Following the description of the model 1, now we have that teams solve the problem in proximal format as follows:

$$\begin{aligned} x^* &= \arg\max_{x \in X} \left\{ -\delta_n t(x) \left\| (x - x^*) \right\|^2 + \alpha_n t(x) f(x, y, x^*, y^*) \right\} \\ y^* &= \arg\max_{y \in Y} \left\{ -\delta_n t(y) \left\| (y - y^*) \right\|^2 + \alpha_n t(y) g(x, y, x^*, y^*) \right\} \end{aligned} \tag{9.4}$$

where

$$f(x, y, x^*, y^*) := \sum_{l=1}^{\mathfrak{n}} \left[\psi^l(x, y^*) - \psi^l(x^*, y^*) \right]$$

$$g(x, y, x^*, y^*) := \sum_{m=1}^{\mathfrak{m}} \left[\varphi^m(x^*, y) - \varphi^m(x^*, y^*) \right]$$

At each step, teams make a new offer according to equation (9.4), both teams solve the bargaining problem together but they reach the equilibrium at different time, the bargaining game continues until the offers (strategies) of all player show convergence.

9.3.1 Formulation of the problem

Consider the following constrained programming problem

$$\begin{aligned} &\max_{x \in X_{\text{adm}}} \ f(x, x_n) \\ &X_{\text{adm}} := \left\{ x \in \mathbb{R}^{\mathfrak{n}} : x \geq 0, \ A_0 x = b_0 \in \mathbb{R}^{M_0}, \ A_1 x \leq b_1 \in \mathbb{R}^{M_1} \right\} \end{aligned} \tag{9.5}$$

where X_{adm} is a bounded set. Introducing the vector $u \in \mathbb{R}^{M_1}$ with components $u_i \geq 0$ for all $i = 1, ..., M_1$, the original problem (9.5) can be rewritten as

$$\begin{aligned} &\max_{x \in X_{\text{adm}}, u \geq 0} \ f(x, x_n) \\ &X_{\text{adm}} := \{ x \in \mathbb{R}^{\mathfrak{n}} : x \geq 0, \ A_0 x = b_0, \ A_1 x - b_1 + u = 0 \} \end{aligned} \tag{9.6}$$

Notice that this problem may have non-unique solution and $\det\left(A_0^{\intercal}A_0\right)=0$. Define by $X^*\subseteq X_{\text{adm}}$ the set of all solutions of the problem (9.6) and consider the objective function

$$\begin{aligned}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right):=-\frac{\delta}{2}t(x)\left\|x-x_n\right\|^2+\alpha t(x)f(x,x_n)-\\ \frac{1}{2}\left\|A_0x-b_0\right\|^2-\frac{1}{2}\left\|A_1x-b_1+u\right\|^2-\frac{\delta}{2}\left\|u\right\|^2\end{aligned}\tag{9.7}$$

where the parameters α, δ are positive. Then, the game theory problem is as follows

$$\max_{x\in X_{\text{adm}},u\geq 0}\ \mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)$$

9.3.2 Convergence analysis

The game consists of a set $\mathcal{N}=\{1,...,\mathfrak{n}\}$ of players. Let $x^l\in X^l$ be the strategy of each player $l=\overline{1,\mathfrak{n}}$ where X^l is the decision space (strategies) of each player. Then, we can define the strategies set of all players as

$$x=(x^1,...,x^{\mathfrak{n}})\in X,\qquad X:=\bigotimes_{l=1}^{\mathfrak{n}}X^l$$

where X is a convex and compact set.

Theorem 9.7 *The bounded set X^* of all solutions of the original game theory problem (9.6) is not empty and the Slater's condition holds, that is, there exists a point $\mathring{x}\in X_{adm}$ such that*

$$A_1\mathring{x}<b_1\tag{9.8}$$

Moreover, the parameters α and δ are time-varying, i.e.,

$$\alpha=\alpha_n,\quad \delta=\delta_n\quad (n=0,1,2,....)$$

such that

$$0<\alpha_n\downarrow 0,\quad \frac{\alpha_n}{\delta_n}\downarrow 0\quad \textit{when } n\to\infty\tag{9.9}$$

Then

$$x_n^*:=x^*\left(\alpha_n,\delta_n\right)\underset{n\to\infty}{\to}x^{**}$$

$$u_n^*:=u^*\left(\alpha_n,\delta_n\right)\underset{n\to\infty}{\to}u^{**}$$

where $x^{**} \in X^*$ *and* $u^{**} \in \mathbb{R}^{M_1}$ *define the solution of the original problem (9.6) with the minimal weighted norm,*

$$\|x^{**}\|^2 + \|u^{**}\|^2 \leq \|x^*\|^2 + \|u^*\|^2$$

for all $x^* \in X^*$, $u^* \in \mathbb{R}^{M_1}$ *and*

$$u^{**} = b_1 - A_1 x^{**}$$

Proof.

1. *First, let us prove that the Hessian matrix* $\mathbb{H}$ *associated with the objective function (9.7) is strictly negative definite for any positive* α *and* δ*, to show that the objective function (9.7) is strictly concave. If the set of solutions of problem (9.6) is non-empty then the objective function (9.7) is strictly concave.*

It should be proven that for all $x \in \mathbb{R}^{\mathfrak{n}}$ and $u \in \mathbb{R}^{M_1}$

$$\mathbb{H} = \begin{bmatrix} \dfrac{\partial^2}{\partial x^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) & \dfrac{\partial^2}{\partial u \partial x}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) \\ \dfrac{\partial^2}{\partial x \partial u}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) & \dfrac{\partial^2}{\partial u^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) \end{bmatrix} < 0,$$

Employing Schur's lemma [74] it is necessary and sufficient to prove that

$$1.\quad \frac{\partial^2}{\partial x^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) < 0, \qquad 2.\quad \frac{\partial^2}{\partial u^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) < 0,$$

$$3.\quad \frac{\partial^2}{\partial x^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) < \frac{\partial^2}{\partial u \partial x}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)\left[\frac{\partial^2}{\partial u^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)\right]^{-1}\frac{\partial^2}{\partial x \partial u}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right).$$

Applying the Schur's lemma over the objective function (9.7) it follows for condition 1

$$\frac{\partial^2}{\partial x^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) = -\delta t(x_n) I_{\mathfrak{n}\times\mathfrak{n}} + \alpha t(x_n)\frac{\partial^2}{\partial x^2} f\left(x,x_n\right) - A_0^{\intercal}A_0 - A_1^{\intercal}A_1 \leq$$

$$\alpha t(x_n)\frac{\partial^2}{\partial x^2} f\left(x,x_n\right) - \delta t(x_n) I_{\mathfrak{n}\times\mathfrak{n}} \leq \delta t(x_n)\left(\frac{\alpha}{\delta}\lambda^{+} - 1\right) I_{\mathfrak{n}\times\mathfrak{n}} < 0,$$

for all $\delta > 0$ where

$$\lambda^{+} := \max_{x \in X_{\text{adm}}}\left[\lambda_{\max}\left(\frac{\partial^2}{\partial x^2} f(x,x_n)\right)\right] < 0,$$

Then, for condition 2 we have

$$\frac{\partial^2}{\partial u^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right) = -\left(1+\delta\right) I_{M_1\times M_1} < 0.$$

By condition 3, it is necessary to satisfy that

$$\frac{\partial^2}{\partial x^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)=-\delta t(x_n)I_{\mathfrak{n}\times\mathfrak{n}}+\alpha t(x_n)\frac{\partial^2}{\partial x^2}f\left(x,x_n\right)-A_0^{\intercal}A_0-A_1^{\intercal}A_1<$$
$$\frac{\partial^2}{\partial u\partial x}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)\left[\frac{\partial^2}{\partial u^2}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)\right]^{-1}\frac{\partial^2}{\partial x\partial u}\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)=-\left(1+\delta\right)^{-1}A_1^{\intercal}A_1,$$

or equivalently,

$$\alpha t(x_n)\frac{\partial^2}{\partial x^2}f\left(x,x_n\right)-\delta t(x_n)I_{\mathfrak{n}\times\mathfrak{n}}-A_0^{\intercal}A_0-\frac{\delta}{1+\delta}A_1^{\intercal}A_1<0,$$

which holds for any $\delta>0$ having

$$t(x_n)\left(\alpha\lambda^{+}-\delta\right)I_{\mathfrak{n}\times\mathfrak{n}}-A_0^{\intercal}A_0-\frac{\delta}{1+\delta}A_1^{\intercal}A_1\leq$$
$$\delta t(x_n)\left(\frac{\alpha}{\delta}\lambda^{+}-1\right)I_{\mathfrak{n}\times\mathfrak{n}}=\delta t(x_n)\left(o(1)-1\right)I_{\mathfrak{n}\times\mathfrak{n}}<0.$$

As a result, the Hessian is $\mathbb{H}<0$ which means that proximal function (9.7) is strictly concave and, hence, has a unique maximal point defined below as $x^*\left(\alpha,\delta\right)$ and $u^*\left(\alpha,\delta\right)$.

2. *If the proximal function (9.7) is strictly concave then the sequence $\{x_n\}$ of the proximal function (9.7) converges when $n\to\infty$, i.e. the proximal function has a maximal point defined by $x^*\left(\alpha,\delta\right)$ and $u^*\left(\alpha,\delta\right)$.*

Following the strictly concavity property (Theorem 9.7) for any $y:=\begin{pmatrix}x\\u\end{pmatrix}$ and any vector $y_n^*:=\begin{pmatrix}x_n^*=x^*\left(\alpha_n,\delta_n\right)\\u_n^*=u^*\left(\alpha_n,\delta_n\right)\end{pmatrix}$ for the function $\mathbb{P}_{\alpha,\delta}\left(x,u|x_n\right)=\mathbb{P}_{\alpha,\delta}\left(y|x_n\right)$ we have

$$
\begin{aligned}
0 &\leq (y_n^* - y)^\intercal \frac{\partial}{\partial y}\mathbb{P}_{\alpha_n,\delta_n}\left(y_n^*|x_n\right) \\
&= (x_n^* - x)^\intercal \frac{\partial}{\partial x}\mathbb{P}_{\alpha_n,\delta_n}\left(x_n^*, u_n^*|x_n\right) + (u_n^* - u)^\intercal \frac{\partial}{\partial u}\mathbb{P}_{\alpha_n,\delta_n}\left(x_n^*, u_n^*|x_n\right) \\
&= (x_n^* - x)^\intercal \Bigg(-\delta_n t(x_n)(x_n^* - x_n) + \alpha_n t(x_n)\frac{\partial}{\partial x} f\left(x_n^*, x_n\right) - A_0^\intercal\left[A_0 x_n^* - b_0\right] \\
&\quad -A_1^\intercal\left[A_1 x_n^* - b_1 + u_n^*\right]) + (u_n^* - u)^\intercal\left(-A_1 x_n^* + b_1 - (1+\delta_n)\, u_n^*\right) \\
&= \alpha_n t(x_n)\left(x_n^* - x\right)^\intercal \frac{\partial}{\partial x} f\left(x_n^*, x_n\right) - \left[A_0\left(x_n^* - x\right)\right]^\intercal\left[A_0 x_n^* - b_0\right] \\
&\quad - \left[A_1\left(x_n^* - x\right)\right]^\intercal\left[A_1 x_n^* - b_1 + u_n^*\right] - \delta_n t(x_n)\left(x_n^* - x\right)^\intercal\left(x_n^* - x_n\right) \\
&\quad - \left(u_n^* - u\right)^\intercal\left[A_1 x_n^* - b_1 + (1+\delta_n)\, u_n^*\right].
\end{aligned}
\tag{9.10}
$$

Now, selecting $x := x^* \in X^*$ (x^* is one of the admissible solutions such that $A_0 x^* = b_0$ and $A_1 x^* = b_1 - u^*$) and $u := (1+\delta_n)^{-1}\left(b_1 - A_1 x_n^*\right)$ we obtain

$$
\begin{aligned}
0 \leq \alpha_n t(x_n)\left(x_n^* - x^*\right)^\intercal \frac{\partial}{\partial x} f\left(x_n^*, x_n\right) - \left[A_0\left(x_n^* - x^*\right)\right]^\intercal\left[A_0 x_n^* - b_0\right] - \\
\left[A_1\left(x_n^* - x^*\right)\right]^\intercal\left[A_1 x_n^* - b_1 + u_n^*\right] - \delta_n t(x_n)\left(x_n^* - x^*\right)^\intercal\left(x_n^* - x_n\right) - \\
(1+\delta_n)^{-1}\left[u_n^*(1+\delta_n) - b_1 + A_1 x_n^*\right]^\intercal\left[A_1 x_n^* - b_1 + (1+\delta_n)\, u_n^*\right] - \\
\delta_n\left(u_n^* - b_1 - A_1 x_n^*\right)^\intercal u_n^*,
\end{aligned}
\tag{9.11}
$$

simplifying eq. (9.11) we have

$$
\begin{aligned}
0 \leq \alpha_n t(x_n)\left(x_n^* - x^*\right)^\intercal \frac{\partial}{\partial x} f\left(x_n^*, x_n\right) - \left\|A_0\left(x_n^* - x^*\right)\right\|^2 - \delta_n t(x_n)\left(x_n^* - x^*\right)^\intercal\left(x_n^* - x_n\right) - \\
\left\|A_1\left(x_n^* - x^*\right)\right\|^2 - (1+\delta_n)^{-1}\left\|A_1 x_n^* - b_1 + u_n^*(1+\delta_n)\right\|^2 - \delta_n\left(u_n^* - b_1 - A_1 x_n^*\right)^\intercal u_n^*.
\end{aligned}
$$

Dividing both sides of this inequality by δ_n we obtain

$$
\begin{aligned}
0 \leq \frac{\alpha_n}{\delta_n} t(x_n)\left(x_n^* - x^*\right)^\intercal \frac{\partial}{\partial x} f\left(x_n^*, x_n\right) - \frac{1}{\delta_n}\left\|A_0\left(x_n^* - x^*\right)\right\|^2 - \\
\frac{1}{\delta_n}\left\|A_1\left(x_n^* - x^*\right)\right\|^2 - \frac{1}{\delta_n}(1+\delta_n)^{-1}\left\|A_1 x_n^* - b_1 + u_n^*(1+\delta_n)\right\|^2 - \\
t(x_n)\left(x_n^* - x^*\right)^\intercal\left(x_n^* - x_n\right) - \left(u_n^x - b_1 - A_1 x_n^*\right)^\intercal u_n^*.
\end{aligned}
\tag{9.12}
$$

Now, taking $x = x_n^*$ and $u = 0$ from Eq. (9.10) one has

$$\begin{aligned}
0 \leq & -\left(u_n^*\right)^{\intercal}\left[A_1 x_n^* - b_1 + \left(1+\delta_n\right)u_n^*\right] \\
& = -\left(u_n^*\right)^{\intercal}\left(A_1 x_n^* - b_1\right) - \left(1+\delta_n\right)\left\|u_n^*\right\|^2 \\
& = -\left(\left\|\sqrt{1+\delta_n}u_n^*\right\|^2 + 2\left(\sqrt{1+\delta_n}u_n^*\right)^{\intercal}\left[\frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right] + \left\|\frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right\|^2\right. \\
& \left. - \left\|\frac{\left(A_1 v_n^* - b_1\right)}{2\sqrt{1\delta_n}}\right\|^2\right) \\
& = -\left\|\sqrt{1+\delta_n}u_n^* + \frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right\|^2 + \left\|\frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right\|^2,
\end{aligned}$$

implying

$$\left\|\frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right\|^2 \geq \left\|\sqrt{1+\delta_n}u_n^* + \frac{\left(A_1 x_n^* - b_1\right)}{2\sqrt{1+\delta_n}}\right\|^2,$$

and

$$1 \geq \left\|e + 2\left(1+\delta_n\right)u_n^*\left\|\left(A_1 x_n^* - b_1\right)\right\|^{-1}\right\|^2,$$

where $\|e\| = 1$. Which means that the sequence $\{u_n^*\}$ is bounded. In view of this and taking into account that by the supposition that $\dfrac{\alpha_n}{\delta_n} \underset{n\to\infty}{\to} 0$, from Eq. (9.12) it follows

$$\begin{gathered}
\text{Const} = \limsup_{n\to\infty}\left(\left|\left(x_n^* - x^*\right)^{\intercal}\left(x_n^* - x_n\right)\right| + \left|\left(u_n^* - b_1 - A_1 x_n^*\right)^{\intercal}u_n^*\right|\right) \geq \\
\limsup_{n\to\infty}\frac{1}{\delta_n}\left(\left\|A_0\left(x_n^* - x^*\right)\right\|^2 + \left\|A_1\left(x_n^* - x^*\right)\right\|^2 + \left(1+\delta_n\right)^{-1}\left\|A_1 x_n^* - b_1 + \left(1+\delta_n\right)u_n^*\right\|^2\right).
\end{gathered} \tag{9.13}$$

From Eq. (9.13) we may conclude that

$$\begin{gathered}
\left\|A_0\left(x_n^* - x^*\right)\right\|^2 + \left\|A_1\left(x_n^* - x^*\right)\right\|^2 + \\
\left(1+\delta_n\right)^{-1}\left\|A_1 x_n^* - b_1 + \left(1+\delta_n\right)u_n^*\right\|^2 = O\left(\delta_n\right),
\end{gathered} \tag{9.14}$$

and

$$A_0 x_\infty^* - A_0 x^* = A_0 x_\infty^* - b_0 = 0,$$

$$A_1 v_\infty^* - A_1 x^* = A_1 x_\infty^* - b_1 + u_\infty^* = 0,$$

where $x_\infty^* \in X^*$ is a partial limit of the sequence $\{x_n^*\}$ which, obviously, may be not unique. The vector u_∞^* is also a partial limit of the sequence $\{u_n^*\}$.

3. *Now, denote by* $\hat{x}_n$ *the projection of* x_n^* *to the set* X_{adm}*, namely,*

$$\hat{x}_n = \underset{X_{\text{adm}}}{\Pr}\left(x_n^*\right),$$

where $\Pr$ *is the projection operator. And show that*

$$\|x_n^* - \hat{x}_n\| \leq C\sqrt{\delta_n}, \quad C = \text{const} > 0. \tag{9.15}$$

From Eq. (9.14) we have that

$$\|A_1 x_n^* - b_1 + u_n^*\| \leq C_1\sqrt{\delta_n}, \quad C_1 = \text{const} > 0,$$

implying

$$A_1 x_n^* - b_1 \leq C_1\sqrt{\delta_n}e - u_n^* \leq C_1\sqrt{\delta_n}e, \quad \|e\| = 1,$$

where the vector inequality is treated in component-wise sense. Since:

$$\|x_n^* - \hat{x}_n\|^2 \leq \max_{y\in X_{\text{adm}}} \min_{A_1 x - b_1 \leq C_1\sqrt{\delta_n}e,\, x\in X_{\text{adm}}} \|x - y\|^2 := d\left(\delta_n\right).$$

Introduce the new variable

$$\tilde{x} := \left(1 - v_n\right)x + v_n\mathring{x} \quad \in X_{\text{adm}},$$

where by Slater's condition given in Eq. (9.8)

$$0 < v_n := \frac{C_1\sqrt{\delta_n}}{C_1\sqrt{\delta_n} + \max\limits_{j=1,\ldots,M_1}\left|\left(A_1\mathring{x} - b_1\right)_j\right|} < 1.$$

For the new variable $x = \dfrac{\tilde{x} - v_n\mathring{x}}{1 - v_n}$ we have

$$\begin{aligned}
A_1\tilde{x} - b_1 &= \left(1 - v_n\right)A_1 x + v_n A_1\mathring{x} - b_1 \\
&= \left(1 - v_n\right)\left(A_1 x - b_1\right) + \left(1 - v_n\right)b_1 + v_n\left(A_1\mathring{x} - b_1\right) + v_n b_1 - b_1 \\
&= \left(1 - v_n\right)\left(A_1 x - b_1\right) + v_n\left(A_1\mathring{x} - b_1\right) \\
&\leq \left(1 - v_n\right)C_1\sqrt{\delta_n}\, e + v_n\left(A_1\mathring{x} - b_1\right) \\
&= \frac{C_1\sqrt{\delta_n}}{C_1\sqrt{\delta_n} + \max\limits_{j=1,\ldots,M_1}\left|\left(A_1\mathring{x} - b_1\right)_j\right|}\left(\max_{j=1,\ldots,M_1}\left|\left(A_1\mathring{x} - b_1\right)_j\right| e + \left(A_1\mathring{x} - b_1\right)\right) \leq 0,
\end{aligned}$$

and therefore

$$\begin{aligned} d(\delta_n) &= \max_{y\in X_{\text{adm}}} \min_{A_1x-b_1\le C_1\sqrt{\delta_n}e,\, x\in X_{\text{adm}}} \|x-y\|^2 \\ &\le \max_{A_1\tilde{x}-b_1\le 0,\, \tilde{x}\in X_{\text{adm}}} \left\| \frac{\tilde{x}-v_n\mathring{x}}{1-v_n} - \tilde{x} \right\|^2 \\ &= \frac{v_n^2}{(1-v_n)^2} \min_{A_1\tilde{x}-b_1\le 0,\, \tilde{x}\in X_{\text{adm}}} \|\tilde{x}-\mathring{x}\|^2 \\ &\le C_2\delta_n, \quad C_2>0. \end{aligned}$$

Given that $\|x_n^*-\hat{x}_n\| \le \sqrt{d(\delta_n)} \le \sqrt{C_2}\sqrt{\delta_n}$ which proves Eq. (9.15).

4. *If the proximal function (9.7) is strictly concave and the sequence $\{x_n\}$ of the proximal function (9.7) converges, then, the necessary and sufficient condition for the point x^* to be the maximum point of the function $\|x_\infty^*\|^2$ on the set X^* is given by*

$$0 \ge (x_\infty^*-x^*)^\intercal (x_\infty^*-x_n) \text{ for any } x_\infty^* \le X^*. \tag{9.16}$$

In addition, this point is unique and it has a minimal norm among all possible partial limits x_∞^*.

From Eq. (9.12) one obtains

$$\begin{aligned} 0 \le\ & t(x_n)(x_n^*-x^*)^\intercal \frac{\partial}{\partial x} f(x_n^*,x_n) - \frac{1}{\alpha_n}\|A_0(x_n^*-x^*)\|^2 - \frac{1}{\alpha_n}\|A_1(x_n^*-x^*)\|^2 \\ & - \frac{1}{\alpha_n}(1+\delta_n)^{-1}\|A_1x_n^*-b_1+u_n^*(1+\delta_n)\|^2 - \frac{\delta_n}{\alpha_n}t(x_n)(x_n^*-x^*)^\intercal(x_n^*-x_n) \\ \le\ & t(x_n)(x_n^*-x^*)^\intercal \frac{\partial}{\partial x} f(x_n^*,x_n) - \frac{\delta_n}{\alpha_n}t(x_n)(x_n^*-x^*)^\intercal(x_n^*-x_n). \end{aligned} \tag{9.17}$$

By the strong concavity property

$$(y-z)^\intercal\left(\frac{\partial}{\partial y}f(y)-\frac{\partial}{\partial y}f(z)\right)\le 0 \text{ for any } y,z\in\mathbb{R}^{\mathbb{N}},$$

which, in view of the property (9.15), implies

$$t(x_n)(x_n^*-\hat{x}_n)^\intercal\frac{\partial}{\partial x}f(x_n^*,x_n) = O\left(\sqrt{\delta_n}\right),$$

$$t(x_n)(\hat{x}_n-x^*)^\intercal\frac{\partial}{\partial x}f(\hat{x}_n,x_n) \le t(x_n)(\hat{x}_n-x^*)^\intercal\frac{\partial}{\partial x}f(x^*,x_n)\le 0,$$

then, we have

$$\begin{aligned}
t(x_n)\left(x_n^*-x^*\right)^{\intercal}\frac{\partial}{\partial x}f\left(x_n^*,x_n\right) &= t(x_n)\left(x_n^*-\hat{x}_n\right)^{\intercal}\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)+t(x_n)\left(\hat{x}_n-x^*\right)^{\intercal}\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)\\
&= O\left(\sqrt{\delta_n}\right)+t(x_n)\left(\hat{x}_n-x^*\right)^{\intercal}\left(\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)-\frac{\partial}{\partial x}f\left(\hat{x}_n,x_n\right)\right)\\
&\quad+t(x_n)\left(\hat{x}_n-x^*\right)^{\intercal}\frac{\partial}{\partial x}f\left(\hat{x}_n,x_n\right)\\
&\leq O\left(\sqrt{\delta_n}\right)+t(x_n)\left(\hat{x}_n-x^*\right)^{\intercal}\left(\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)-\frac{\partial}{\partial x}f\left(\hat{x}_n,x_n\right)\right)\\
&\quad+t(x_n)\left(\hat{x}_n-x^*\right)^{\intercal}\frac{\partial}{\partial x}f\left(x^*,x_n\right)\\
&\leq O\left(\sqrt{\delta_n}\right)+t(x_n)\left\|\hat{x}_n-x^*\right\|\left\|\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)-\frac{\partial}{\partial x}f\left(\hat{x}_n,x_n\right)\right\|.
\end{aligned}$$

Since any function is Lipschitz-continuous on any bounded compact set, we can conclude that

$$\left\|\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)-\frac{\partial}{\partial x}f\left(\hat{x}_n,x_n\right)\right\|\leq \text{Const}\left\|x_n^*-\hat{x}_n\right\|=O\left(\sqrt{\delta_n}\right),$$

which gives

$$t(x_n)\left(x_n^*-\hat{x}_n\right)^{\intercal}\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)=O\left(\sqrt{\delta_n}\right),$$

that by Eq. (9.17) leads to

$$\begin{aligned}
0 &\leq t(x_n)\left(x_n^*-\hat{x}_n\right)^{\intercal}\frac{\partial}{\partial x}f\left(x_n^*,x_n\right)-\frac{\delta_n}{\alpha_n}t(x_n)\left(x_n^*-x^*\right)^{\intercal}\left(x_n^*-x_n\right)\\
&= O\left(\sqrt{\delta_n}\right)-\frac{\delta_n}{\alpha_n}t(x_n)\left(x_n^*-x^*\right)^{\intercal}\left(x_n^*-x_n\right).
\end{aligned}\tag{9.18}$$

Dividing both sides of the inequality (9.18) by $\frac{\alpha_n}{\delta_n}$, taking $t(x_n)=1$, and given that $\left\|x_n^*-\hat{x}_n\right\|\leq \kappa\sqrt{\delta_n}$ by Eq. (9.15) we obtain that

$$0\leq O\left(\frac{\alpha_n}{\sqrt{\delta_n}}\right)-\left(x_n^*-x^*\right)^{\intercal}\left(x_n^*-x_n\right)=o\left(1\right)\sqrt{\delta_n}-\left(x_n^*-x^*\right)^{\intercal}\left(x_n^*-x_n\right),$$

which, by Eq. (9.9), for $n\to\infty$ leads to Eq. (9.16). Finally, for any $x^*\leq X^*$ it implies

$$0\geq\left(x_\infty^*-x^*\right)^{\intercal}\left(x_\infty^*-x_n\right)=$$

$$\left\|x_\infty^*-x^*\right\|^2+\left(x_\infty^*-x^*\right)^{\intercal}\left(x^*-x_n\right)\geq\left(x_\infty^*-x^*\right)^{\intercal}\left(x^*-x_n\right).$$

■

9.4 Bargaining with Markov chains

Consider a game with players' strategies denoted by $x^l \in X^l \left(l = \overline{1, \mathfrak{n}}\right)$ where $X := \bigotimes_{l=1}^{\mathfrak{n}} X^l$ is a convex and compact set,

$$x^l := \text{col}\,(c^l), \quad X^l := C^l_{\text{adm}}$$

where *col* is the column operator and C_{adm} satisfies the restrictions (2.7, 2.8 and 2.9).

Denote by $x = (x^1, ..., x^{\mathfrak{n}})^\top \in X$, the joint strategy of the players and $x^{\hat{l}}$ is a strategy of the rest of the players adjoint to x^l, namely,

$$x^{\hat{l}} := \left(x^1, ..., x^{l-1}, x^{l+1}, ..., x^{\mathfrak{n}}\right)^\top \in X^{\hat{l}} := \bigotimes_{h=1,\, h\neq l}^{\mathfrak{n}} X^h$$

such that $x = (x^l, x^{\hat{l}})$, $l = \overline{1, \mathfrak{n}}$.

The process to solve the non-cooperative bargaining game consists of two main steps: firstly to find the initial point of the negotiation (an ideal agreement that players can reach if they negotiate cooperatively, this point is the Pareto optimal solution of the bargaining game), the formulation and solution for this problem is called the strong Nash equilibrium (for the complete formulation, solution and convergence analysis see Chapter 2); while for the solution of the non-cooperative bargaining process we follow the different models presented in section 9.3.

9.4.1 The Pareto optimal solution of the bargaining problem

Consider that players try to reach the strong Nash equilibrium, that is, to find a joint strategy $x^* = (x^{1*}, ..., x^{\mathfrak{n}*}) \in X$ satisfying for any admissible $x^l \in X^l$ and any $l = \overline{1, \mathfrak{n}}$

$$G_{L_p}\left(x(\lambda), \hat{x}(x, \lambda)\right) := \left[\sum_{l=1}^{\mathfrak{n}} \left|\lambda^l \left[\psi^l\left(x^l, x^{\hat{l}}\right) - \psi^l\left(\bar{x}^l, x^{\hat{l}}\right)\right]\right|^p\right]^{1/p}$$

where $\hat{x}(x, \lambda) = (x^{\hat{1}\top}, ..., x^{\hat{\mathfrak{n}}\top})^\top \in \hat{X} \subseteq \mathbb{R}^{\mathfrak{n}(\mathfrak{n}-1)}$, $p \geq 1$ and $\bar{x}^l$ is the utopia point (3.2). Here $\psi^l\left(x^l, x^{\hat{l}}\right)$ is the cost-function of player l which plays the strategy $x^l \in X^l$ and the rest of players the strategy $x^{\hat{l}} \in X^{\hat{l}}$. The functions $\psi^l\left(x^l, x^{\hat{l}}\right)$, $l = \overline{1, \mathfrak{n}}$, are assumed to be concave in all their arguments.

Remark 9.8 *The function* $G_{L_p}\left(u(\lambda),\hat{u}(u,\lambda)\right)$ *satisfies the Nash condition*

$$\psi^l\left(x^l,x^{\hat{l}}\right)-\psi^l\left(\bar{x}^l,x^{\hat{l}}\right)\leq 0$$

for any $x^l\in X^l$ *and all* $l=\overline{1,\mathfrak{n}}$

Applying the Lagrange principle we may conclude

$$x^*_{L_p}=\arg\max_{x\in X,\hat{x}(x)\in\hat{X},\lambda\in\mathcal{S}^{\mathfrak{n}}}\min_{\mu\geq 0,\xi\geq 0,\eta\geq 0}\mathcal{L}_\delta(x,\hat{x}(x),\lambda,\mu,\xi,\eta) \tag{9.19}$$

where

$$\mathcal{L}_\delta(x,\hat{x}(x),\lambda,\mu,\xi,\eta):=G_{L_p,\delta}\left(x(\lambda),\hat{x}(x,\lambda)\right)-\sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N}\mu^l_{(j)}h^l_{(j)}(x^l)-$$

$$\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\xi^l_{(j)}q^l_{(j|i,k)}x^l_{(i,k)}-\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{k=1}^{M}\eta^l\left(x^l_{(i,k)}-1\right)+\frac{\delta}{2}\left(\|\mu\|^2+\|\xi\|^2+\|\eta\|^2\right)$$

and

$$G_{L_p,\delta}\left(x(\lambda),\hat{x}(x,\lambda)\right)=\left[\sum_{l=1}^{\mathfrak{n}}\left|\lambda^l\left[\psi^l\left(x^l,x^{\hat{l}}\right)-\psi^l\left(\bar{x}^l,x^{\hat{l}}\right)\right]\right|^p\right]^{1/p}-\frac{\delta}{2}(\|x\|^2+\|\hat{x}(x)\|^2+\|\lambda\|^2)$$

In order to find the Pareto optimal solution, the relation (9.19) can be expressed in the proximal format as

$$\begin{aligned}
\mu^*_\delta&=\arg\min_{\mu\geq 0}\left\{\tfrac{1}{2}\|\mu-\mu^*_\delta\|^2+\gamma\mathcal{L}_\delta(x^*_\delta,\hat{x}^*_\delta(x),\lambda^*_\delta,\mu,\xi^*_\delta,\eta^*_\delta)\right\}\\
\xi^*_\delta&=\arg\min_{\xi\geq 0}\left\{\tfrac{1}{2}\|\xi-\xi^*_\delta\|^2+\gamma\mathcal{L}_\delta(x^*_\delta,\hat{x}^*_\delta(x),\lambda^*_\delta,\mu^*_\delta,\xi,\eta^*_\delta)\right\}\\
\eta^*_\delta&=\arg\min_{\eta\geq 0}\left\{\tfrac{1}{2}\|\eta-\eta^*_\delta\|^2+\gamma\mathcal{L}_\delta(x^*_\delta,\hat{x}^*_\delta(x),\lambda^*_\delta,\mu^*_\delta,\xi^*_\delta,\eta)\right\}\\
x^*_\delta&=\arg\max_{x\in X}\left\{-\tfrac{1}{2}\|x-x^*_\delta\|^2+\gamma\mathcal{L}_\delta(x,\hat{x}^*_\delta(x),\lambda^*_\delta,\xi^*_\delta)\right\}\\
\hat{x}^*_\delta(x)&=\arg\max_{\hat{x}\in\hat{X}}\left\{-\tfrac{1}{2}\|\hat{x}(x)-\hat{x}^*_\delta(x)\|^2+\gamma\mathcal{L}_\delta(x^*_\delta,\hat{x}(x),\lambda^*_\delta,\xi^*_\delta)\right\}\\
\lambda^*_\delta&=\arg\max_{\lambda\in\mathcal{S}^{\mathcal{N}}}\left\{-\tfrac{1}{2}\|\lambda-\lambda^*_\delta\|^2+\gamma\mathcal{L}_\delta(x^*_\delta,\hat{x}^*_\delta(x),\lambda,\xi^*_\delta)\right\}
\end{aligned} \tag{9.20}$$

where the solutions x^*_δ, $\hat{x}^*_\delta(x)$, λ^*_δ, μ^*_δ, ξ^*_δ and η^*_δ depend on the small parameters $\delta,\gamma>0$.

9.4.2 The non-cooperative bargaining solution

Bargaining model 1

In order to find the non-cooperative bargaining solution, let us define a time function that depends of the transition rates between states of each player as follows

$$\tau^l_{(j|i,k)} := \begin{cases} \dfrac{1}{\left| \sum\limits_{i \neq j}^{N} q^l_{(j|i,k)} \right|} & \text{if} \quad i = j \\ \dfrac{1}{q^l_{(j|i,k)}}, & \text{if} \quad i \neq j \end{cases} \tag{9.21}$$

Also, redefined the utility function in eq. (2.3) to involves the previous time function (9.21)

$$W^l_{(i,k)} = \sum_{j=1}^{N} \left(\tau^l_{(j|i,k)}\right)^{-1} U^l_{(j,i,k)} \pi^l_{(j|i,k)}$$

so that the average utility function in the stationary regime can be expressed as

$$\psi^l(x) := \sum_{i=1}^{N} \sum_{k=1}^{M} W^l_{(i,k)} \prod_{l=1}^{\mathfrak{n}} c^l_{(i,k)} \tag{9.22}$$

Then, define the norm of the strategies x that depends on the transition time cost of each player as follows

$$\|(x - x^*)\|^2_{\Lambda} = \sum_{l=1}^{\mathfrak{n}} \sum_{k=1}^{M} \left\| \left(x^l_{(k)} - x^{l*}_{(k)} \right) \right\|^2 = \sum_{l=1}^{\mathfrak{n}} \sum_{k=1}^{M} \left(x^l_{(k)} - x^{l*}_{(k)} \right)^T \Lambda^l_{(k)} \left(x^l_{(k)} - x^{l*}_{(k)} \right)$$

where

$$x^l_{(k)} = (c^l_{(1,k)}, ..., c^l_{(N,k)})^T \in \mathbb{R}^N, \quad k = \overline{1, M}$$

and

$$\Lambda^l_{(k)} := \frac{1}{2} \left[\tilde{\Lambda}^l_{(k)} + \tilde{\Lambda}^{l\intercal}_{(k)} \right], \quad \tilde{\Lambda}^l_{(k)} := \left[\tau^l_{(j|i,k)} \right], \quad \tilde{\Lambda}^l_{(k)} \in \mathbb{R}^{N \times N}$$

Considering the utility function that depends on the average utility function $\psi^l(x)$ defined as follows

$$F^l(x, \mu, \xi, \eta) := \psi^l(x) - \psi^l(x^*) - \tfrac{1}{2} \sum_{l=1}^{\mathfrak{n}} \sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(x^l) -$$

$$\tfrac{1}{2} \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \xi^l_{(j)} q^l_{(j|i,k)} x^l_{(i,k)} - \tfrac{1}{2} \sum_{l=1}^{\mathfrak{n}} \sum_{i=1}^{N} \sum_{k=1}^{M} \eta^l \left(x^l_{(i,k)} - 1 \right)$$

we may conclude that

$$x^* = \arg\max_{x\in X} \min_{\mu\geq 0,\xi\geq 0,\eta\geq 0} F^l(x,\mu,\xi,\eta)$$

Finally we have that the player in turn has to fix the strategies according to the solution of the non-cooperative bargaining problem in proximal format defined as follows

$$\begin{aligned}
\mu^* &= \arg\min_{\mu\geq 0} \left\{\delta^l \|\mu-\mu^*\|^2 + \alpha^l F^l\left(x^*,\mu,\xi^*,\eta^*\right)\right\}\\
\xi^* &= \arg\min_{\xi\geq 0} \left\{\delta^l \|\xi-\xi^*\|^2 + \alpha^l F^l\left(x^*,\mu^*,\xi,\eta^*\right)\right\}\\
\eta^* &= \arg\min_{\eta\geq 0} \left\{\delta^l \|\eta-\eta^*\|^2 + \alpha^l F^l\left(x^*,\mu^*,\xi^*,\eta\right)\right\}\\
x^* &= \arg\max_{x\in X} \left\{-\delta^l \|(x-x^*)\|_\Lambda^2 + \alpha^l F^l\left(x,\mu^*,\xi^*,\eta^*\right)\right\}
\end{aligned} \tag{9.23}$$

Bargaining model 2

Consider that players try to reach the Nash equilibrium of the bargaining problem, that is, to find a joint strategy $x^* = (x^{1*},...,x^{\mathfrak{n}*}) \in X$ satisfying for any admissible $x^l \in X^l$ and any $l = \overline{1,\mathfrak{n}}$

$$f(x,\hat{x}(x)) := \sum_{l=1}^{\mathfrak{n}} \left[\psi^l\left(x^l, x^{\hat{l}}\right) - \psi^l\left(\bar{x}^l, x^{\hat{l}}\right)\right]$$

where $\hat{x} = (x^{\hat{1}\top},...,x^{\hat{\mathfrak{n}}\top})^\top \in \hat{X} \subseteq \mathbb{R}^{\mathfrak{n}(\mathfrak{n}-1)}$, $\bar{x}^l$ is the utopia point defined as eq. (3.2) and $\psi^l\left(x^l, x^{\hat{l}}\right)$ is the concave cost-function of player l which plays the strategy $x^l \in X^l$ and the rest of players the strategy $x^{\hat{l}} \in X^{\hat{l}}$ defined as eq. (9.22) considering the time function.

Remark 9.9 *The function $f(x,\hat{x}(x))$ satisfies the Nash condition*

$$\psi^l\left(x^l, x^{\hat{l}}\right) - \psi^l\left(\bar{x}^l, x^{\hat{l}}\right) \leq 0$$

for any $x^l \in X^l$ and all $l = \overline{1,\mathfrak{n}}$

We redefine the utility function that depends of the average utility function of all players as follows

$$F(x,\hat{x}(x)) := f(x,\hat{x}(x)) - \tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N} \mu^l_{(j)} h^l_{(j)}(x^l) -$$

$$\tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M} \xi^l_{(j)} q^l_{(j|i,k)} x^l_{(i,k)} - \tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{k=1}^{M} \eta^l\left(x^l_{(i,k)} - 1\right)$$

then, we may conclude that

$$x^* = \arg \max_{x \in X, \hat{x} \in \hat{X}} \min_{\mu \geq 0, \xi \geq 0, \eta \geq 0} F(x, \hat{x}(x), \mu, \xi, \eta)$$

Finally we have that at each step of the bargaining process, players calculate the Nash equilibrium (but they reach the equilibrium at different time) according to the solution of the non-cooperative bargaining problem in proximal format defined as follows

$$\begin{aligned}
\mu^* &= \arg \min_{\mu \geq 0} \{-\delta \|\mu - \mu^*\|^2 + \alpha F\left(x^*, \hat{x}^*(x), \mu, \xi^*, \eta^*\right)\} \\
\xi^* &= \arg \min_{\xi \geq 0} \{-\delta \|\xi - \xi^*\|^2 + \alpha F\left(x^*, \hat{x}^*(x), \mu^*, \xi, \eta^*\right)\} \\
\eta^* &= \arg \min_{\eta \geq 0} \{-\delta \|\eta - \eta^*\|^2 + \alpha F\left(x^*, \hat{x}^*(x), \mu^*, \xi^*, \eta\right)\} \\
x^* &= \arg \max_{x \in X} \left\{-\delta \left\|(x - x^*)\right\|_\Lambda^2 + \alpha F\left(x, \hat{x}^*(x), \mu^*, \xi^*, \eta^*\right)\right\} \\
\hat{x}^* &= \arg \max_{\hat{x} \in \hat{X}} \left\{-\delta \left\|(\hat{x} - \hat{x}^*)\right\|_\Lambda^2 + \alpha F\left(x^*, \hat{x}(x), \mu^*, \xi^*, \eta^*\right)\right\}
\end{aligned} \tag{9.24}$$

Bargaining model 3

For this model, in the same way that we define the strategies $x \in X$, consider a set of strategies denoted by $y^m \in Y^m \left(m = \overline{1, \mathfrak{m}}\right)$ where $Y := \bigotimes_{m=1}^{\mathfrak{m}} Y^l$ is a convex and compact set,

$$y^m := \text{col}\,(c^m), \quad Y^m := C_{adm}^m$$

where *col* is the column operator.

Denote by $y = (y^1, ..., y^{\mathfrak{m}})^\top \in Y$, the joint strategy of the players and $y^{\hat{m}}$ is a strategy of the rest of the players adjoint to y^m, namely,

$$y^{\hat{m}} := \left(y^1, ..., y^{m-1}, y^{m+1}, ..., y^{\mathfrak{m}}\right)^\top \in Y^{\hat{m}} := \bigotimes_{h=1,\, h \neq m}^{\mathfrak{m}} Y^h$$

such that $y = (y^m, y^{\hat{m}})$, $m = \overline{1, \mathfrak{m}}$.

Consider that players of team A try to reach the Nash equilibrium of the bargaining problem, that is, to find a joint strategy $x^* = (x^{1*}, ..., x^{\mathfrak{n}*}) \in X$ satisfying for any admissible $x^l \in X^l$ and any $l = \overline{1, \mathfrak{n}}$

$$f(x, \hat{x}(x)|y) := \sum_{l=1}^{\mathfrak{n}} \left[\psi^l\left(x^l, x^{\hat{l}}|y\right) - \psi^l\left(\bar{x}^l, x^{\hat{l}}|y\right)\right]$$

where $\hat{x} = (x^{\hat{1}\top}, ..., x^{\hat{\mathfrak{n}}\top})^\top \in \hat{X} \subseteq \mathbb{R}^{\mathfrak{n}(\mathfrak{n}-1)}$, $\bar{x}^l$ is the utopia point (3.2) and $\psi^l\left(x^l, x^{\hat{l}}|y\right)$ is the concave cost-function of player l which plays the strategy $x^l \in X^l$ and the rest of players the strategy $x^{\hat{l}} \in X^{\hat{l}}$ fixing the strategies $y \in Y$ of team B, and it is defined as eq. (9.22) considering the time function.

Similarly, consider that players of team B also try to reach the Nash equilibrium of the bargaining problem, that is, to find a joint strategy $y^* = (y^{1*}, ..., y^{\mathfrak{m}*}) \in Y$ satisfying for any admissible $y^m \in Y^m$ and any $m = \overline{1, \mathfrak{m}}$

$$g(y, \hat{y}(y)|x) := \sum_{m=1}^{\mathfrak{m}} \left[\psi^m\left(y^m, y^{\hat{m}}|x\right) - \psi^m\left(\bar{y}^m, y^{\hat{m}}|x\right)\right]$$

where $\hat{y} = (y^{\hat{1}\top}, ..., y^{\hat{\mathfrak{m}}\top})^\top \in \hat{Y} \subseteq \mathbb{R}^{\mathfrak{m}(\mathfrak{m}-1)}$, $\bar{y}^m$ is the utopia point (3.2) and $\psi^m\left(y^m, y^{\hat{m}}|x\right)$ is the concave cost-function of player m which plays the strategy $y^m \in Y^m$ and the rest of players the strategy $y^{\hat{m}} \in Y^{\hat{m}}$ fixing the strategies $x \in X$ of team A, and it is defined as eq. (9.22) considering the time function.

Then, we have that a strategy $x^* \in X$ of team A together with the collection $y^* \in Y$ of team B are defined as the equilibrium of a strictly concave bargaining problem if

$$(x^*, y^*) \;=\; \arg \max_{x \in X_{\text{adm}}, y \in Y_{adm}} \; \{f(x, \hat{x}(x)|y) \leq 0, g(y, \hat{y}(y)|x) \leq 0\}$$

We redefine the utility function that depends of the average utility function of all players as follows

$$F(x, \hat{x}(x), y, \hat{y}(y)) := f(x, \hat{x}(x)|y) + g(y, \hat{y}(y)|x) - \tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{j=1}^{N} \mu_{(j)}^l h_{(j)}^l(x^l) -$$

$$\tfrac{1}{2}\sum_{m=1}^{\mathfrak{m}}\sum_{j=1}^{N} \mu_{(j)}^m h_{(j)}^m(y^m) - \tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M} \xi_{(j)}^l q_{(j|i,k)}^l x_{(i,k)}^l - \tfrac{1}{2}\sum_{m=1}^{\mathfrak{m}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M} \xi_{(j)}^m q_{(j|i,k)}^m y_{(i,k)}^m -$$

$$\tfrac{1}{2}\sum_{l=1}^{\mathfrak{n}}\sum_{i=1}^{N}\sum_{k=1}^{M} \eta^l\left(x_{(i,k)}^l - 1\right) - \tfrac{1}{2}\sum_{m=1}^{\mathfrak{m}}\sum_{i=1}^{N}\sum_{k=1}^{M} \eta^m\left(y_{(i,k)}^m - 1\right)$$

then, we may conclude that

$$(x^*, y^*) = \arg \max_{x \in X, \hat{x} \in \hat{X}, y \in Y, \hat{y} \in \hat{Y}} \; \min_{\mu \geq 0, \xi \geq 0, \eta \geq 0} \; F(x, \hat{x}(x), y, \hat{y}(y), \mu, \xi, \eta)$$

Finally we have that at each step of the bargaining process, players calculate their equilibrium according to the solution of the non-cooperative bargaining problem in proximal format defined

as follows

$$
\begin{aligned}
\mu^* &= \arg\min_{\mu\geq 0}\left\{-\delta\|\mu-\mu^*\|^2+\alpha F\left(x^*,\hat{x}^*(x),y^*,\hat{y}^*(y),\mu,\xi^*,\eta^*\right)\right\}\\
\xi^* &= \arg\min_{\xi\geq 0}\left\{-\delta\|\xi-\xi^*\|^2+\alpha F\left(x^*,\hat{x}^*(x),y^*,\hat{y}^*(y),\mu^*,\xi,\eta^*\right)\right\}\\
\eta^* &= \arg\min_{\eta\geq 0}\left\{-\delta\|\eta-\eta^*\|^2+\alpha F\left(x^*,\hat{x}^*(x),y^*,\hat{y}^*(y),\mu^*,\xi^*,\eta\right)\right\}\\
x^* &= \arg\max_{x\in X}\left\{-\delta\left\|(x-x^*)\right\|_\Lambda^2+\alpha F\left(x,\hat{x}^*(x),y^*,\hat{y}^*(y),\mu^*,\xi^*,\eta^*\right)\right\}\\
\hat{x}^* &= \arg\max_{\hat{x}\in \hat{X}}\left\{-\delta\left\|(\hat{x}-\hat{x}^*)\right\|_\Lambda^2+\alpha F\left(x^*,\hat{x}(x),y^*,\hat{y}^*(y),\mu^*,\xi^*,\eta^*\right)\right\}\\
y^* &= \arg\max_{y\in Y}\left\{-\delta\left\|(y-y^*)\right\|_\Lambda^2+\alpha F\left(x^*,\hat{x}^*(x),y,\hat{y}^*(y),\mu^*,\xi^*,\eta^*\right)\right\}\\
\hat{y}^* &= \arg\max_{\hat{y}\in \hat{Y}}\left\{-\delta\left\|(\hat{y}-\hat{y}^*)\right\|_\Lambda^2+\alpha F\left(x^*,\hat{x}^*(x),y^*,\hat{y}(y),\mu^*,\xi^*,\eta^*\right)\right\}
\end{aligned}
\tag{9.25}
$$

9.5 Numerical Example

Our goal is to analyze a three-player non-cooperative bargaining situation in a class of continuous time Markov chains. Consider a transfer pricing approach which divide the revenue of a passenger between members of an airline alliance. The set of origin-destination time are made up of itineraries. The itineraries are either a direct flight or a series of connecting flights within the supply chain represented by the airlines network. The game penalizes the revenue taking into account the total time that a passenger takes for reaching the final destination. We are taking into account only round trips so the Markov chain game is ergodic.

Let the number of states $N = 3$ and the number of actions $M = 2$ for each airline. The individual utility for each airline are defined by

$$
U^1_{(i,j,1)}=\begin{bmatrix}10 & 8 & 12\\ 6 & 11 & 19\\ 10 & 14 & 13\end{bmatrix}\qquad
U^2_{(i,j,1)}=\begin{bmatrix}7 & 9 & 11\\ 5 & 10 & 14\\ 9 & 6 & 10\end{bmatrix}\qquad
U^3_{(i,j,1)}=\begin{bmatrix}17 & 9 & 6\\ 19 & 13 & 11\\ 3 & 2 & 8\end{bmatrix}
$$

$$
U^1_{(i,j,2)}=\begin{bmatrix}12 & 10 & 5\\ 20 & 16 & 14\\ 18 & 9 & 11\end{bmatrix}\qquad
U^2_{(i,j,2)}=\begin{bmatrix}15 & 6 & 9\\ 15 & 8 & 9\\ 12 & 10 & 7\end{bmatrix}\qquad
U^3_{(i,j,2)}=\begin{bmatrix}10 & 12 & 3\\ 4 & 10 & 9\\ 20 & 17 & 19\end{bmatrix}
$$

The transition rate matrices, i.e. the matrices with the information about the behavior of each airline, are defined as follows

$$q^1_{(j|i,1)}=\begin{bmatrix} -0.2230 & 0.0581 & 0.1649 \\ 0.1166 & -0.3131 & 0.1965 \\ 0.0504 & 0.0531 & -0.1034 \end{bmatrix} \quad q^1_{(j|i,2)}=\begin{bmatrix} -0.8918 & 0.2323 & 0.6595 \\ 0.4664 & -1.2526 & 0.7862 \\ 0.2014 & 0.2122 & -0.4137 \end{bmatrix}$$

$$q^2_{(j|i,1)}=\begin{bmatrix} -0.9336 & 0.7250 & 0.2086 \\ 0.4673 & -0.9428 & 0.4755 \\ 0.0862 & 0.6542 & -0.7405 \end{bmatrix} \quad q^2_{(j|i,2)}=\begin{bmatrix} -0.2334 & 0.1813 & 0.0521 \\ 0.1168 & -0.2357 & 0.1189 \\ 0.0216 & 0.1636 & -0.1851 \end{bmatrix}$$

$$q^3_{(j|i,1)}=\begin{bmatrix} -0.3297 & 0.2872 & 0.0426 \\ 0.0473 & -0.1738 & 0.1265 \\ 0.2912 & 0.2401 & -0.5313 \end{bmatrix} \quad q^3_{(j|i,2)}=\begin{bmatrix} -0.7694 & 0.6700 & 0.0993 \\ 0.1103 & -0.4056 & 0.2953 \\ 0.6794 & 0.5602 & -1.2396 \end{bmatrix}$$

First let us calculate the starting point of the bargaining process applying the proximal method (9.20) to find the strong Nash equilibrium. We obtain the convergence of the strategies in terms of the variable $c^l_{(i,k)}$ for each player (airline) $l = \overline{1,\mathfrak{n}}$ (see Figures 9.2, 9.3 and 9.4) and the convergence of the parameter λ (see Figure 9.5).

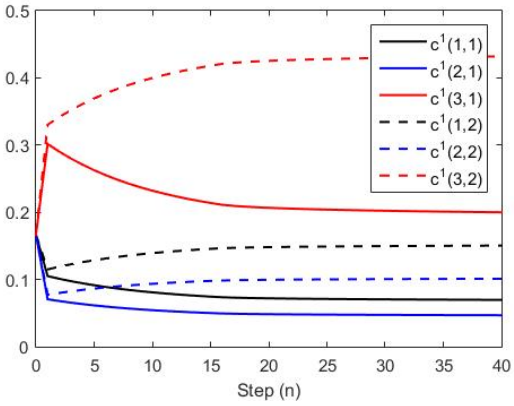


Figure 9.2 SNE Strategies of player 1.

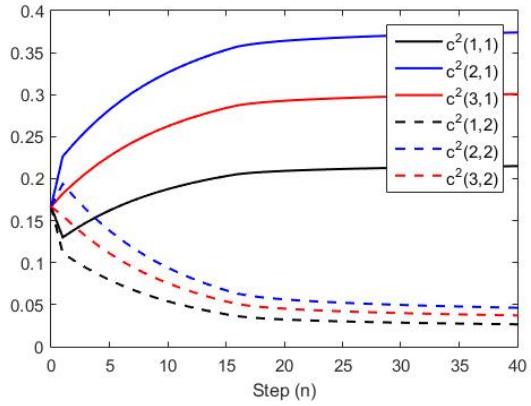


Figure 9.3 SNE Strategies of player 2.

The strong Nash equilibrium reached for all players is as follows:

$$c^1 = \begin{bmatrix} 0.0691 & 0.1510 \\ 0.0464 & 0.1015 \\ 0.1984 & 0.4336 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.2163 & 0.0253 \\ 0.3764 & 0.0440 \\ 0.3026 & 0.0354 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.0071 & 0.2237 \\ 0.0187 & 0.5876 \\ 0.0050 & 0.1579 \end{bmatrix}$$

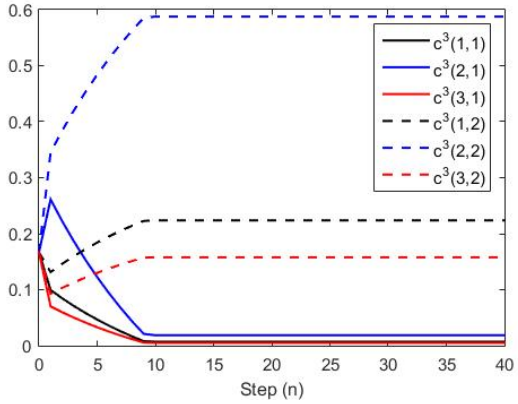


Figure 9.4 SNE Strategies of player 3.

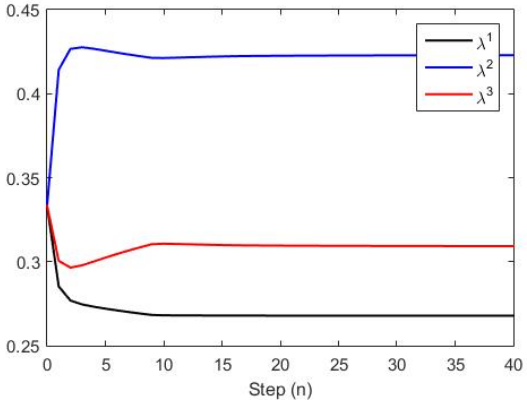


Figure 9.5 Convergence of λ.

The utilities for each player in the strong Nash equilibrium are $\psi^1(c^1, c^2, c^3) = 3842.4$, $\psi^2(c^1, c^2, c^3) = 2961.7$ and $\psi^3(c^1, c^2, c^3) = 3560.3$. Once the starting point is set, the negotiation process between players begins, calculating the strategies until they converge. Then, the results obtained in each of the models presented above are shown:

Bargaining model 1

In this model each player calculates the strategies independently and alternately following the relation (9.23) until they reach an agreement. Figures 9.6, 9.7 and 9.8 show the behavior of the offers (strategies) during the bargaining process.

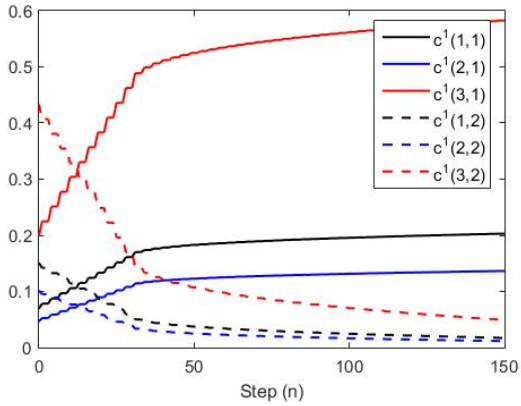


Figure 9.6 Strategies of player 1 in the bargaining model 1.

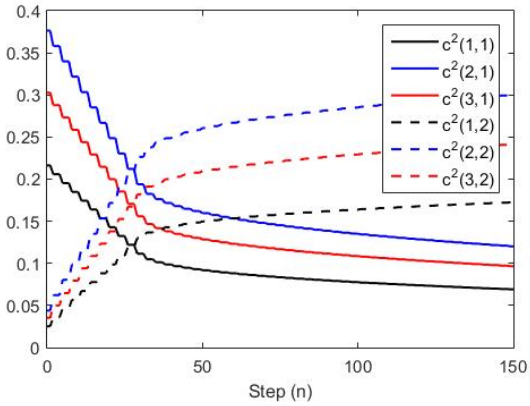


Figure 9.7 Strategies of player 2 in the bargaining model 1.

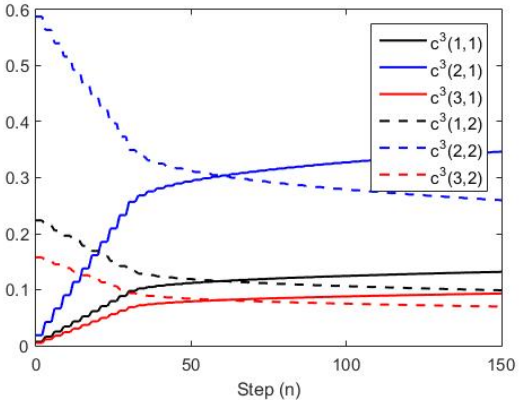


Figure 9.8 Strategies of player 3 in the bargaining model 1.

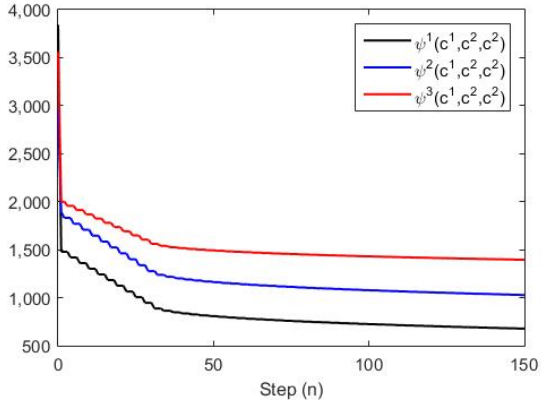


Figure 9.9 Behavior of players' utilities in the bargaining model 1.

Finally, the agreement reached is as follows:

$$c^1 = \begin{bmatrix} 0.2028 & 0.0173 \\ 0.1363 & 0.0116 \\ 0.5824 & 0.0495 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.0691 & 0.1725 \\ 0.1202 & 0.3001 \\ 0.0967 & 0.2413 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.1320 & 0.0988 \\ 0.3469 & 0.2594 \\ 0.0932 & 0.0697 \end{bmatrix}$$

Following (2.6) the mixed strategies obtained for players are as follows

$$d^1 = \begin{bmatrix} 0.9216 & 0.0784 \\ 0.9216 & 0.0784 \\ 0.9216 & 0.0784 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.2860 & 0.7140 \\ 0.2860 & 0.7140 \\ 0.2860 & 0.7140 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.5721 & 0.4279 \\ 0.5721 & 0.4279 \\ 0.5721 & 0.4279 \end{bmatrix}$$

With the strategies calculated at each step of the negotiation process, the utilities of each player showed a decreasing behavior as shown in the Figure 9.9, i.e., at each step of the bargaining process, the utility of each player decreases until they reach an agreement. At the end of the bargaining process, the resulting utilities are as follows $\psi^1(c^1, c^2, c^3) = 678.2$, $\psi^2(c^1, c^2, c^3) = 1028.0$ and $\psi^2(c^1, c^2, c^3) = 1394.3$ for each player.

Bargaining model 2

In this model each player calculates the strategies according the Nash equilibrium formulation where players calculate the Nash equilibrium simultaneously, but with the characteristic

that they reach the equilibrium at different time, following the relation (9.24) until they reach an agreement (strategies show convergence). Figures 9.10, 9.11 and 9.12 show the behavior of the offers (strategies) during the bargaining process.

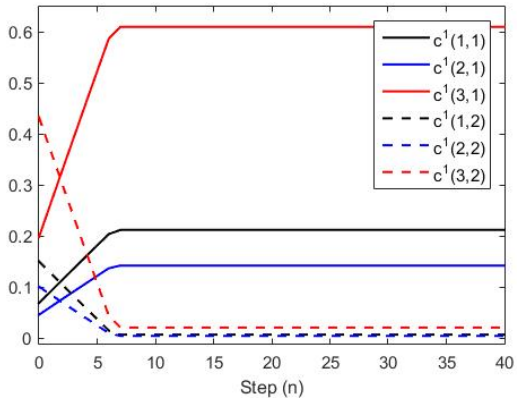


Figure 9.10 Strategies of player 1 in the bargaining model 2.

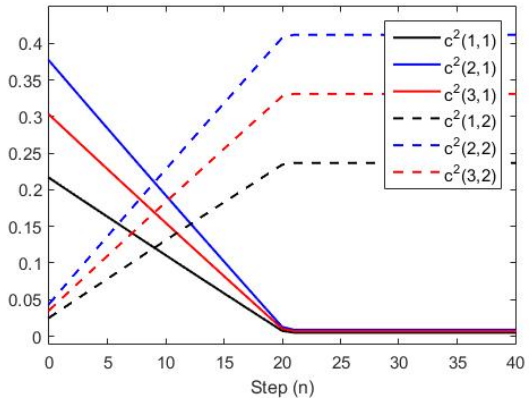


Figure 9.11 Strategies of player 2 in the bargaining model 2.

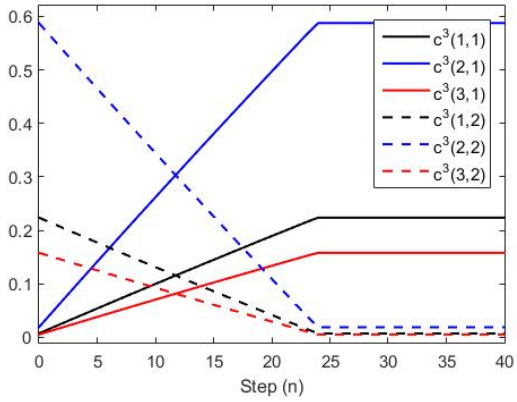


Figure 9.12 Strategies of player 3 in the bargaining model 2.

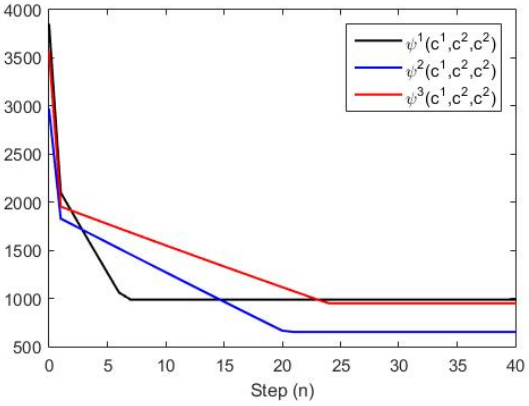


Figure 9.13 Behavior of players' utilities in the bargaining model 2.

Finally, the agreement reached is as follows:

$$c^1 = \begin{bmatrix} 0.2127 & 0.0074 \\ 0.1429 & 0.0050 \\ 0.6106 & 0.0214 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.0050 & 0.2366 \\ 0.0087 & 0.4117 \\ 0.0070 & 0.3310 \end{bmatrix} \qquad c^3 = \begin{bmatrix} 0.2237 & 0.0071 \\ 0.5877 & 0.0186 \\ 0.1579 & 0.0050 \end{bmatrix}$$

Following (2.6) the mixed strategies obtained for players are as follows

$$d^1 = \begin{bmatrix} 0.9662 & 0.0338 \\ 0.9662 & 0.0338 \\ 0.9662 & 0.0338 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.0207 & 0.9793 \\ 0.0207 & 0.9793 \\ 0.0207 & 0.9793 \end{bmatrix} \qquad d^3 = \begin{bmatrix} 0.9693 & 0.0307 \\ 0.9693 & 0.0307 \\ 0.9693 & 0.0307 \end{bmatrix}$$

With the strategies calculated at each step of the negotiation process, the utilities of each player showed a decreasing behavior as shown in the Figure 9.13, i.e., at each step of the bargaining process, the utility of each player decreases until they reach an agreement. At the end of the bargaining process, the resulting utilities are as follows $\psi^1(c^1, c^2, c^3) = 986.8936$, $\psi^2(c^1, c^2, c^3) = 651.4633$ and $\psi^2(c^1, c^2, c^3) = 949.6980$ for each player.

Bargaining model 3

For this model players make teams, in this example as we have three players the team 1 is only formed by player 1 while team 2 is composed of players 2 and 3. Although the players calculate the strategies together following the relation (9.25), we consider that players reach the equilibrium at different times. Figures 9.14, 9.15 and 9.16 show the behavior of the offers (strategies) during the bargaining process.

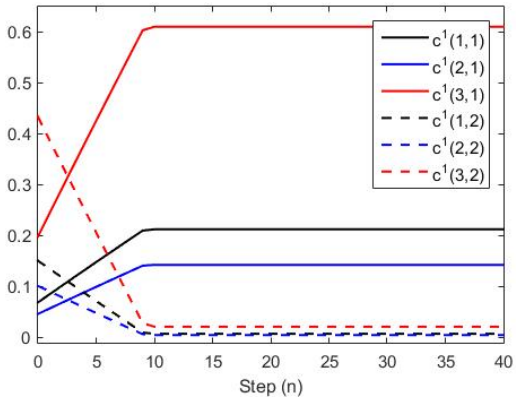


Figure 9.14 Strategies of player 1 in the bargaining model 3.

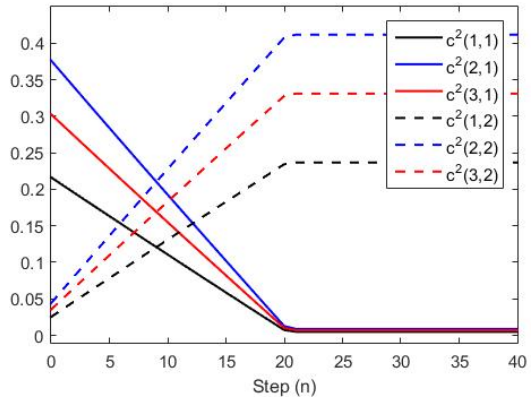


Figure 9.15 Strategies of player 2 in the bargaining model 3.

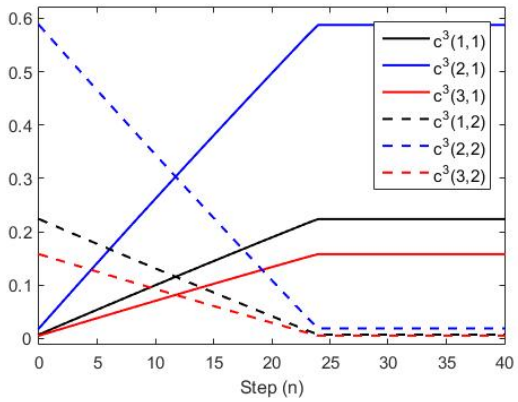


Figure 9.16 Strategies of player 3 in the bargaining model 3.

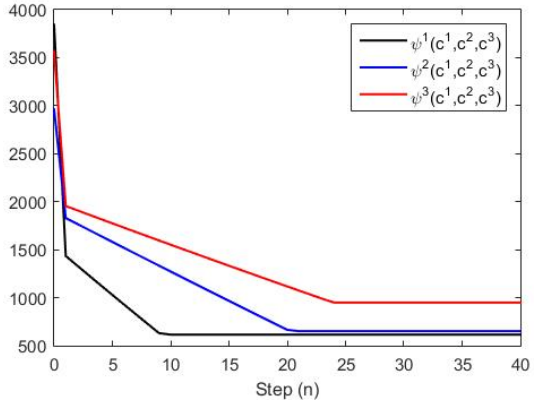


Figure 9.17 Behavior of players' utilities in the bargaining model 3.

Finally, the agreement reached is as follows:

$$c^1 = \begin{bmatrix} 0.2127 & 0.0074 \\ 0.1429 & 0.0050 \\ 0.6106 & 0.0214 \end{bmatrix} \quad c^2 = \begin{bmatrix} 0.0050 & 0.2366 \\ 0.0087 & 0.4117 \\ 0.0070 & 0.3310 \end{bmatrix} \quad c^3 = \begin{bmatrix} 0.2237 & 0.0071 \\ 0.5877 & 0.0186 \\ 0.1579 & 0.0050 \end{bmatrix}$$

Following (2.6) the mixed strategies obtained for players are as follows

$$d^1 = \begin{bmatrix} 0.9662 & 0.0338 \\ 0.9662 & 0.0338 \\ 0.9662 & 0.0338 \end{bmatrix} \quad d^2 = \begin{bmatrix} 0.0207 & 0.9793 \\ 0.0207 & 0.9793 \\ 0.0207 & 0.9793 \end{bmatrix} \quad d^3 = \begin{bmatrix} 0.9693 & 0.0307 \\ 0.9693 & 0.0307 \\ 0.9693 & 0.0307 \end{bmatrix}$$

With the strategies calculated at each step of the negotiation process, the utilities of each player showed a decreasing behavior as shown in the Figure 9.17, i.e., at each step of the bargaining process, the utility of each player decreases until they reach an agreement. At the end of the bargaining process, the resulting utilities are as follows $\psi^1(c^1, c^2, c^3) = 986.8936$, $\psi^2(c^1, c^2, c^3) = 651.4631$ and $\psi^2(c^1, c^2, c^3) = 949.6978$ for each player.

The following figure shows the behavior of the utilities at each of the applied models, we can see that the utilities begin at the same point, the strong Nash equilibrium, and then decrease until the strategies converge (see Figure 9.18). From the results obtained we observed that model 1 favors the utilities of players 2 and 3, while model 2 and 3 are better for player 1. We also observed that even if models 2 and 3 reach the same agreement (equilibrium point) the

strategies and, as a consequence, the utilities have a different behavior during the bargaining process.

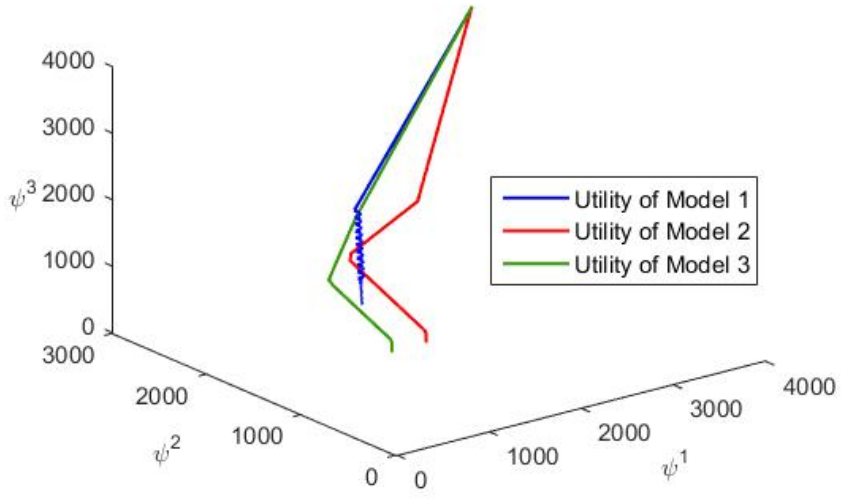


Figure 9.18 Behavior of the utilities at each model.

Chapter 10

Conclusions

This thesis presented models to establish cooperative and non-cooperative strategies for solving different games. It was suggested a novel method for computing the L_p-Nash and the Strong L_p-Nash equilibrium in case of a metric state space. Under mild assumptions, it is shown the existence of L_p-Nash and the Strong L_p-Nash equilibrium characterized as a strong Pareto policy, which is the closest in the Euclidean norm, to the virtual minimum. Following these concepts it was presented a method for computing the strong L_p-Stackelberg/Nash equilibrium, where leaders and followers together are in a Stackelberg game: the model involves two cooperatively Nash games restricted by a Stackelberg game. We should also note that our solution approach essentially simplifies the convergence to a strong Nash equilibrium and to a strong Stackelberg/Nash equilibrium.

A very interesting problem in game theory is the security games that can be described under the Stackelberg formulation. This work addressed dynamic execution uncertainty in security resources allocation presenting a novel approach for adapting attackers and defenders preferred patrolling strategies using a RL process based-on average rewards for Stackelberg security games. More specifically, we presented several contributions. First, we proposed a general RL architecture that combines three different paradigms in reinforcement learning: prior knowledge, imitation and temporal-difference method. The RL architecture involved two components: the Adaptive Primary Learning architecture and the Actor-critic architecture. We showed that the Adaptive Primary Learning architecture accelerates the reinforcement learning process while the Actor-critic architecture determines if rewards are better or worse than

expected based on a game theory solution. For the solution of the game, we considered that defenders and attackers conform coalitions in the Stackelberg security game, respectively. The coalition of the defenders and attackers are reached by computing in case of a metric state space the Strong Lp-Stackelberg/Nash equilibrium. The key result is that this contribution can employ real information available about security uncertainty and generate strategies for scheduling random patrols for different domains of application.

A method to find the equilibria in cooperative and non-cooperative bargaining games was also presented. With respect to cooperative solutions, we examined the bargaining approach from a theoretical perspective and provided a computational solution of the bargaining game for the Nash and Kalai-Smorodinsky models. We first proposed a solution for the disagreement point considered as a Nash equilibrium. Then, to solve the cooperative bargaining problem finding a new agreement point we employed the Nash and the Kalai-Smorodinsky models. We encapsulated both models, first focusing on some of the early results suggested in the literature, and then extending the Nash and Kalai-Smorodinsky analysis to continuous-time Markov games.

Following the results in Nash, Stackelberg and bargaining games, we proposed a new equilibrium point for game theory called the manipulation equilibrium point conceptualized under the Machiavellianism social theory. We employed this equilibrium for proposing a novel approach in solving the bargaining problem. The dynamics and the rationality proposed for the manipulation game correspond with many real-world manipulation situations. The manipulation game is determined by a Stackelberg game model consisting of manipulating and manipulated players that employ manipulation strategies to achieve power situations with the disposition to not become attached to a conventional moral. We represented the Stackelberg game model as Nash game for relaxing the interpretation of the game and the equilibrium selection problem where the weights of the players for the Nash solution are determined by their role in the Stackelberg game. We proposed an analytical formula for solving the manipulation game which arises as the maximum of the quotient of two Nash products which under a feasibility condition is a manipulation equilibrium point. Since a manipulation solution for the

bargaining problem is a particular case of a single-valued function, we analyzed the rationality of the players in the game solution.

In relation to non-cooperative bargaining, a proximal algorithm to solve the non-cooperative bargaining game between two o more unsophisticated players as if they were forward-looking players was presented. To achieve this goal we considered a time penalization related with the time spent for each player for the decision-making at each step of the negotiation process as well as their deviation from the previous best response strategy. We presented three different approaches for the non-cooperative bargaining problem: 1. a game where players are individual-rational and compute the strategies thinking only of their own interests, 2. we consider a game where players calculate the Nash equilibrium simultaneously but they reach the equilibrium point at different times, and 3. a game where players make teams and alternately each team makes an offer to the others until they reach an equilibrium. It was shown that our work complements traditional bargaining literature for myopic agents, but also enlarges the class of processes and functions where Rubinstein's non-cooperative bargaining solutions might be defined and applied.

Our solution approaches are supported by the proximal and extraproximal method. We proposed a set of nonlinear equations represented by the Lagrange optimization method involving the Tikhonov's regularization approach for ensuring the convergence of the solution method to one of the equilibria of the problem. We employed the c-variable method for making the problem computationally tractable. We restricted the solution to a class of Markov chains games. The effectiveness of the proposed methods was validated theoretically and illustrated with some numerical examples.

Appendix A
Proximal Constrained Optimization Approach with Time Penalization

This chapter concerns a proximal-point algorithm with time penalization [105]. The case where the cost to move from one position to a better one is penalized by the time taken by the agent for the decision-making is being studied and the restriction employing the penalty method is incorporated. It is shown that the method converges monotonically with respect to the minimal weighted norm to a unique minimal point under mild assumptions. The gradient method is employed for solving the objective function, and its convergence is proven. The rate of convergence of the method is also estimated by computing the optimal parameters. The effectiveness of the method is illustrated by a numerical optimization example employing continuous-time Markov chains.

To the best of our knowledge, the proximal constrained optimization approach with time penalization is still an open problem. The method presents a main advantage: it involves a cost to move from one position to a better one with penalization of time taken by the agent for the decision-making. This chapter presents the following results:

- Shows that the method converges monotonically with respect to the minimal weighted norm to a unique minimal point under mild assumptions.
- Employs the projection gradient method for solving the objective function.
- Estimates the rate of convergence of the method by computing the optimal parameters.
- Illustrates the method by a numerical example employing continuous-time Markov chains.

A.1 Introduction

Let φ be a convex twice-differentiable real-valued function defined on V, which is a compact and convex subset of $\mathbb{R}^{\mathbb{N}}$. The optimization problem

$$\min_{v \in V} \varphi(v),$$

means the problem of finding the minimal point v^* such that $\varphi(v^*) \leq \varphi(v), \forall\, v \in V$. The optimization problem in this chapter involves an additional penalization or cost to move from one position to a better one, which is related with the proximal point algorithms [4, 5, 70]. The proximal point algorithms for solving non-smooth constrained optimization problems were initially proposed by [57, 79, 53]. Several applications reported in the literature employ proximal point algorithms, for instance see [17, 19, 21, 27]. Indeed, when one uses proximal algorithms suppose choosing an arbitrary initial point $v_0 \in V$ and builds the sequence (v_n), where v_n is the unique solution of the optimization problem of the form

$$\min_{v \in V} \left[\varphi(v) + \frac{\delta_n}{2} \left\| v - v_n \right\|^2 \right],\ \delta_n > 0,\ \delta_n \to 0 \text{ and } n \in \mathbb{N}. \tag{A.1}$$

The term $\|v - v_n\|^2$ ensures that the objective function (A.1) is strictly positive definite [93, 94] and it is introduced in order to improve convergence of some iterative methods. In addition, $\|v - v_n\|^2$ minimizes the distance between v_{n+1} and v_n (v_{n+1} is not far from v_n). Because, $\delta_n > 0$ and $\delta_n \to 0$ the final result obtained is not affected by the quadratic term. An iterative approach, like the projected gradient method, can be employed to solve the objective function (A.1).

The case where the cost to move from one position to a better one is penalized by the time taken by the agent for the decision-making [6, 59, 9] is studied herein. Let $V \subseteq \mathbb{R}^{\mathbb{N}}$ be the decision space (strategies) and define the behavior of an agent as a sequence $(v_n)_{n \in \mathbb{N}}$ where there are possible changes, $v_j \neq v_i$ or remaining in the same, $v_j = v_i$. Then, at each step $n \in \mathbb{N}$ the agent chooses to change or to remain in $v_i \in V$. The function φ represents the cost function that determines the decision to move from v_i. A cost-to-go is defined as a function $\Lambda : V \times V \to \mathbb{R}$ which can be interpreted as a distance function $\Lambda(v_i, v_j) = \kappa(v_i, v_j)$ where $\kappa(v_i, v_j) = 0$ if $v_j = v_i$ or $\kappa(v_i, v_j) > 0$ if $v_i \neq v_j$. In general, the cost to go function can

be reexpressed as $\Lambda(v_i, v_j) := t(v_i, v_j)\kappa(v_i, v_j)$ where $t(v_i, v_j) \geq 0$ is the time spent to move from v_i to v_j and $\kappa(v_i, v_j)$ is the one-step cost-to-go function.

In the simplest (separable) case, $\varphi(v_i) - \varphi(v_j) \geq 0$ is the advantage to change from v_i to v_j given $t(v_j) > 0$, the time spent to benefit of this advantage, and $\alpha(v_i)$ being the weight the agent puts on his advantages to change. Thus, the respective advantages to change from v_i to v_j are given by $A(v_i, v_j) = \alpha(v_i)t(v_j)\left(\varphi(v_i) - \varphi(v_j)\right)$.

The dynamics of the cost to go model is as follows. At each step, the agent considers to change from v_i to v_j, $v_i, v_j \in V$. A transition from v_i to v_j is acceptable if the advantages to change $A(v_i, v_j)$ from v_i to v_j are determined by $\delta(v_i) \in [0, 1]$ (degree of acceptability) of the costs to move $\Lambda(v_i, v_j)$ from v_i to v_j. Then, the set of strategies that minimizes the general cost to go is defined by

$$G(v_i) = \left\{v_j \in V : \alpha(v_i)t(v_j)\left[\varphi(v_i) - \varphi(v_j)\right] \geq \delta(v_i)t(v_j)\kappa(v_i, v_j)\right\}.$$

One can associate a discrete dynamic on V to this relation, whose trajectories $(v_n)_{n\in\mathbb{N}}$ satisfies that $v_{n+1} \in G(v_n)$. Then, in this context, a utility function $\varphi : V \to \mathbb{R}$ such that the impact of experience on cost is constant and limited to the most recent element v_n on the trajectory $(v_n)_{n\in\mathbb{N}}$ is defined. In addition, the advantages to change $A(v_n, v_{n+1})$ are determined by the degree of acceptability $\delta_n(v_n) \in [0, 1]$ of the costs to move $\Lambda(v_n, v_{n+1})$.

Thus, the acceptance criterion to change or stay process satisfies the condition

$$\alpha_n(v_n)t(v_{n+1})\left[\varphi(v_n) - \varphi(v_{n+1})\right] \geq \delta_n(v_n)t(v_{n+1})\kappa(v_n, v_{n+1}).$$

These algorithms are naturally linked with several classical proximal algorithms given in Eq. (A.1). That is, by fixing $v_n = v$ and taking $\delta(v)t(v)\kappa(v, v^*) = \delta_n t(v)\left\|(v - v^*)\right\|^2$ and $A(v, v^*) := \alpha_n t(v)\left[\varphi(v) - \varphi(v^*)\right]$ as one has in proximal format that

$$v^* = \arg\min_{v\in V}\left\{\delta_n t(v)\left\|(v - v^*)\right\|^2 + \alpha_n t(v)\left[\varphi(v) - \varphi(v^*)\right]\right\}.$$

A.2 Formulation of the problem

Consider the following constrained programming problem

$$\begin{gathered}\min_{v\in V_{\text{adm}}}\ \varphi(v)\,,\\ V_{\text{adm}}:=\left\{v\in\mathbb{R}^{\mathbb{N}}:v\geq 0,\ A_{\text{eq}}v=b_{\text{eq}}\in\mathbb{R}^{M_0},\ A_{\text{ineq}}v\leq b_{\text{ineq}}\in\mathbb{R}^{M_1}\right\}.\end{gathered}\tag{A.2}$$

where V_{adm} is a bounded set. Introducing the "slack" vectors $u\in\mathbb{R}^{M_1}$ with nonnegative components, that is, $u_j\geq 0$ for all $j=1,...,M_1$, the original problem (A.2) can be rewritten as

$$\begin{gathered}\min_{v\in V_{\text{adm}},\ u\geq 0}\ \varphi(v)\,,\\ V_{\text{adm}}:=\left\{v\in\mathbb{R}^{\mathbb{N}}:v\geq 0,\ A_{\text{eq}}v=b_{\text{eq}},\ A_{\text{ineq}}v-b_{\text{ineq}}+u=0\right\}.\end{gathered}\tag{A.3}$$

Define by $V^*\subseteq V_{\text{adm}}$ the set of all solutions of the problem (A.3).

Consider the *objective function* given by

$$\begin{gathered}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right):=t(v_n)\tfrac{\delta}{2}\left\|v-v_n\right\|^2+\alpha t(v_n)\left(\varphi(v)-\varphi(v_n)\right)+\\ \tfrac{1}{2}\left\|A_{\text{eq}}v-b_{\text{eq}}\right\|^2+\tfrac{1}{2}\left\|A_{\text{ineq}}v-b_{\text{ineq}}+u\right\|^2+\tfrac{\delta}{2}\left\|u\right\|^2\end{gathered}\tag{A.4}$$

where $\alpha,\ \delta>0$. The problem of calculating the fixed point of the extremal mapping will be considered

$$\min_{v\in V_{\text{adm}},\ u\geq 0}\ \mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)\ ,$$

such that

$$v_n^*:=v^*\left(\alpha_n,\delta_n\right)\underset{n\to\infty}{\to}v^{**},\qquad u_n^*:=u^*\left(\alpha_n,\delta_n\right)\underset{n\to\infty}{\to}u^{**},$$

considering that the parameters α and δ are time-varying, *i.e.*,

$$\alpha=\alpha_n,\quad \delta=\delta_n,\quad (n=0,1,2,....)\,,$$

and

$$0<\alpha_n\to 0,\quad \frac{\alpha_n}{\delta_n}\to 0,\quad \text{when } n\to\infty.\tag{A.5}$$

In addition, $v^{**}\left(\alpha,\delta\right)\in V^*$ is the solution of the original problem (A.3) with minimal weighted norm, *i.e.*,

$$\left\|v^{**}\right\|\leq\left\|v^*\right\|\ \text{ for all } v^*\in V^*,$$

and

$$u^* = b_{\text{ineq}} - A_{\text{ineq}} v^*.$$

Moreover, the bounded set V^* of all solutions of the original optimization problem given in Eq. (A.3) is not empty and Slater's condition holds [74], that is, there exists a point $\mathring{v} \in V_{\text{adm}}$ such that

$$A_{\text{ineq}} \mathring{v} < b_{\text{ineq}}. \tag{A.6}$$

A.3 Convergence analysis

The behavior of the iterative proximal method of the original problem (A.3) will be studied. First, it will be proven that the Hessian matrix $\mathbb{H}$ associated with the objective function (A.4) is strictly positive definite for any positive α and δ, to show that the objective function (A.4) is strictly convex. The following theorem if formulated.

Theorem A.1 *If the set of solutions of problem (*A.*3) is non-empty then the objective function (*A.*4) is strictly convex.*

Proof. It should be proven that for all $v \in \mathbb{R}^{\mathbb{N}}$ and $u \in \mathbb{R}^{M_1}$

$$\mathbb{H} = \begin{bmatrix} \frac{\partial^2}{\partial v^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) & \frac{\partial^2}{\partial u \partial v}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) \\ \frac{\partial^2}{\partial v \partial u}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) & \frac{\partial^2}{\partial u^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) \end{bmatrix} > 0,$$

Employing Schur's lemma [74] it is necessary and sufficient to prove that

$$\begin{gathered} \frac{\partial^2}{\partial v^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) > 0,\ \frac{\partial^2}{\partial u^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) > 0, \\ \frac{\partial^2}{\partial v^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) > \frac{\partial^2}{\partial u \partial v}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)\left[\frac{\partial^2}{\partial u^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)\right]^{-1}\frac{\partial^2}{\partial v \partial u}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right). \end{gathered} \tag{A.7}$$

Then, applying the Schur's lemma over the objective function (A.4) it follows that

$$\frac{\partial^2}{\partial v^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right) = \alpha t(v_n)\frac{\partial^2}{\partial v^2}\left(\varphi(v) - \varphi(v_n)\right) + A_{\text{eq}}^{\intercal}A_{\text{eq}} + A_{\text{ineq}}^{\intercal}A_{\text{ineq}} + t(v_n)\delta I_{\mathbb{N}\times\mathbb{N}} \geq$$

$$\alpha t(v_n)\frac{\partial^2}{\partial v^2}\left(\varphi(v) - \varphi(v_n)\right) + t(v_n)\delta I_{\mathbb{N}\times\mathbb{N}} \geq t(v_n)\delta\left(1 + \frac{\alpha}{\delta}\lambda^{-}\right) I_{\mathbb{N}\times\mathbb{N}} > 0\ \forall\ \delta_n > 0,$$

where

$$\lambda^{-} := \min_{v \in V_{\text{adm}}} \lambda_{\min}\left(\frac{\partial^2}{\partial v^2}\left(\varphi(v) - \varphi(v_n)\right)\right),$$

and

$$\frac{\partial^2}{\partial u^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)=\left(1+\delta\right)I_{M_1\times M_1}>0.$$

By Eq. (A.7) it is necessary to satisfy that

$$\frac{\partial^2}{\partial v^2}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)=\alpha t(v_n)\frac{\partial^2}{\partial v^2}\left(\varphi(v)-\varphi(v_n)\right)+A_{\text{eq}}^{\intercal}A_{\text{eq}}+A_{\text{ineq}}^{\intercal}A_{\text{ineq}}+t(v_n)\delta I_{\mathbb{N}\times\mathbb{N}}>$$
$$\frac{\partial^2}{\partial u\partial v}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)\left[\frac{\partial^2}{\partial u^2}\mathbb{F}_{\alpha,\delta}\left(v,u\right)\right]^{-1}\frac{\partial^2}{\partial v\partial u}\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)=\left(1+\delta\right)^{-1}A_{\text{ineq}}^{\intercal}A_{\text{ineq}},$$

or equivalently,

$$\alpha t(v_n)\frac{\partial^2}{\partial v^2}\left(\varphi(v)-\varphi(v_n)\right)+A_{\text{eq}}^{\intercal}A_{\text{eq}}+\frac{\delta}{1+\delta}A_{\text{ineq}}^{\intercal}A_{\text{ineq}}+t(v_n)\delta I_{\mathbb{N}\times\mathbb{N}}>0,$$

which holds for any $\delta>0$ having

$$t(v_n)\left(\alpha\lambda^{-}+\delta\right)I_{\mathbb{N}\times\mathbb{N}}+A_{\text{eq}}^{\intercal}A_{\text{eq}}+\frac{\delta}{1+\delta}A_{\text{ineq}}^{\intercal}A_{\text{ineq}}\geq$$
$$t(v_n)\delta\left(1+\tfrac{\alpha}{\delta}\lambda^{-}\right)I_{\mathbb{N}\times\mathbb{N}}=t(v_n)\delta\left(1+o(1)\right)I_{\mathbb{N}\times\mathbb{N}}>0.$$

As a result, the Hessian is $\mathbb{H}>0$ which means that proximal function (A.4) is strictly convex.

■

Remark A.2 *The Hessian $\mathbb{H}>0$ is a sufficient condition for the convergence to a unique minimal point defined $v^*\left(\alpha,\delta\right)$ and $u^*\left(\alpha,\delta\right)$ for the proximal function (A.4).*

Next, objective function (A.4) is considered to be strictly convex and it is shown that it converges to a unique minimal point that depends of the parameters α and δ.

Theorem A.3 *If the proximal function (A.4) is strictly convex then the sequence $\{v_n\}$ of the proximal function (A.4) converges when $n\to\infty$, i.e. the proximal function (A.4) has a minimal point defined by $v^*\left(\alpha,\delta\right)$ and $u^*\left(\alpha,\delta\right)$.*

Proof. The theorem will be proven in two parts.

i) Following the strictly convexity property (Theorem A.1) for any $w:=\begin{pmatrix}v\\u\end{pmatrix}$ and any vector $w_n^*:=\begin{pmatrix}v_n^*=v^*\left(\alpha_n,\delta_n\right)\\u_n^*=u^*\left(\alpha_n,\delta_n\right)\end{pmatrix}$ for the function $\mathbb{F}_{\alpha,\delta}\left(v,u|v_n\right)=\mathbb{F}_{\alpha,\delta}\left(w|v_n\right)$ one has

$$
\begin{gathered}
0 \geq \left(w_n^* - w\right)^\intercal \frac{\partial}{\partial w} \mathbb{F}_{\alpha_n,\delta_n}\left(w_n^* | v_n\right) = \left(v_n^* - v\right)^\intercal \frac{\partial}{\partial v} \mathbb{F}_{\alpha_n,\delta_n}\left(v_n^*, u_n^* | v_n\right) + \\
\left(u_n^* - u\right)^\intercal \frac{\partial}{\partial u} \mathbb{F}_{\alpha_n,\delta_n}\left(v_n^*, u_n^* | v_n\right) = \left(v_n^* - v\right)^\intercal \Bigg(\alpha_n t(v_n) \frac{\partial}{\partial v} (\varphi\left(v_n^*\right) - \varphi\left(v_n\right)) + \\
A_{\text{eq}}^\intercal \left[A_{\text{eq}} v_n^* - b_{\text{eq}}\right] + A_{\text{ineq}}^\intercal \left[A_{\text{ineq}} v_n^* - b_{\text{ineq}} + u_n^*\right] + t(v_n)\delta_n (v_n^* - v_n)\Big) + \\
\left(u_n^* - u\right)^\intercal \left(A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right) = \\
\alpha_n t(v_n) \left(v_n^* - v\right)^\intercal \frac{\partial}{\partial v} (\varphi\left(v_n^*\right) - \varphi\left(v_n\right)) + \left[A_{\text{eq}}\left(v_n^* - v\right)\right]^\intercal \left[A_{\text{eq}} v_n^* - b_{\text{eq}}\right] + \\
\left[A_{\text{ineq}}\left(v_n^* - v\right)\right]^\intercal \left[A_{\text{ineq}} v_n^* - b_{\text{ineq}} + u_n^*\right] + t(v_n)\delta_n \left(v_n^* - v\right)^\intercal \left(v_n^* - v_n\right) + \\
\left(u_n^* - u\right)^\intercal \left[A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right].
\end{gathered}
\tag{A.8}
$$

Selecting in Eq. (A.8) $v := v^* \in V^*$ (v^* is one of the admissible solutions such that $A_{\text{eq}} v^* = b_{\text{eq}}$) and $u := (1+\delta_n)^{-1}\left(b_{\text{ineq}} - A_{\text{ineq}} v_n^*\right)$ one obtains

$$
\begin{gathered}
0 \geq \alpha_n t(v_n) \left(v_n^* - v^*\right)^\intercal \frac{\partial}{\partial v} (\varphi\left(v_n^*\right) - \varphi\left(v_n\right)) + \left[A_{\text{eq}}\left(v_n^* - v^*\right)\right]^\intercal \left[A_{\text{eq}} v_n^* - b_{\text{eq}}\right] + \\
\left[A_{\text{ineq}}\left(v_n^* - v^*\right)\right]^\intercal \left[A_{\text{ineq}} v_n^* - b_{\text{ineq}} + u_n^*\right] + t(v_n)\delta_n \left(v_n^* - v^*\right)^\intercal \left(v_n^* - v_n\right) + \\
(1+\delta_n)^{-1} \left(u_n^* (1+\delta_n) - b_{\text{ineq}} + A_{\text{ineq}} v_n^*\right)^\intercal \left(A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right) + \\
\delta_n \left(u_n^* - b_{\text{ineq}} - A_{\text{ineq}} v_n^*\right)^\intercal u_n^*,
\end{gathered}
\tag{A.9}
$$

simplifying Eq. (A.9) one obtains

$$
\begin{gathered}
0 \geq \alpha_n t(v_n) \left(v_n^* - v^*\right)^\intercal \frac{\partial}{\partial v} (\varphi\left(v_n^*\right) - \varphi\left(v_n\right)) + \left\|A_{\text{eq}}\left(v_n^* - v^*\right)\right\|^2 + \left\|A_{\text{ineq}}\left(v_n^* - v^*\right)\right\|^2 + \\
t(v_n)\delta_n \left(v_n^* - v^*\right)^\intercal \left(v_n^* - v_n\right) + (1+\delta_n)^{-1} \left\|A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right\|^2 + \\
\delta_n \left(u_n^* - b_{\text{ineq}} - A_{\text{ineq}} v_n^*\right)^\intercal u_n^*.
\end{gathered}
$$

Dividing both sides of this inequality by δ_n one gets

$$
\begin{gathered}
0 \geq \frac{\alpha_n}{\delta_n} t(v_n) \left(v_n^* - v^*\right)^\intercal \frac{\partial}{\partial v} (\varphi\left(v_n^*\right) - \varphi\left(v_n\right)) + \\
\frac{1}{\delta_n} \left(\left\|A_{\text{eq}}\left(v_n^* - v^*\right)\right\|^2 + \left\|A_{\text{ineq}}\left(v_n^* - v^*\right)\right\|^2 + (1+\delta_n)^{-1} \left\|A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right\|^2\right) + \\
t(v_n) \left(v_n^* - v^*\right)^\intercal \left(v_n^* - v_n\right) + \left(u_n^* - b_{\text{ineq}} - A_{\text{ineq}} v_n^*\right)^\intercal u_n^*.
\end{gathered}
\tag{A.10}
$$

Now, taking $v = v_n^*$ and $u = 0$ from Eq. (A.8) one has

$$0 \geq (u_n^*)^\intercal \left[A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right] = (u_n^*)^\intercal \left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right) + (1+\delta_n) \left\|u_n^*\right\|^2 =$$

$$\left(\left\|\sqrt{1+\delta_n} u_n^*\right\|^2 + 2\left(\sqrt{1+\delta_n} u_n^*\right)^\intercal \left[\frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right] + \left\|\frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2 - \right.$$

$$\left. \left\|\frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2 \right) = \left[\left\|\sqrt{1+\delta_n} u_n^* + \frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2 - \left\|\frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2 \right],$$

implying

$$\left\|\frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2 \geq \left\|\sqrt{1+\delta_n} u_n^* + \frac{\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)}{2\sqrt{1+\delta_n}}\right\|^2,$$

and

$$1 \geq \left\| e + 2\,(1+\delta_n)\, u_n^* \left\|\left(A_{\text{ineq}} v_n^* - b_{\text{ineq}}\right)\right\|^{-1} \right\|^2, \quad \|e\| = 1,$$

which means that the sequence $\{u_n^*\}$ is bounded. If it is assumed that $\dfrac{\alpha_n}{\delta_n} \underset{n\to\infty}{\to} 0$, from Eq. (A.10) it follows

$$\begin{aligned} \text{Const} = \limsup_{n\to\infty} \left(\left|(v_n^* - v^*)^\intercal (v_n^* - v_n)\right| + \left|(u_n^* - b_{\text{ineq}} - A_{\text{ineq}} v_n^*)^\intercal u_n^*\right|\right) \geq \\ \limsup_{n\to\infty} \frac{1}{\delta_n} \left(\left\|A_{\text{eq}} v_n^* - b_{\text{eq}}\right\|^2 + \left\|A_{\text{ineq}} (v_n^* - v^*)\right\|^2 + \right. \\ \left. (1+\delta_n)^{-1} \left\|A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right\|^2\right). \end{aligned} \tag{A.11}$$

From Eq. (A.11) one may conclude that

$$\begin{aligned} \left\|A_{\text{eq}} v_n^* - b_{\text{eq}}\right\|^2 + \left\|A_{\text{ineq}} (v_n^* - v^*)\right\|^2 + \\ (1+\delta_n)^{-1} \left\|A_{\text{ineq}} v_n^* - b_{\text{ineq}} + (1+\delta_n)\, u_n^*\right\|^2 = O(\delta_n), \end{aligned} \tag{A.12}$$

and

$$A_{\text{eq}} v_\infty^* - b_{\text{eq}} = 0,$$

$$A_{\text{ineq}} v_\infty^* - A_{\text{ineq}} v^* = A_{\text{ineq}} v_\infty^* - b_{\text{ineq}} + u_\infty^* = 0,$$

where $v_\infty^* \in V^*$ is a partial limit of the sequence $\{v_n^*\}$ which, obviously, may be not unique. The vector u_∞^* is also a partial limit of the sequence $\{u_n^*\}$.

ii) Denote by $\hat{v}_n$ the projection of v_n^* to the set V_{adm}, *i.e.*

$$\hat{v}_n = \Pr_{V_{\text{adm}}} (v_n^*),$$

where Pr is the projection operator. It is shown that

$$\|v_n^* - \hat{v}_n\| \leq \kappa\sqrt{\delta_n}, \quad \kappa = \text{const} > 0. \tag{A.13}$$

Given Eq. (A.12) one has that

$$\|A_{\text{ineq}} v_n^* - b_{\text{ineq}} + u_n^*\| \leq \kappa_1\sqrt{\delta_n}, \quad \kappa_1 = \text{const} > 0,$$

implying

$$A_{\text{ineq}} v_n^* - b_{\text{ineq}} \leq \kappa_1\sqrt{\delta_n} e - u_n^* \leq \kappa_1\sqrt{\delta_n} e, \quad \|e\| = 1,$$

where the vector inequality is treated in component-wise sense:

$$\|v_n^* - \hat{v}_n\|^2 \leq \max_{A_{\text{ineq}} v_n - b_{\text{ineq}} \leq \kappa_1\sqrt{\delta_n} e,\, v \in V_{\text{adm}}} \min_{z \in V_{\text{adm}}} \|v - z\|^2 := d(\delta_n).$$

Define

$$\tilde{v} := (1 - x_n)\, v + x_n \mathring{v} \quad \in V_{\text{adm}},$$

by Slater's condition given in Eq. (A.6) one obtains that

$$0 < x_n := \frac{\kappa_1\sqrt{\delta_n}}{\kappa_1\sqrt{\delta_n} + \min\limits_{j=1,\ldots,M_1} \left|\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right)_j\right|} < 1.$$

For the variable $v = \dfrac{\tilde{v} - x_n\mathring{v}}{1 - x_n}$ one has

$$A_{\text{ineq}}\tilde{v} - b_{\text{ineq}} = (1 - x_n)\, A_{\text{ineq}} v + x_n A_{\text{ineq}}\mathring{v} - b_{\text{ineq}} =$$

$$(1 - x_n)\left(A_{\text{ineq}} v - b_{\text{ineq}}\right) + (1 - x_n)\, b_{\text{ineq}} + x_n\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right) + x_n b_{\text{ineq}} - b_{\text{ineq}} =$$

$$(1 - x_n)\left(A_{\text{ineq}} v - b_{\text{ineq}}\right) + x_n\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right) \leq$$

$$(1 - x_n)\,\kappa_1\sqrt{\delta_n} e + \frac{\kappa_1\sqrt{\delta_n}}{\kappa_1\sqrt{\delta_n} + \min\limits_{j=1,\ldots,M_1} \left|\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right)_j\right|}\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right) =$$

$$\frac{\kappa_1\sqrt{\delta_n}}{\kappa_1\sqrt{\delta_n} + \min\limits_{j=1,\ldots,M_1} \left|\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right)_j\right|}\left(\min_{j=1,\ldots,M_1} \left|\left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right)_j\right| e + \left(A_{\text{ineq}}\mathring{v} - b_{\text{ineq}}\right)\right) \leq 0,$$

then

$$d(\delta_n) = \max_{A_{\text{ineq}}v - b_{\text{ineq}} \le \kappa_1\sqrt{\delta_n}e,\ v\in V_{\text{adm}}} \min_{y\in V_{\text{adm}}} \|x-y\|^2 \le$$

$$\max_{A_{\text{ineq}}\tilde{v} - b_{\text{ineq}} \le 0,\ \tilde{v}\in V_{\text{adm}}} \left\| \frac{\tilde{v} - x_n\mathring{v}}{1-x_n} - \tilde{x} \right\|^2 =$$

$$\frac{x_n^2}{(1-x_n)^2} \max_{A_{\text{ineq}}\tilde{v} - b_{\text{ineq}} \le 0,\ \tilde{v}\in V_{\text{adm}}} \|\tilde{v} - \mathring{v}\|^2 \le \kappa_1\delta_n, \quad 0 < \kappa_1 < \infty.$$

Given that $\|v_n^* - \hat{v}_n\| \le \sqrt{d(\delta_n)} \le \sqrt{\kappa_1 = \text{const} > 0}\sqrt{\delta_n}$ proving Eq. (A.13). ■

Finally, in the following theorem it is shown that the sequence $\{v_n\}$ converges with minimal norm to v^*.

Theorem A.4 *If the proximal function (*A.*4) is strictly convex and the sequence* $\{v_n\}$ *of the proximal function (*A.*4) converges, then, the necessary and sufficient condition for the point* v^* *to be the minimum point of the function* $\|v_\infty^*\|^2$ *on the set* V^* *is given by*

$$0 \ge (v_\infty^* - v^*)^\intercal (v_\infty^* - v_n) \text{ for any } v_\infty^* \le V^*. \tag{A.14}$$

In addition, this point is unique and it has a minimal norm among all possible partial limits v_∞^*.

Proof. From Eq. (A.10) one obtains

$$\begin{aligned} 0 \ge t(v_n)(v_n^* - v^*)^\intercal \frac{\partial}{\partial v}(\varphi(v_n^*) - \varphi(v_n)) + \frac{1}{\alpha_n}\left(\|A_{\text{eq}}v_n^* - b_{\text{eq}}\|^2 + \|A_{\text{ineq}}(v_n^* - v^*)\|^2\right) + \\ \frac{\delta_n}{\alpha_n} t(v_n)(v_n^* - v^*)^\intercal (v_n^* - v_n) + \frac{1}{\alpha_n}\|A_{\text{ineq}}v_n^* - b_{\text{ineq}} + (1+\delta_n)u_n^*\|^2 \ge \\ t(v_n)(v_n^* - v^*)^\intercal \frac{\partial}{\partial v}(\varphi(v_n^*) - \varphi(v_n)) + \frac{\delta_n}{\alpha_n} t(v_n)(v_n^* - v^*)^\intercal (v_n^* - v_n). \end{aligned} \tag{A.15}$$

By the strong convexity property (see Corollary 21.4 in [74]) it follows that

$$(x-y)^\intercal \left(\frac{\partial}{\partial x}\varphi(x) - \frac{\partial}{\partial x}\varphi(y)\right) \ge 0 \text{ for any } x, y \in \mathbb{R}^{\mathbb{N}},$$

which, in view of the property (A.13), implies

$$t(v_n)\,(v_n^* - \hat{v}_n)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) = O\left(\sqrt{\delta_n}\right),$$

$$t(v_n)\,(\hat{v}_n - v^*)^\intercal \frac{\partial}{\partial v}(\varphi\,(\hat{v}_n) - \varphi\,(v_n)) \geq 0,$$

$$t(v_n)\,(\hat{v}_n - v^*)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) \geq t(v_n)\,(v_n^* - v^*)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) =$$
$$t(v_n)\,(v_n^* - \hat{v}_n)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) + t(v_n)\,(\hat{v}_n - v^*)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) \geq$$
$$O\left(\sqrt{\delta_n}\right) + t(v_n)\,(\hat{v}_n - v^*)^\intercal \cdot$$
$$\left(\frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) - \frac{\partial}{\partial v}(\varphi\,(\hat{v}_n) - \varphi\,(v_n))\right) + t(v_n)\,(\hat{v}_n - v^*)^\intercal \frac{\partial}{\partial v}(\varphi\,(\hat{v}_n) - \varphi\,(v_n)) \geq$$
$$O\left(\sqrt{\delta_n}\right) - t(v_n)\,\|\hat{v}_n - v^*\| \left\| \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) - \frac{\partial}{\partial v}(\varphi\,(\hat{v}_n) - \varphi\,(v_n)) \right\|.$$

Since any function is Lipschitz-continuous on any bounded compact set, one can conclude that

$$\left\| \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) - \frac{\partial}{\partial v}(\varphi\,(\hat{v}_n) - \varphi\,(v_n)) \right\| \leq \text{Const}\,\|v_n^* - \hat{v}_n\| = O\left(\sqrt{\delta_n}\right),$$

which gives

$$t(v_n)\,(v_n^* - \hat{v}_n)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) = O\left(\sqrt{\delta_n}\right),$$

which, by Eq. (A.15) leads to

$$\begin{gathered} 0 \geq t(v_n)\,(v_n^* - \hat{v}_n)^\intercal \frac{\partial}{\partial v}(\varphi\,(v_n^*) - \varphi\,(v_n)) + \frac{\delta_n}{\alpha_n} t(v_n)\,(v_n^* - v^*)^\intercal\,(v_n^* - v_n) = \\ O\left(\sqrt{\delta_n}\right) + \frac{\delta_n}{\alpha_n} t(v_n)\,(v_n^* - v^*)^\intercal\,(v_n^* - v_n). \end{gathered} \tag{A.16}$$

Dividing both sides of the inequality (A.16) by $\frac{\alpha_n}{\delta_n}$, taking $t(v_n) = 1$, and given that $\|v_n^* - \hat{v}_n\| \leq \kappa\sqrt{\delta_n}$ by Eq. (A.13) one has that

$$0 \geq O\left(\frac{\alpha_n}{\sqrt{\delta_n}}\right) + (v_n^* - v^*)^\intercal\,(v_n^* - v_n) = o\,(1)\,\sqrt{\delta_n} + (v_n^* - v^*)^\intercal\,(v_n^* - v_n),$$

which, by Eq. (A.5), for $n \to \infty$ leads to Eq. (A.14). Finally, for any $v^* \leq V^*$ it implies

$$0 \geq (v_\infty^* - v^*)^\intercal\,(v_\infty^* - v_n) =$$

$$\|v_\infty^* - v^*\|^2 + (v_\infty^* - v^*)^\intercal\,(v^* - v_n) \geq (v_\infty^* - v^*)^\intercal\,(v^* - v_n).$$

■

A.4 Gradient solver

Consider the *proximal function* for finding the unique minimal point defined by $v^*(\alpha,\delta)$ and $u^*(\alpha,\delta)$

$$\mathbb{F}_{\alpha,\delta}(v,u|v_n) := t(v_n)\tfrac{\delta_n}{2}\|v-v_n\|^2 + \alpha_n t(v_n)(\varphi(v)-\varphi(v_n)) + \tfrac{1}{2}\|A_{\text{eq}}v - b_{\text{eq}}\|^2 + \tfrac{1}{2}\|A_{\text{ineq}}v - b_{\text{ineq}} + u\|^2 + \tfrac{\delta_n}{2}\|u\|^2.$$

When φ is smooth, one could use iterative methods to solve $\mathbb{F}_{\alpha,\delta}(v,u|v_n)$.

Theorem A.5 *Consider the following iterative procedure for finding the extremal point* $w^{**} = \begin{pmatrix} v^{**} \\ u^{**} \end{pmatrix}$*:*

$$w_n = \left[w_{n-1} - \gamma_n \frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n)\right]_+, \tag{A.17}$$

where

$$[z]_+ = \begin{cases} z & \text{if} \quad z \geq 0, \\ 0 & \text{if} \quad z < 0. \end{cases}$$

If

$$\sum_{n=0}^{\infty}\gamma_n\delta_n = \infty, \quad \frac{\gamma_n}{\delta_n} \underset{n\to\infty}{\to} 0, \quad \frac{|\alpha_n-\alpha_{n-1}|+|\delta_n-\delta_{n-1}|}{\gamma_n\delta_n} \underset{n\to\infty}{\to} 0, \tag{A.18}$$

then

$$\Xi_n := \|w_n - w_n^*\|^2 \underset{n\to\infty}{\to} 0. \tag{A.19}$$

Proof. From the iterative procedure given in Eq. (A.17) one has that

$$\begin{aligned}
\Xi_n = \left\|\left[w_{n-1} - \gamma_n\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n)\right]_+ - w_n^*\right\|^2 \leq \\
\left\|\left(w_{n-1}-w_{n-1}^*\right) - \gamma_n\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n) + \left(w_{n-1}^*-w_n^*\right)\right\|^2 = \\
\Xi_{n-1} + \gamma_n^2\left\|\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n)\right\|^2 + \left\|\left(w_{n-1}^*-w_n^*\right)\right\|^2 - \\
2\gamma_n\left(w_{n-1}-w_{n-1}^*\right)^{\intercal}\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n) + 2\left(w_{n-1}-w_{n-1}^*\right)^{\intercal}\left(w_{n-1}^*-w_n^*\right) - \\
2\gamma_n\left(w_{n-1}^*-w_n^*\right)^{\intercal}\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}(w_{n-1}|v_n).
\end{aligned} \tag{A.20}$$

By the inequalities (see the inequalities (21.17) and (21.36) in [74]) it can be concluded that

$$\left\|\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w_{n-1}|v_n\right)\right\|^2 =$$

$$\left\|\left[\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w_{n-1}|v_n\right)-\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w^*_{n-1}|v_n\right)\right]+\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w^*_{n-1}|v_n\right)\right\|^2 \leq$$

$$\left(1+\vartheta_n\right)\left\|\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w_{n-1}|v_n\right)-\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w^*_{n-1}|v_n\right)\right\|^2+$$

$$\left(1+\vartheta_n^{-1}\right)\left\|\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w^*_{n-1}|v_n\right)\right\|^2 \leq \left(1+\vartheta_n\right)L_\nabla\Xi_{n-1}+\left(1+\vartheta_n^{-1}\right)d,$$

where $\left\|\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w^*_{n-1}|v_n\right)\right\|^2 \leq d$, and

$$\left(w_{n-1}-w^*_{n-1}\right)^\intercal\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w_{n-1}|v_n\right) \geq l_n\Xi_{n-1},\ l_n=\left(\alpha_n\lambda^-+\delta_n\right)$$

$$\left|\left(w_{n-1}-w^*_{n-1}\right)^\intercal\left(w^*_{n-1}-w^*_n\right)\right| \leq \left\|w^*_{n-1}-w^*_n\right\|\sqrt{\Xi_{n-1}}$$

$$\left|\left(w^*_{n-1}-w^*_n\right)^\intercal\frac{\partial}{\partial w}\mathbb{F}_{\alpha_n,\delta_n}\left(w_{n-1}|v_n\right)\right| \overset{\vartheta>0}{\leq} \left\|w^*_{n-1}-w^*_n\right\|\sqrt{\left(1+\vartheta\right)L_\nabla\Xi_{n-1}+\left(1+\vartheta^{-1}\right)d} \leq$$

$$\left\|w^*_{n-1}-w^*_n\right\|\left[\left(1+\vartheta^{1/2}\right)\sqrt{L_\nabla}\sqrt{\Xi_{n-1}}+\left(1+\vartheta^{-1/2}\right)\sqrt{d}\right].$$

Then, from Eq. (A.20) it follows that

$$\Xi_n \leq \Xi_{n-1}+\gamma_n^2\left[\left(1+\vartheta\right)L_\nabla\Xi_{n-1}+\left(1+\vartheta^{-1}\right)d\right]+2\left(\kappa_1^2\left|\alpha_n-\alpha_{n-1}\right|^2+\kappa_2^2\left|\delta_n-\delta_{n-1}\right|^2\right)$$

$$-2\gamma_n\left(\alpha_n\lambda^-+\delta_n\right)\Xi_{n-1}+2\left(\kappa_1\left|\alpha_n-\alpha_{n-1}\right|+\kappa_2\left|\delta_n-\delta_{n-1}\right|\right)\sqrt{\Xi_{n-1}}+$$

$$2\gamma_n\left(\kappa_1\left|q_n-q_{n-1}\right|+\kappa_2\left|\delta_n-\delta_{n-1}\right|\right)\cdot\left[\left(1+\vartheta^{1/2}\right)\sqrt{L_\nabla}\sqrt{\Xi_{n-1}}+\left(1+\vartheta^{-1/2}\right)\sqrt{d}\right],$$

or, equivalently,

$$\Xi_n \leq \Xi_{n-1}\left(1-q_{n-1}\right)+\bar{\delta}_{n-1}\sqrt{\Xi_{n-1}}+\omega_{n-1},$$

where

$$q_{n-1} = 2\gamma_n\left(\alpha_n\lambda^- + \delta_n\right) - \gamma_n^2\left(1+\vartheta\right)L_\nabla = 2\gamma_n\left(\alpha_n\lambda^- + \delta_n\right)\left[1 - \frac{\gamma_n\left(1+\vartheta\right)L_\nabla}{2\left(\alpha_n\lambda^- + \delta_n\right)}\right] \geq$$
$$\gamma_n\delta_n 2\left(1+o\left(1\right)\right)\left[1 - \frac{\gamma_n\left(1+\vartheta\right)L_\nabla}{2\delta_n\left(o\left(1\right)+1\right)}\right] \geq \kappa_q\gamma_n\delta_n,$$

$$\bar{\delta}_{n-1} = 2\left(\kappa_1\left|\alpha_n - \alpha_{n-1}\right| + \kappa_2\left|\delta_n - \delta_{n-1}\right|\right)\cdot\left[1 + \gamma_n\left(1+\vartheta^{1/2}\right)\sqrt{L_\nabla}\right]$$
$$\leq \kappa_\delta\left(\left|\alpha_n - \alpha_{n-1}\right| + \left|\delta_n - \delta_{n-1}\right|\right),$$

$$\omega_{n-1} = \gamma_n^2\left(1+\vartheta^{-1}\right)d + \left(\kappa_1^2\left|\alpha_n - \alpha_{n-1}\right|^2 + \kappa_2^2\left|\delta_n - \delta_{n-1}\right|^2\right)$$
$$+ 2\gamma_n\left(\kappa_1\left|\alpha_n - \alpha_{n-1}\right| + \kappa_2\left|\delta_n - \delta_{n-1}\right|\right)\left(1+\vartheta^{-1/2}\right)\sqrt{d} \leq$$
$$\gamma_n^2\kappa_{\omega,1} + \gamma_n\left(\left|\alpha_n - \alpha_{n-1}\right| + \left|\delta_n - \delta_{n-1}\right|\right)\kappa_{\omega,2} + \left(\left|\alpha_n - \alpha_{n-1}\right|^2 + \left|\delta_n - \delta_{n-1}\right|^2\right)\kappa_{\omega,3}.$$

Using the inequality

$$\Xi_n^r \leq \left(1-r\right)\theta_n^r + \frac{p}{\theta_n^{1-r}}\Xi_n, \quad p\in\left(0,1\right), \quad \theta_n > 0,$$

for $p = 1/2$ and $\sqrt{\theta_n} = \dfrac{\bar{\delta}_{n-1}}{2q_{n-1}\left(1-\eta\right)}$, $\eta \in (0,1)$, the inequality can be reduced to the following one

$$\begin{aligned}\Xi_n \leq \Xi_{n-1}\left(1 - q_{n-1}\left[1 - \frac{\bar{\delta}_{n-1}}{2q_{n-1}\sqrt{\theta_n}}\right]\right) + \left[\omega_{n-1} + \tfrac{1}{2}\bar{\delta}_{n-1}\sqrt{\theta_n}\right] = \\ \Xi_{n-1}\left(1 - q_{n-1}\eta\right) + \left[\omega_{n-1} + \frac{\bar{\delta}_{n-1}^2}{4\left(1-\eta\right)q_{n-1}}\right].\end{aligned} \tag{A.21}$$

By Theorem 16.14 in [74] $\Xi_n \underset{n\to\infty}{\to} 0$ if

$$\sum_{n=0}^{\infty} q_n = \infty, \quad \frac{\omega_{n-1}}{q_{n-1}} + \frac{\bar{\delta}_{n-1}^2}{q_{n-1}^2} \underset{n\to\infty}{\to} 0,$$

which is equivalent to Eq. (A.19). Theorem is proven. ■

A.5 Rate of convergence

Select the parameters of the algorithm (A.17) as follows:

$$\delta_n = \begin{cases} \delta_0 & \text{if} \quad n \le n_0 \\ \delta_0 \frac{[1+\ln(n-n_0)]}{(1+n-n_0)^{\delta}} & \text{if} \quad n > n_0 \end{cases}, \quad \alpha_n = \begin{cases} \alpha_0 & \text{if} \quad n < n_0 \\ \frac{\alpha_0}{(1+n-n_0)^{\alpha}} & \text{if} \quad n \ge n_0 \end{cases},$$

$$\gamma_n = \begin{cases} \gamma_0 & \text{if} \quad n < n_0 \\ \frac{\gamma_0}{(1+n-n_0)^{\gamma}} & \text{if} \quad n \ge n_0 \end{cases}, \qquad \delta,\ \alpha,\ \gamma > 0, \quad \delta_0,\ \alpha_0,\ \gamma_0 > 0, \tag{A.22}$$

To guarantee the convergence of the suggested procedure, by the property $\frac{\alpha_n}{\delta_n} \underset{n\to\infty}{\to} 0$ and by the conditions (A.18), the parameters of the algorithm should satisfy that

$$\delta \le \alpha, \quad \gamma \ge \delta, \quad \gamma + \delta \le 1. \tag{A.23}$$

Lemma A.6 *Suppose that for a nonnegative sequence* $\{s_n\}$ *the following recurrent inequality holds*

$$s_n \le s_{n-1}(1 - q_n) + \omega_n,$$

where numerical sequences $\{q_n\}$ *and* $\{\omega_n\}$ *satisfies*

$$q_n \in (0, 1], \quad \omega_n \ge 0, \quad \upsilon_n > 0 \ \ \textit{for all } n = 0, 1...$$

$$\sum_{n=0}^{\infty} q_n = \infty, \quad \sum_{n=0}^{\infty} \omega_n \upsilon_n < \infty, \quad \lim_{n\to\infty} \frac{\upsilon_n - \upsilon_{n-1}}{q_n \upsilon_n} := \theta < 1.$$

Then

$$s_n = o\left(\upsilon_n^{-1}\right). \tag{A.24}$$

Proof. For $\tilde{s}_n = \upsilon_n s_n$ it follows

$$\tilde{s}_n \le \tilde{s}_{n-1}(1 - q_n)\upsilon_n \upsilon_{n-1}^{-1} + \upsilon_n \omega_n = \tilde{s}_{n-1}[1 - q_n(1 - \theta + o(1))] + \upsilon_n \omega_n,$$

which by the same Theorem 16.14 in [74] implies Eq. (A.24). ■

Theorem A.7 *If the proximal function (*A.*4) is strictly convex and the sequence* υ_n *of the proximal function (*A.*4) converges, then the optimal parameters are given by*

$$\gamma = \gamma^* = \frac{1}{2}, \quad \delta = \delta^* = \frac{1}{2}, \quad \alpha^* = \frac{3}{4}, \quad \xi^* = 1.$$

Proof. By (A.21) and (A.22) one has

$$q_n = O\left(\frac{1}{n^{\gamma+\delta}}\right),$$

$$\left[\omega_{n-1} + \frac{\bar{\delta}_{n-1}^2}{4\left(1-\eta\right)\alpha_{n-1}}\right] = O\left(\frac{1}{n^{2\gamma}}\right) + O\left(\frac{1}{n^{2(\alpha+1)}}\right) + O\left(\frac{1}{n^{2(\delta+1)}}\right) + O\left(\frac{1}{n^{\gamma+\alpha+1}}\right) + O\left(\frac{1}{n^{\gamma+\delta+1}}\right) O\left(\frac{1}{n^{2(\alpha+1)-\gamma-\delta}}\right) + O\left(\frac{1}{n^{2(\delta+1)-\gamma-\delta}}\right) = O\left(\frac{1}{n^{2\gamma}}\right) + O\left(\frac{1}{n^{\gamma+\delta+1}}\right) + O\left(\frac{1}{n^{\delta+2-\gamma}}\right).$$

As a result,

$$\Xi_n \leq \Xi_{n-1}\left[1 - \left|O\left(\frac{1}{n^{\gamma+\delta}}\right)\right|\right] + O\left(\frac{1}{n^{2\gamma}}\right) + O\left(\frac{1}{n^{\gamma+\delta+1}\left[1+\ln\left(n-n_0\right)\right]}\right) + O\left(\frac{1}{n^{2(\delta+1)-\gamma-\delta}\left[1+\ln\left(n-n_0\right)\right]}\right),$$

and for $\upsilon_n = n^v$ it follows that $\Xi_n = O\left(n^{-v}\right)$, if $v \in (0,1]$ satisfies

$$\gamma + \delta \leq 1, \;\; v \leq 2\gamma, \;\; v \leq \delta + 2 - \gamma, \tag{A.25}$$

or, equivalently, $0 < v \leq \min\left\{2\gamma, \delta+2-\gamma\right\}$.

So, the rate of convergence for $\tilde{\Xi}_n := \left\|w_n - w^{**}\right\|^2$ will be estimated by the following relation

$$\tilde{\Xi}_n = \left\|\left(w_n - w_n^*\right) + \left(w_n^* - w^{**}\right)\right\|^2 \leq 2\Xi_n + 2\left\|\left(w_n^{**} - w^{**}\right)\right\|^2 = 2\Xi_n + O\left(\frac{\alpha_n^2}{\delta_n}\right) = o(n^{-v}) + O\left(\frac{\alpha_n^2}{\delta_n}\right) \underset{n\to\infty}{\to} 0,$$

which leads to the following conclusion: the best rate $n^{-\xi^*}$ of the convergence $\tilde{\Xi}_n$ to zero is defined as

$$\tilde{\Xi}_n = O\left(n^{-\xi}\right),$$

where $\xi^* = \max\min\left\{v,\; 2\alpha - \delta\right\} = \max\min\left\{2\gamma, \delta+2-\gamma, 2\alpha-\delta, 1\right\}$.

Since $\delta + 2 - \gamma \geq \delta + 2 - (1-\delta) = 2\delta + 1 > 1$, it follows that

$$\min\left\{2\gamma, \delta+2-\gamma, 2\alpha-\delta, 1\right\} = \min\left\{2\gamma, 2\alpha-\delta, 1\right\}.$$

Under constrains (A.23) and (A.25) the maximal upper estimate is achieved when $2\gamma = 2\alpha - \delta = 1$, implying $\gamma = \frac{1}{2}$, $\delta = 2\alpha - 1 \leq 1 - \gamma = \frac{1}{2}$ and $2\alpha \leq \frac{3}{2}$. Finally one obtains

$$\gamma = \gamma^* = \frac{1}{2}, \quad \delta = \delta^* = \frac{1}{2}, \quad \alpha^* = \frac{3}{4}, \quad \xi^* = 1.$$

■

A.6 Production planning example

Consider the Continuous-Time Markov Chains theory presented in Chapter 2. Then, the *joint strategy variable* $c_{(i,k)}$ which belongs to the set of matrices $c \in C_{\text{adm}}$ is restricted by Eqs. (2.7, 2.8 and 2.9). Introducing

$$\|(v - v^*)\|^2_{\Lambda=\text{diag}(\Lambda_1,...,\Lambda_M)} = \sum_{k=1}^{M} \left\| \left(v_{(k)} - v^*_{(k)}\right) \right\|^2 = \sum_{k=1}^{M} \left(v_{(k)} - v^*_{(k)}\right)^{\intercal} \Lambda_k \left(v_{(k)} - v^*_{(k)}\right),$$

where

$$v = \left(v_{(1)}, ..., v_{(M)}\right)^{\intercal} \in \mathbb{R}^{N \times M}, \quad v_{(k)} = \left(c_{(1,k)}, ..., c_{(N,k)}\right)^{\intercal} \in \mathbb{R}^{N},$$

for $k = 1, ..., M$ and

$$\Lambda_k := \tfrac{1}{2}\left[\tilde{\Lambda}_k + \tilde{\Lambda}_k^{\intercal}\right], \quad \tilde{\Lambda}_k := \left[\tau_{(j|i,k)}\right], \quad \tilde{\Lambda}_k \in \mathbb{R}^{N \times N},$$

$$\tau_{(j|i,k)} := \begin{cases} \dfrac{1}{\left|\sum\limits_{i \neq j}^{N} q_{(j|i,k)}\right|} & \text{if } i = j, \\ \dfrac{1}{q_{(j|i,k)}} & \text{if } i \neq j. \end{cases}$$

Then, one has that

$$v^* = \arg\min_{v \in V} \left\{ \frac{\delta_n}{2} \|(v - v^*)\|^2_{\Lambda=\text{diag}(\Lambda_1,...,\Lambda_M)} + \gamma_n \left(F_{\alpha,\delta}\left(v, u | v_n\right)\right) \right\}.$$

The production planning and scheduling are very important processes that directly influence the success of production companies, this models are usually formulated as optimization problems subject to uncertainties derived from events as fluctuation of demand, equipment failures, quantity of surplus production, among other factors. There is a growing interest in

applying these models in manufacturing and remanufacturing systems in different industries or companies (see for example, [52, 47, 26, 40, 48]).

Consider a Production Planning Model where the state variable is taken as the surplus amount of the production system that is determined by both demand rate and production rate, which in turn is governed by the production capacity [109].

The manufacturing system produces M different products. The system is given by a differential equation, which states that the rates of change of the surplus, the inventory/shortage level $x(t) \in \mathbb{R}^M$, constitute the difference between the production rates $v(t) \in \mathbb{R}^M$ which depends on the random machine capacity, and the random demand rates $z(t) \in \mathbb{R}^M$ (see Figure A.1). The objective is to find the optimal production rate v^* to minimize the cost function subject to the system dynamics, the machine capacity $y(t)$, and other operating conditions.

The usefulness of implementing the optimization method presented in this article is that with the time penalization, the losses that the industry/company has in the manufacturing process due to different factors, for example the continuous deterioration of the machines, can be modeled.

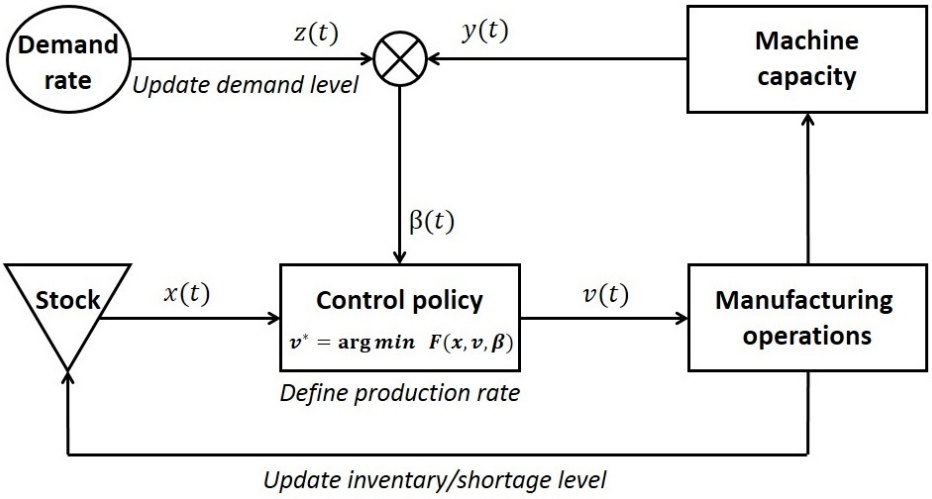


Figure A.1 A schematic representation of a manufacturing system.

Let define demand process as a finite-state Markov chain $z = \{z(t) : t \geq 0\}$ having state space $Z = \{z_1, ..., z_{N_2}\}$. Considering the possible random breakdown and repair, the machine capacity is modeled by a continuous-time, finite-state Markov chain $y = \{y(t) : t \geq 0\}$ with

state space $Y = \{y_1, ..., y_{N_1}\}$. At any given time the production capacity determines the set of all possible rates of production $v(t)$. For each state of the capacity, $y_1 \leq y_i \leq y_{N_1}$, the production rate $u_k = (u_1, u_2, ..., u_M) \geq 0$ must satisfy the constraint

$$p \cdot u' \leq \alpha_i, \qquad i = 0, ..., N_1,$$

where $p = (p_1, ..., p_M) \geq 0$ is a given constant vector with each p_k $(k = 1, ..., M)$ representing the amount of capacity needed to produce one unit of product k.

The generators of the Markov chains y and z following an action denoted by k (the decision of generate a type of product k) are given by $Q^y = \left[q^y_{(j|i,k)}\right]_{i,j=\overline{1,N_1},k=\overline{1,M}}$ and $Q^z = \left[q^z_{(j|i,k)}\right]_{i,j=\overline{1,N_2},k=\overline{1,M}}$, respectively.

Let there be a manufacturing model with two machine producing three different products, because of this, the machines can follow three actions $k = \overline{1, M}$, $M = 3$, this means to generate product k. Suppose that one only has flexible machines which require no setup-time consumption when switching from production of one type of product to the production of another. In this example, consider the demand process as a two-states Markov chain, $z(t) \in Z = \{z_1, z_2\}$, $N_2 = 2$, that means that z_1 is a low level and z_2 is a high level of demand. The generator for each action k of the demand process is as follows

$$Q^z_{(j|i,1)} = \begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix} \qquad Q^z_{(j|i,2)} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \qquad Q^z_{(j|i,3)} = \begin{bmatrix} -5 & 5 \\ 3 & -3 \end{bmatrix}$$

Consider that the two parallel machines are subject to breakdown and repair. If the machine is up, then it can produce parts with production rate $v(t)$ and its production rate is zero if the machine is under repair, so each having capacities $y^1 \in \{0, y_1\}$ and $y^2 \in \{0, y_2\}$, the overall state space of the four-state Markov chain capacity is $Y = \{(0,0), (y_1, 0), (0, y_2), (y_1, y_2)\}$ which contains all possible combinations $N_1 = 4$ between the states of the two machines. For simplicity, suppose each of the machines is either in operating condition (denoted by 1) or under repair (denoted by 0), then it follows that $Y = \{(0,0), (1,0), (0,1), (1,1)\}$. Let the rates of each machine going down be λ_1 and λ_2, and the rates of resumption be μ_1 and μ_2. The

generators for any action (product) k of each machine capacity are as follows

$$Q^1_{(j|i,k)} = \begin{bmatrix} -\mu_1 & \mu_1 & 0 & 0 \\ \lambda_1 & -\lambda_1 & 0 & 0 \\ 0 & 0 & -\mu_1 & \mu_1 \\ 0 & 0 & \lambda_1 & -\lambda_1 \end{bmatrix} \qquad Q^2_{(j|i,k)} = \begin{bmatrix} -\mu_2 & 0 & \mu_2 & 0 \\ 0 & -\mu_2 & 0 & \mu_2 \\ \lambda_2 & 0 & -\lambda_2 & 0 \\ 0 & \lambda_2 & 0 & -\lambda_2 \end{bmatrix}$$

To determine the generator for the overall capacity process, it is satisfied that only one machine may change its state during a single transition. Therefore

$$Q^y_{(j|i,k)} = \begin{bmatrix} -(\mu_1+\mu_2) & \mu_1 & \mu_2 & 0 \\ \lambda_1 & -(\lambda_1+\mu_2) & 0 & \mu_2 \\ \lambda_2 & 0 & -(\lambda_2+\mu_1) & \mu_1 \\ 0 & \lambda_2 & \lambda_1 & -(\lambda_1+\lambda_2) \end{bmatrix}$$

In this example, consider $\lambda_1 = 3.53$, $\lambda_2 = 4.8$, $\mu_1 = 120$ and $\mu_2 = 120$, then the matrix Q^y for any action k gives

$$Q^y_{(j|i,k)} = \begin{bmatrix} -240 & 120 & 120 & 0 \\ 3.53 & -123.53 & 0 & 120 \\ 4.8 & 0 & -124.8 & 120 \\ 0 & 4.8 & 3.53 & -8.33 \end{bmatrix}$$

Then, the production system is subject to a joint stochastic process, $\beta(t) = (y(t), z(t))$ consisting of the capacity and demand pair. Observe that β is also a Markov chain that has a state space of size $N = N_1 \times N_2$

$$B = \{(y_1, z_1), ..., (y_{N_1}, z_1), ..., (y_1, z_{N_2}), ..., (y_{N_1}, z_{N_2})\}.$$

In this example the total number of states is $N = N_1 \times N_2 = 8$

$$B = \{(0,0,z_1), (1,0,z_1), (0,1,z_1), (1,1,z_1), (0,0,z_2), (1,0,z_2), (0,1,z_2), (1,1,z_2)\},$$

that is, all possible combination between the states of the demand process and the states of the capacity process.

In many manufacturing processes, the machine capacity (breakdown and repair) take place much more frequently than the changes in demand. To reflect the differences in transition rates between the matrices Q^y and Q^z, i.e., the weak and strong interactions of the systems, a timescale separation by introducing a small parameter $\varepsilon > 0$ into the system [108] is implemented. The generator of the chain $\beta(t)$ is of the form $Q = \left[q_{(j|i,k)}\right]_{i,j=\overline{1,N},k=\overline{1,M}}$ where for action $k = \overline{1, M}, M = 3$ and a scale factor $\varepsilon = 0.01$ is given by

$$Q = \frac{1}{\varepsilon}\tilde{Q} + \hat{Q} = \frac{1}{\varepsilon}\begin{bmatrix} Q^y & & \\ & \ddots & \\ & & Q^y \end{bmatrix} + \begin{bmatrix} q^z_{(1,1)}I_{N_1} & q^z_{(1,2)}I_{N_1} & \cdots & q^z_{(1,N_2)}I_{N_1} \\ \vdots & \vdots & \ddots & \vdots \\ q^z_{(N_2,1)}I_{N_1} & q^z_{(N_2,2)}I_{N_1} & \cdots & q^z_{(N_2,N_2)}I_{N_1} \end{bmatrix},$$

where $\tilde{Q} = \text{diag}(Q^y, ..., Q^y)$ is a block-diagonal matrix representing the fast motion and the Kronecker product $\hat{Q} = Q^z \otimes I_{N_1}$ governs the slow varying part.

The dynamic system of the manufacturing process is given by

$$\begin{aligned} \dot{x}(t) &= y(t)v(t) - z(t), \\ x(0) &= x, \end{aligned} \tag{A.26}$$

where $\beta = (y, z)$ is the initial state of the Markov chain and $x \in \mathbb{R}^M$ is the initial surplus level that is positive when it represents inventory and negative when it represents shortage. Define the cost functional as

$$F\left(x(t), v(t), \beta(t)\right) = E\int_0^\infty e^{-\rho t} G\left(x(t), v(t), \beta(t)\right) dt, \tag{A.27}$$

where $G\left(x(t), v(t), \beta(t)\right)$ is the running cost of having surplus $x(t)$, a normalized production rate $v(t)$, a Markov chain $\beta\left(y(t), z(t)\right)$ and a discount rate $\rho > 0$ (old data have less impact into the overall cost). The goal is to find the optimal policy or the optimal production rate $v^*(t) \in \mathbb{R}^{N\times M}$, to minimize the objective function (Eq. A.27), subject to dynamics described by Eq. A.26, the capacity $y(t)$, and other production constraints for the given initial conditions.

Let the utility matrix $u_{(j|i,k)}$ of the production process that depends on the transition between the states of the Markov chain $\beta(y, z)$ and the product being manufactured be as follows:

$$u_{(j|i,1)} = \begin{bmatrix} 19 & 10 & 18 & 11 & 18 & 11 & 10 & 6 \\ 15 & 19 & 3 & 17 & 18 & 3 & 15 & 8 \\ 12 & 1 & 9 & 17 & 5 & 3 & 7 & 18 \\ 20 & 4 & 6 & 17 & 2 & 18 & 9 & 9 \\ 17 & 10 & 20 & 5 & 10 & 13 & 11 & 5 \\ 20 & 11 & 13 & 11 & 17 & 6 & 18 & 12 \\ 13 & 2 & 6 & 18 & 10 & 18 & 16 & 10 \\ 8 & 14 & 3 & 3 & 9 & 2 & 2 & 18 \end{bmatrix}$$

$$u_{(j|i,2)} = \begin{bmatrix} 19 & 17 & 2 & 18 & 14 & 2 & 10 & 20 \\ 17 & 11 & 9 & 7 & 12 & 15 & 10 & 5 \\ 15 & 8 & 3 & 1 & 13 & 7 & 19 & 16 \\ 15 & 5 & 9 & 3 & 18 & 14 & 8 & 12 \\ 18 & 11 & 18 & 2 & 14 & 14 & 3 & 9 \\ 2 & 6 & 8 & 15 & 18 & 11 & 5 & 11 \\ 7 & 2 & 3 & 10 & 10 & 15 & 14 & 10 \\ 1 & 2 & 12 & 14 & 3 & 11 & 17 & 14 \end{bmatrix}$$

$$u_{(j|i,3)} = \begin{bmatrix} 15 & 19 & 18 & 1 & 20 & 15 & 2 & 4 \\ 7 & 18 & 2 & 8 & 11 & 11 & 14 & 1 \\ 2 & 20 & 20 & 19 & 20 & 17 & 13 & 9 \\ 17 & 18 & 19 & 11 & 10 & 11 & 5 & 10 \\ 8 & 1 & 12 & 10 & 9 & 12 & 9 & 4 \\ 16 & 11 & 9 & 10 & 5 & 5 & 13 & 14 \\ 4 & 20 & 7 & 7 & 4 & 12 & 12 & 1 \\ 3 & 11 & 15 & 20 & 17 & 3 & 3 & 3 \end{bmatrix}$$

The cost matrix $r_{(i,k)}$ for each state and action, that depends on the utility matrix $u_{(j|i,k)}$ and the transition matrix $\pi_{(j|i,k)}$ that represent the behavior of the Markov chain $\beta(y, z)$, is defined as follows

$$r_{ik} = \sum_{j=1}^{N} u_{j|ik} \pi_{j|ik},$$

then, the production cost function $J(v, \beta)$ of the manufacturing process is given by

$$J\left(v(t), \beta(t)\right) = \sum_{i=1}^{N} \sum_{k=1}^{M} r_{(i,k)} v_{(i,k)}.$$

For the integral cost function $G\left(x, v, \beta\right)$, it is also considered the holding cost, which are the costs associated with storing and maintaining a piece of inventory that remains unsold over the course of time and that depends only of the total surplus for product k

$$h\left(x(t)\right) = \sum_{k=1}^{M} \left(0.01 x_k^+ + 0.7 x_k^-\right),$$

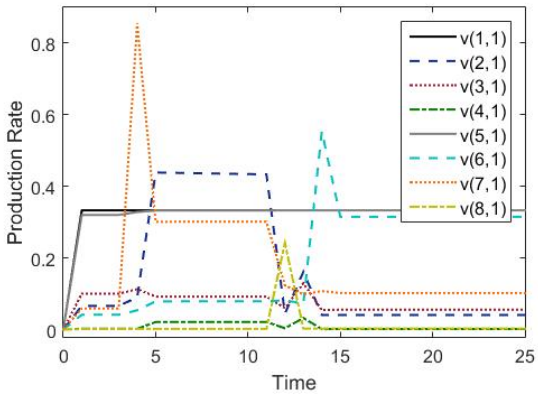
where $x_k^+ = \max\{0, x_k\}$ and $x_k^- = \max\{0, -x_k\}$. Finally, the overall cost function for the manufacturing model is $G\left(x(t), v(t), \beta(t)\right) = h(x(t)) + J\left(v(t), \beta(t)\right)$. Applying the proposed optimization method with given initial values of surplus $x = (45, -15, 5)$ and following Eq. (2.6), the optimal values v^* are calculated. Figure A.2 shows the convergence of the production rate v^* for each state and product $k = \overline{1, M}, M = 3$.

Once the method converges, the optimal production rate v^* is as follows

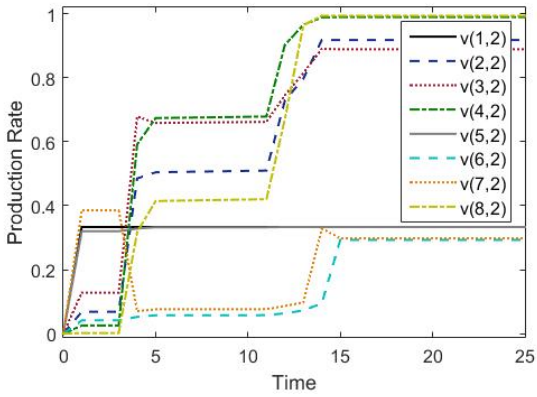
$$v^*_{(i,k)} = \begin{bmatrix} 0.3333 & 0.3333 & 0.3333 \\ 0.0409 & 0.9180 & 0.0411 \\ 0.0558 & 0.8892 & 0.0550 \\ 0.0019 & 0.9896 & 0.0085 \\ 0.3333 & 0.3333 & 0.3333 \\ 0.3152 & 0.2932 & 0.3917 \\ 0.1023 & 0.2979 & 0.5998 \\ 0.0032 & 0.9937 & 0.0032 \end{bmatrix}$$

For example, for state 3 (this means that only one machine is in operating condition and the rate of demand is low for all products) one has that in a working day the 0.0558 is dedicated to manufacture product $k = 1$, 0.8892 to product $k = 2$ and 0.0550 to product $k = 3$, these production rates are due to the fact that we have a shortage level of product 2 ($x_2 = -15$); and in general, for all states in the production rate matrix, it is observed that there is a greater emphasis on compensating the shortage of the product 2.

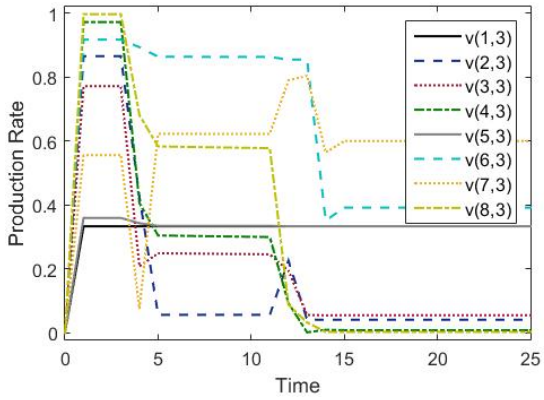
Finally, Figure A.3 shows the behavior of the objective cost function, which with the use of the presented method, converges in a lower cost than the initial one.



(a) Convergence for product $k = 1$.



(b) Convergence for product $k = 2$.



(c) Convergence for product $k = 3$.

Figure A.2 Convergence of the production rate v_{ik}.

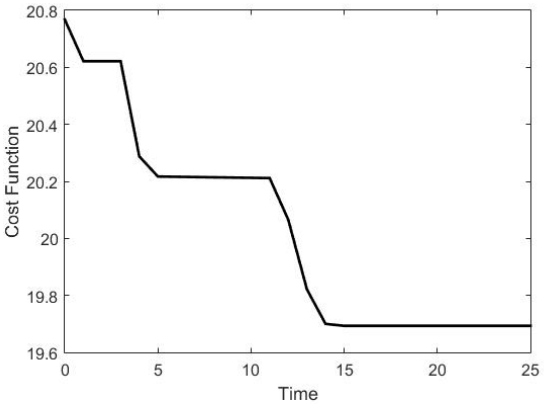


Figure A.3 Convergence of the cost function.

Appendix B
The Nash vs. Kalai-Smorodinsky solution

This Chapter presents a numerical example in order to appreciate the difference between the solution presented by Nash and the one presented by Kalai and Smorodinsky [95].

Consider a two-person bargaining problem in a class of continuous time controllable Markov chains. Let us denote the disagreement cost that depends on the strategies $c^l_{(i,k)}$ $(l = 1, 2)$ for players 1 and 2 as $\phi^1(c^1, c^2)$ and $\phi^2(c^1, c^2)$ respectively, and the solution for the bargaining problem as the point (ψ^1, ψ^2).

Let the states $N = 6$, and the number of actions $M = 3$. The individual utility for each player are defined by

$$U^1_{(i,j|1)} = \begin{bmatrix} 34 & 45 & 1 & 28 & 7 & 23 \\ 27 & 43 & 25 & 47 & 26 & 24 \\ 15 & 45 & 14 & 15 & 43 & 48 \\ 36 & 47 & 12 & 17 & 20 & 5 \\ 20 & 41 & 22 & 43 & 35 & 14 \\ 29 & 29 & 18 & 18 & 32 & 23 \end{bmatrix} \quad U^2_{(i,j|1)} = \begin{bmatrix} 31 & 1 & 30 & 38 & 2 & 17 \\ 18 & 41 & 10 & 13 & 42 & 11 \\ 5 & 8 & 34 & 33 & 12 & 31 \\ 2 & 44 & 13 & 43 & 3 & 40 \\ 25 & 5 & 22 & 5 & 28 & 10 \\ 13 & 18 & 7 & 29 & 48 & 3 \end{bmatrix}$$

$$U^1_{(i,j|2)} = \begin{bmatrix} 30 & 44 & 14 & 47 & 25 & 31 \\ 44 & 24 & 45 & 37 & 11 & 30 \\ 24 & 25 & 12 & 20 & 32 & 22 \\ 22 & 25 & 44 & 50 & 12 & 33 \\ 38 & 12 & 36 & 33 & 27 & 22 \\ 24 & 5 & 44 & 45 & 37 & 1 \end{bmatrix} \quad U^2_{(i,j|2)} = \begin{bmatrix} 15 & 15 & 43 & 9 & 18 & 14 \\ 13 & 13 & 2 & 36 & 32 & 30 \\ 25 & 25 & 15 & 42 & 18 & 22 \\ 39 & 23 & 45 & 2 & 11 & 5 \\ 18 & 41 & 27 & 38 & 40 & 2 \\ 29 & 5 & 7 & 18 & 17 & 25 \end{bmatrix}$$

$$U^1_{(i,j|3)}=\begin{bmatrix} 28 & 27 & 48 & 8 & 16 & 27 \\ 43 & 47 & 33 & 24 & 22 & 28 \\ 21 & 37 & 19 & 28 & 15 & 42 \\ 24 & 29 & 24 & 3 & 50 & 42 \\ 42 & 49 & 46 & 33 & 31 & 42 \\ 50 & 42 & 51 & 45 & 13 & 11 \end{bmatrix} \quad U^2_{(i,j|3)}=\begin{bmatrix} 14 & 11 & 31 & 48 & 50 & 11 \\ 17 & 34 & 14 & 39 & 39 & 20 \\ 15 & 23 & 28 & 31 & 24 & 2 \\ 9 & 22 & 48 & 48 & 35 & 24 \\ 20 & 9 & 36 & 3 & 21 & 17 \\ 35 & 10 & 34 & 14 & 20 & 49 \end{bmatrix}$$

The transition rate matrices for each player are defined as follows

$$q^1_{(i,j|1)}=\begin{bmatrix} -0.5371 & 0.0444 & 0.2305 & 0.0946 & 0.0705 & 0.0970 \\ 0.0208 & -0.5381 & 0.0294 & 0.0665 & 0.0471 & 0.3743 \\ 0.1179 & 0.0965 & -0.6554 & 0.0939 & 0.1042 & 0.2429 \\ 0.1871 & 0.0965 & 0.1622 & -0.5826 & 0.0285 & 0.1083 \\ 0.0825 & 0.1871 & 0.0671 & 0.0431 & -0.4624 & 0.0827 \\ 0.0831 & 0.1685 & 0.1221 & 0.3425 & 0.0432 & -0.7593 \end{bmatrix}$$

$$q^1_{(i,j|2)}=\begin{bmatrix} -1.6112 & 0.1333 & 0.6916 & 0.2839 & 0.2114 & 0.2911 \\ 0.0624 & -1.6142 & 0.0881 & 0.1996 & 0.1412 & 1.1228 \\ 0.3538 & 0.2894 & -1.9662 & 0.2817 & 0.3127 & 0.7287 \\ 0.5614 & 0.2894 & 0.4867 & -1.7477 & 0.0855 & 0.3248 \\ 0.2474 & 0.5614 & 0.2012 & 0.1292 & -1.3873 & 0.2482 \\ 0.2492 & 0.5055 & 0.3662 & 1.0275 & 0.1295 & -2.2780 \end{bmatrix}$$

$$q^1_{(i,j|3)}=\begin{bmatrix} -0.5371 & 0.0444 & 0.2305 & 0.0946 & 0.0705 & 0.0970 \\ 0.0208 & -0.5381 & 0.0294 & 0.0665 & 0.0471 & 0.3743 \\ 0.1179 & 0.0965 & -0.6554 & 0.0939 & 0.1042 & 0.2429 \\ 0.1871 & 0.0965 & 0.1622 & -0.5826 & 0.0285 & 0.1083 \\ 0.0825 & 0.1871 & 0.0671 & 0.0431 & -0.4624 & 0.0827 \\ 0.0831 & 0.1685 & 0.1221 & 0.3425 & 0.0432 & -0.7593 \end{bmatrix}$$

$$q^2_{(i,j|1)}=\begin{bmatrix} -0.8499 & 0.2201 & 0.3707 & 0.1271 & 0.0374 & 0.0947 \\ 0.3467 & -0.6729 & 0.1271 & 0.0376 & 0.0970 & 0.0644 \\ 0.2831 & 0.0856 & -0.6306 & 0.0706 & 0.0376 & 0.1537 \\ 0.0703 & 0.1577 & 0.1369 & -0.8573 & 0.3673 & 0.1250 \\ 0.3727 & 0.0964 & 0.0944 & 0.1298 & -0.8026 & 0.1092 \\ 0.1627 & 0.1095 & 0.1237 & 0.0754 & 0.4537 & -0.9250 \end{bmatrix}$$

$$q^2_{(i,j|2)}=\begin{bmatrix} -0.8499 & 0.2201 & 0.3707 & 0.1271 & 0.0374 & 0.0947 \\ 0.3467 & -0.6729 & 0.1271 & 0.0376 & 0.0970 & 0.0644 \\ 0.2831 & 0.0856 & -0.6306 & 0.0706 & 0.0376 & 0.1537 \\ 0.0703 & 0.1577 & 0.1369 & -0.8573 & 0.3673 & 0.1250 \\ 0.3727 & 0.0964 & 0.0944 & 0.1298 & -0.8026 & 0.1092 \\ 0.1627 & 0.1095 & 0.1237 & 0.0754 & 0.4537 & -0.9250 \end{bmatrix}$$

$$q^2_{(i,j|3)}=\begin{bmatrix} -1.1332 & 0.2934 & 0.4942 & 0.1694 & 0.0498 & 0.1263 \\ 0.4623 & -0.8972 & 0.1694 & 0.0502 & 0.1294 & 0.0859 \\ 0.3774 & 0.1141 & -0.8408 & 0.0942 & 0.0501 & 0.2050 \\ 0.0938 & 0.2102 & 0.1825 & -1.1431 & 0.4898 & 0.1667 \\ 0.4970 & 0.1286 & 0.1258 & 0.1730 & -1.0701 & 0.1456 \\ 0.2169 & 0.1460 & 0.1650 & 0.1005 & 0.6049 & -1.2334 \end{bmatrix}$$

The process to solve the bargaining problem consists of two main steps, firstly to find the disagreement point we define it as the Nash equilibrium point of the problem [63]; while for the solution of the bargaining process we follow the models presented by Nash and Kalai-Smorodinsky.

B.1 The disagreement point

Given δ and γ and applying the extraproximal method we obtain the convergence of the strategies for the disagreement point in terms of the variable $c^1_{(i,k)}$ for the player 1 (see Figure B.1) and the convergence of the strategies $c^2_{(i,k)}$ for the player 2 (see Figure B.2).

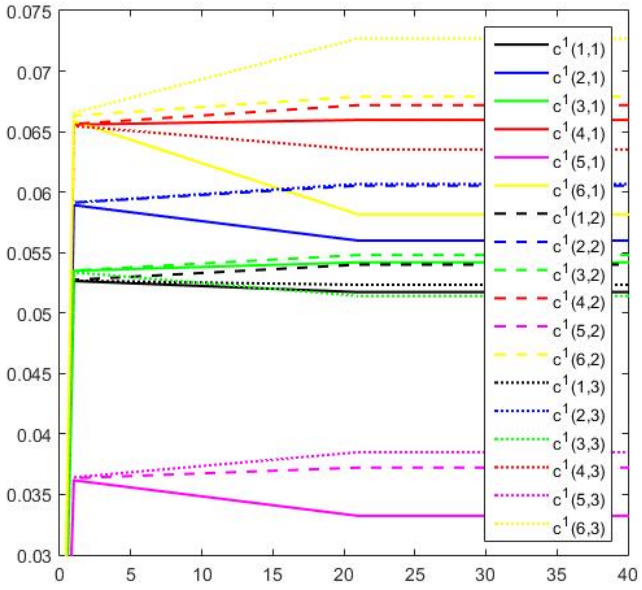


Figure B.1 Convergence of the strategies for player 1 in the disagreement point.

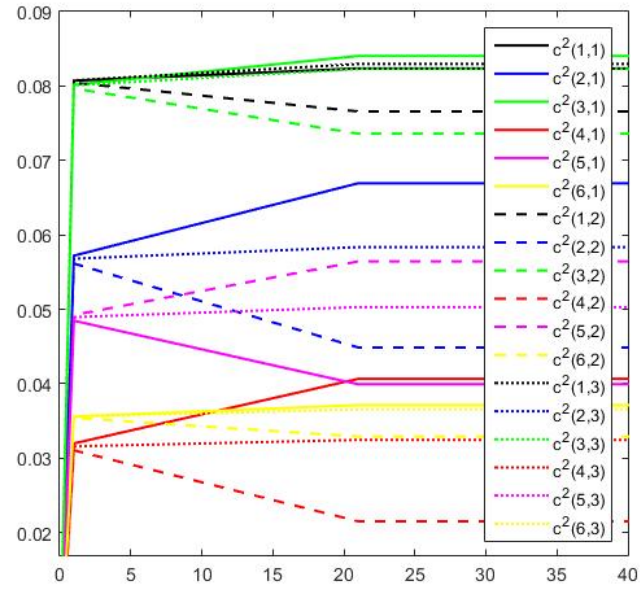


Figure B.2 Convergence of the strategies for player 2 in the disagreement point.

$$c^1 = \begin{bmatrix} 0.0517 & 0.0540 & 0.0523 \\ 0.0560 & 0.0605 & 0.0607 \\ 0.0542 & 0.0548 & 0.0514 \\ 0.0660 & 0.0672 & 0.0635 \\ 0.0332 & 0.0372 & 0.0385 \\ 0.0582 & 0.0679 & 0.0727 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.0824 & 0.0766 & 0.0830 \\ 0.0669 & 0.0449 & 0.0584 \\ 0.0840 & 0.0736 & 0.0823 \\ 0.0407 & 0.0215 & 0.0325 \\ 0.0399 & 0.0564 & 0.0503 \\ 0.0371 & 0.0329 & 0.0366 \end{bmatrix}$$

Following (2.6) the mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.3273 & 0.3416 & 0.3311 \\ 0.3160 & 0.3416 & 0.3424 \\ 0.3378 & 0.3416 & 0.3205 \\ 0.3354 & 0.3416 & 0.3230 \\ 0.3051 & 0.3416 & 0.3533 \\ 0.2926 & 0.3416 & 0.3658 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.3405 & 0.3166 & 0.3429 \\ 0.3933 & 0.2637 & 0.3429 \\ 0.3503 & 0.3068 & 0.3429 \\ 0.4295 & 0.2275 & 0.3429 \\ 0.2723 & 0.3847 & 0.3429 \\ 0.3484 & 0.3087 & 0.3429 \end{bmatrix}$$

With the strategies calculated, the resulting utilities following in the disagreement point for each player $\phi^l(c^1, c^2)$, are as follows:

$$\phi^1(c^1, c^2) = 905.6447 \qquad \phi^2(c^1, c^2) = 704.2493$$

B.2 The Nash bargaining solution

Given δ, γ, α^l and applying the extraproximal method for the Nash bargaining solution, we obtain the convergence of the strategies in terms of the variable $c^1_{(i,k)}$ for the player 1 (see Figure B.3) and the convergence of the strategies $c^2_{(i,k)}$ for the player 2 (see Figure B.4).

$$c^1 = \begin{bmatrix} 0.0281 & 0.0677 & 0.0623 \\ 0.0010 & 0.0758 & 0.1003 \\ 0.0907 & 0.0686 & 0.0010 \\ 0.1115 & 0.0842 & 0.0010 \\ 0.0010 & 0.0466 & 0.0613 \\ 0.0010 & 0.0851 & 0.1127 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.1227 & 0.0350 & 0.0842 \\ 0.1100 & 0.0010 & 0.0592 \\ 0.1555 & 0.0010 & 0.0835 \\ 0.0607 & 0.0010 & 0.0329 \\ 0.0010 & 0.0946 & 0.0510 \\ 0.0663 & 0.0032 & 0.0371 \end{bmatrix}$$

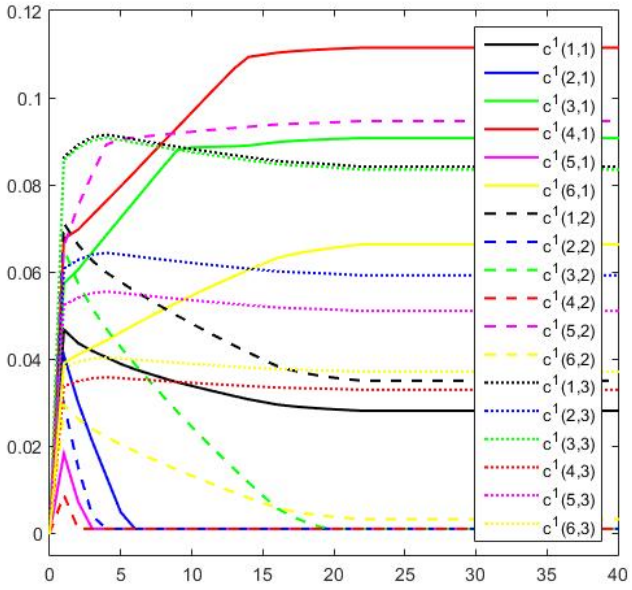


Figure B.3 Convergence of the strategies for player 1 in the Nash solution.

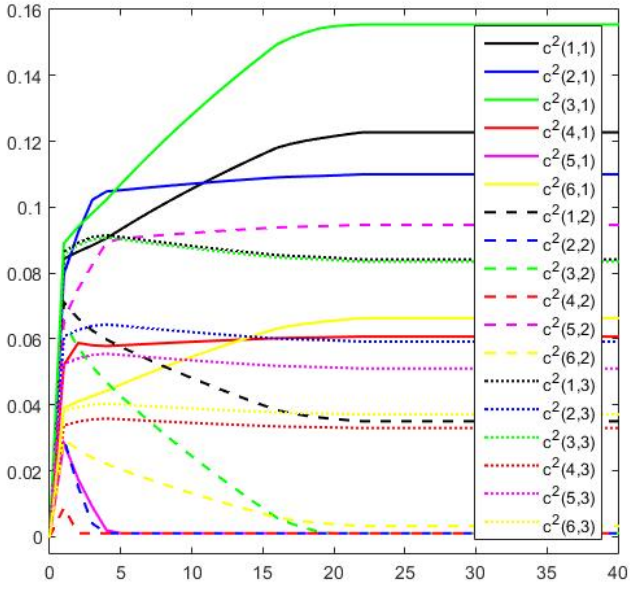


Figure B.4 Convergence of the strategies for payer 2 in the Nash solution.

The mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.1778 & 0.4280 & 0.3942 \\ 0.0056 & 0.4280 & 0.5663 \\ 0.5658 & 0.4280 & 0.0062 \\ 0.5669 & 0.4280 & 0.0051 \\ 0.0092 & 0.4280 & 0.5628 \\ 0.0050 & 0.4280 & 0.5670 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.5073 & 0.1447 & 0.3479 \\ 0.6462 & 0.0059 & 0.3479 \\ 0.6479 & 0.0042 & 0.3479 \\ 0.6415 & 0.0106 & 0.3479 \\ 0.0068 & 0.6453 & 0.3479 \\ 0.6221 & 0.0300 & 0.3479 \end{bmatrix}$$

With the strategies calculated, the resulting utilities in the Nash bargaining solution for each player, are as follows:

$$\psi^1(c^1, c^2) = 958.0281 \qquad \psi^2(c^1, c^2) = 813.2879$$

B.3 The Kalai-Smorodinsky bargaining solution

Given δ, γ, α^l and applying the extraproximal method for the Kalai-Smorodinsky bargaining solution with the L_1-norm, we obtain the convergence of the strategies in terms of the variable $c^1_{(i,k)}$ for the player 1 (see Figure B.5) and the convergence of the strategies $c^2_{(i,k)}$ for the player 2 (see Figure B.6).

$$c^1 = \begin{bmatrix} 0.0010 & 0.0432 & 0.1139 \\ 0.0010 & 0.0484 & 0.1278 \\ 0.1156 & 0.0438 & 0.0010 \\ 0.1420 & 0.0537 & 0.0010 \\ 0.0010 & 0.0297 & 0.0782 \\ 0.0010 & 0.0543 & 0.1435 \end{bmatrix} \qquad c^2 = \begin{bmatrix} 0.2061 & 0.0010 & 0.0349 \\ 0.1447 & 0.0010 & 0.0245 \\ 0.2044 & 0.0010 & 0.0346 \\ 0.0800 & 0.0010 & 0.0136 \\ 0.0010 & 0.1245 & 0.0211 \\ 0.0903 & 0.0010 & 0.0154 \end{bmatrix}$$

The mixed strategies obtained for the players are as follows

$$d^1 = \begin{bmatrix} 0.0063 & 0.2730 & 0.7207 \\ 0.0056 & 0.2730 & 0.7213 \\ 0.7207 & 0.2730 & 0.0062 \\ 0.7219 & 0.2730 & 0.0051 \\ 0.0092 & 0.2730 & 0.7178 \\ 0.0050 & 0.2730 & 0.7219 \end{bmatrix} \qquad d^2 = \begin{bmatrix} 0.8518 & 0.0041 & 0.1441 \\ 0.8500 & 0.0059 & 0.1441 \\ 0.8517 & 0.0042 & 0.1441 \\ 0.8454 & 0.0106 & 0.1441 \\ 0.0068 & 0.8491 & 0.1441 \\ 0.8465 & 0.0094 & 0.1441 \end{bmatrix}$$

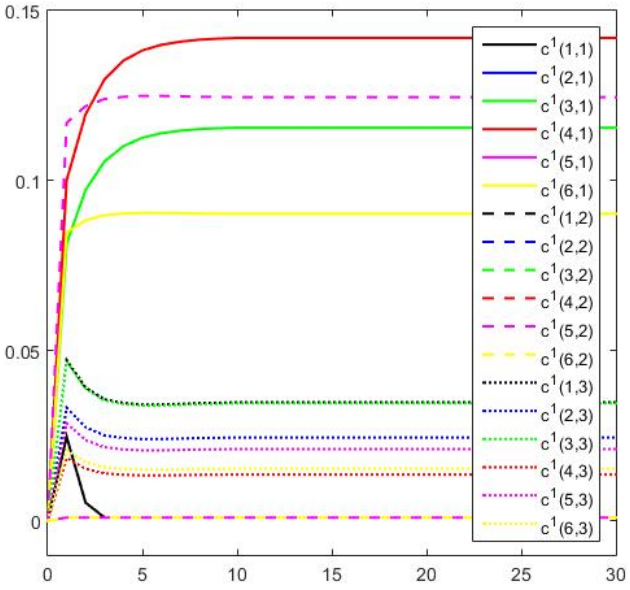


Figure B.5 Convergence of the strategies for player 1 in the KS solution.

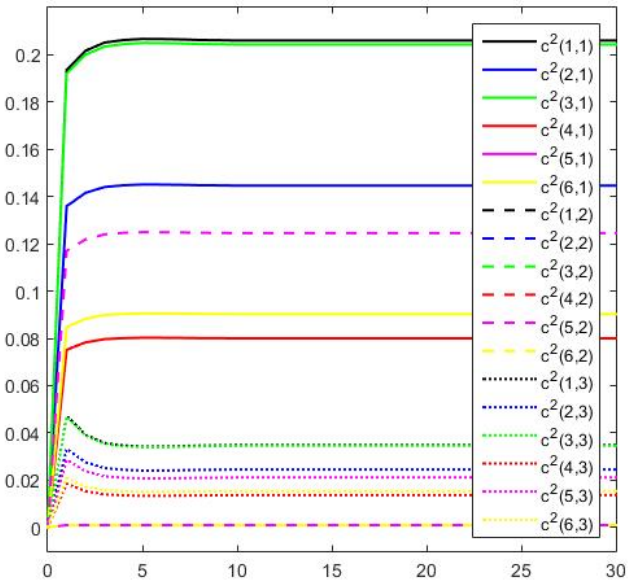


Figure B.6 Convergence of the strategies for payer 2 in the KS solution.

With the strategies calculated, the resulting utilities in the Kalai-Smorodinsky bargaining solution for each player are as follows:

$$\psi^1(c^1, c^2) = 960.5554 \qquad \psi^2(c^1, c^2) = 841.0831$$

Figure B.7 shows the straight line linking the utilities obtained at the disagreement point and those obtained at the utopia point. We can also observe that the Nash solution approaches this line while the Kalai-Smorodinsky solution is exactly on this line. The utilities on the utopia point for the bargaining problem are for each player as follows:

$$\psi^{1*}(c^1, c^2) = 964.3472 \qquad \psi^{2*}(c^1, c^2) = 849.8365$$

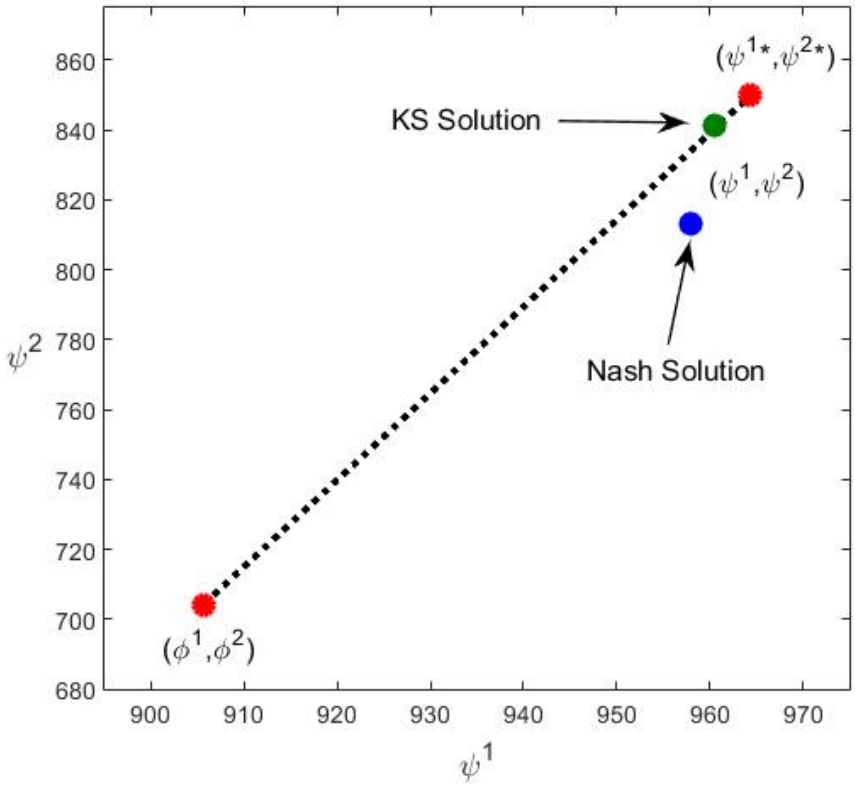


Figure B.7 The bargaining Solution.

Appendix C
Convergence Analysis of the Extraproximal Method

Lemma C.1 *Let $f(z)$ be a convex function defined on the convex set Z. If $z \in Z$ and z^* is a minimizer of function $\varphi(z) = \frac{1}{2}\|z - x\|^2 + \alpha f(z)$ on Z where x and z are fixed. Then, $f(z)$ satisfies the inequality:*

$$\frac{1}{2}\|z^* - x\|^2 + \alpha f(z^*) \leq \frac{1}{2}\|z - x\|^2 + \alpha f(z) - \frac{1}{2}\|z - z^*\|^2 \tag{C.1}$$

Proof. A necessary condition for a minimum at z^* can be written as

$$\langle z^* - x + \alpha \nabla f(z^*), z - z^* \rangle \geq 0$$

and the convexity condition for $f(z)$ is as follows

$$f(z) \geq f(z^*) + \langle \nabla f(z^*), z - z^* \rangle$$

Employing the necessary condition for a minimum at z^*, we have

$$0 \leq \langle z^* - x + \alpha \nabla f(z^*), z - z^* \rangle$$

$$= \langle z^* - x + \alpha, z - z^* \rangle + \langle \alpha \nabla f(z^*), z - z^* \rangle$$

$$= \langle z^* - x + \alpha, z - z^* \rangle + \alpha \langle \nabla f(z^*), z - z^* \rangle$$

Then, by the convexity condition for $f(z)$, we have $\langle \nabla f(z^*), z - z^* \rangle \leq f(z) - f(z^*)$. Now, combining the inequalities it follows that

$$0 \leq \langle z^* - x + \alpha, z - z^* \rangle + \alpha f(z) - \alpha f(z^*)$$

$$\langle z^* - x + \alpha, z - z^* \rangle \geq \alpha f(z^*) - \alpha f(z)$$

Using the identity

$$\frac{1}{2}\|z-x\|^2 = \frac{1}{2}\|z-z^*\|^2 + \langle z-z^*, z^*-x\rangle + \frac{1}{2}\|z^*-x\|^2$$

we have

$$\frac{1}{2}\|z-x\|^2 \geq \frac{1}{2}\|z-z^*\|^2 + \alpha f(z^*) - \alpha f(z) + \frac{1}{2}\|z^*-x\|^2$$
$$\frac{1}{2}\|z^*-x\|^2 + \alpha f(z^*) \leq \frac{1}{2}\|z-x\|^2 + \alpha f(z) - \frac{1}{2}\|z-z^*\|^2$$

Then, inequality (C.1) is proven. ■

Lemma C.2 *Consider the set of regularized solutions of a non-empty game. The behavior of the regularized function is described by the following inequality:*

$$L_\delta(\tilde{w}, \tilde{w}) - L_\delta(\tilde{v}^*_\delta, \tilde{w}) \geq \delta\|\tilde{w} - \tilde{v}^*_\delta\| \tag{C.2}$$

for all $\tilde{w} \in \left\{\tilde{w} \mid \tilde{w} \in \tilde{U} \times \tilde{Z}\right\}$ *and* $\delta > 0$.

Proof. The function $\tilde{L}_\delta(\tilde{u}, \tilde{z}^*_\delta)$ is strictly convex, then we have

$$\tilde{L}_\delta(\tilde{u}, \tilde{z}^*_\delta) - \tilde{L}_\delta(\tilde{u}^*_\delta, \tilde{z}) = \left[\tilde{L}_\delta(\tilde{u}, \tilde{z}^*_\delta) - \tilde{L}_\delta(\tilde{u}^*_\delta, \tilde{z}^*_\delta)\right] + \left[\tilde{L}_\delta(\tilde{u}^*_\delta, \tilde{z}^*_\delta) - \tilde{L}_\delta(\tilde{u}^*_\delta, \tilde{z})\right]$$
$$\geq \delta\left(\|\tilde{u} - \tilde{u}^*_\delta\|^2 + \|\tilde{z} - \tilde{z}^*_\delta\|^2\right)$$

Then, we have that

$$L_\delta(\tilde{w}, \tilde{w}) - L_\delta(\tilde{v}^*_\delta, \tilde{w}) = L_\delta(\tilde{w}_1, \tilde{v}^*_{2,\delta}) - L_\delta(\tilde{v}^*_{1,\delta}, \tilde{w}_2)$$
$$\geq \delta\left(\|\tilde{w}_1 - \tilde{v}^*_{1,\delta}\|^2 + \|\tilde{w}_2 - \tilde{v}^*_{2,\delta}\|^2\right) = \delta\|\tilde{w} - \tilde{v}^*_\delta\|^2$$

■

Lemma C.3 *Let* $\tilde{\mathcal{L}}_\delta(\tilde{u}, \tilde{z})$ *be differentiable in* $\tilde{u}$ *and* $\tilde{z}$*, whose partial derivative with respect to* $\tilde{z}$ *satisfies the Lipschitz condition with positive constant* C_0*. Then,*

$$\|\tilde{v}_{n+1} - \hat{v}_n\| \leq \gamma C_0\|\tilde{v}_n - \hat{v}_n\| \tag{C.3}$$

Proof. Consider the following inequality C.1

$$\frac{1}{2}\|z^* - x\|^2 + \alpha f(z^*) \leq \frac{1}{2}\|z - x\|^2 + \alpha f(z) - \frac{1}{2}\|z - z^*\|^2$$

and let for time n assign the following variables to Eq. (C.1)

$$\begin{gathered} \alpha = \gamma,\ z = \tilde{w},\ x = \tilde{v}_n,\ z^* = \hat{v}_n, \\ f(z) = L_\delta(\tilde{w}, \tilde{v}_n),\ f(z^*) = L_\delta(\hat{v}_n, \tilde{v}_n) \end{gathered} \tag{C.4}$$

Then, we can rewrite the first step of the equivalent extraproximal method in an equivalent form to Eq. (C.1) replacing the variables (C.4) as follows

$$\frac{1}{2}\|\hat{v}_n - \tilde{v}_n\|^2 + \gamma L_\delta(\hat{v}_n, \tilde{v}_n) \leq \frac{1}{2}\|\tilde{w} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{w}, \tilde{v}_n) - \frac{1}{2}\|\tilde{w} - \hat{v}_n\|^2 \tag{C.5}$$

As well, let for time $n+1$ assign the following variables to Eq. (C.1)

$$\begin{gathered} z = \tilde{w},\ x = \tilde{v}_n,\ z^* = \tilde{v}_{n+1}, \\ f(z) = L_\delta(\tilde{w}, \hat{v}_n),\ f(z^*) = L_\delta(\tilde{v}_{n+1}, \hat{v}_n) \end{gathered} \tag{C.6}$$

Then, we can rewrite the second step of the equivalent extraproximal method in an equivalent form to Eq. (C.1) replacing the variables (C.6) as follows

$$\frac{1}{2}\|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_{n+1}, \hat{v}_n) \leq \frac{1}{2}\|\tilde{w} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{w}, \hat{v}_n) - \frac{1}{2}\|\tilde{w} - \tilde{v}_{n+1}\|^2 \tag{C.7}$$

Assigning $\tilde{w} = \tilde{v}_{n+1}$ and replacing in (C.5) we obtain

$$\frac{1}{2}\|\hat{v}_n - \tilde{v}_n\|^2 + \gamma \Psi_\delta(\hat{v}_n, \tilde{v}_n) \leq \frac{1}{2}\|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_{n+1}, \tilde{v}_n) - \frac{1}{2}\|\tilde{v}_{n+1} - \hat{v}_n\|^2 \tag{C.8}$$

as well, replacing $\tilde{w} = \hat{v}_n$ in (C.7) we obtain

$$\frac{1}{2}\|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_{n+1}, \hat{v}_n) \leq \frac{1}{2}\|\hat{v}_n - \tilde{v}_n\|^2 + \gamma L_\delta(\hat{v}_n, \hat{v}_n) - \frac{1}{2}\|\hat{v}_n - \tilde{v}_{n+1}\|^2 \tag{C.9}$$

Adding (C.8) with (C.9) we obtain

$$\gamma L_\delta(\hat{v}_n, \tilde{v}_n) + \gamma L_\delta(\tilde{v}_{n+1}, \hat{v}_n) \leq \gamma L_\delta(\tilde{v}_{n+1}, \tilde{v}_n) - \frac{1}{2}\|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \gamma L_\delta(\hat{v}_n, \hat{v}_n) - \frac{1}{2}\|\hat{v}_n - \tilde{v}_{n+1}\|^2$$

Then, we have that

$$\tfrac{1}{2}\|\tilde{v}_{n+1}-\hat{v}_n\|^2 \le \gamma L_\delta(\tilde{v}_{n+1},\tilde{v}_n)+\gamma L_\delta(\hat{v}_n,\hat{v}_n)-\gamma L_\delta(\hat{v}_n,\tilde{v}_n)-\gamma L_\delta(\tilde{v}_{n+1},\hat{v}_n)\le$$

$$\gamma\left([L_\delta(\tilde{v}_{n+1},\tilde{v}_n)-L_\delta(\hat{v}_n,\tilde{v}_n)]+[L_\delta(\hat{v}_n,\hat{v}_n)-L_\delta(\tilde{v}_{n+1},\hat{v}_n)]\right)$$

Now, assign the following variables

$$\tilde{w}+h=\tilde{v}_{n+1},\ \tilde{w}=\hat{v}_n,\ \tilde{v}+t=\tilde{v}_n,\ \tilde{v}=\hat{v}_n,\ h=\tilde{v}_{n+1}-\hat{v}_n,\ t=\tilde{v}_n-\hat{v}_n$$

Because all partial derivative of $\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z})$ satisfy the Lipschitz condition with positive constant C, the following Lipschitz-type condition holds:

$$\|[L_\delta(\tilde{w}+h,\tilde{v}+t)-L_\delta(\tilde{w},\tilde{v}+t)]-[L_\delta(\tilde{w}+h,\tilde{v})-L_\delta(\tilde{w},\tilde{v})]\|\le C\|h\|\|t\| \qquad \text{(C.10)}$$

for any $\tilde{w},h,\tilde{v},t\in\tilde{U}\times\tilde{Z}$. Then, employing Eq. (C.10) we conclude

$$\|\tilde{v}_{n+1}-\hat{v}\|^2\le\gamma[L_\delta(\tilde{v}_{n+1},\tilde{v}_n)-L_\delta(\hat{v}_n,\tilde{v}_n)]-\gamma[L_\delta(\tilde{v}_{n+1},\hat{v}_n)-L_\delta(\hat{v}_{n+1},\hat{v}_n)]\le$$

$$\gamma C\|\tilde{v}_{n+1}-\hat{v}_n\|\|\tilde{v}_n-\hat{v}_n\|$$

which implies

$$\|\tilde{v}_{n+1}-\hat{v}_n\|\le\gamma C\|\tilde{v}_n-\hat{v}_n\|$$

■

Theorem C.4 ***[Convergence and Rate of Convergence]*** *Let $\tilde{\mathcal{L}}_\delta(\tilde{u},\tilde{z})$ be differentiable in $\tilde{u}$ and $\tilde{z}$, whose partial derivative with respect to $\tilde{z}$ satisfies the Lipschitz condition with positive constant C. Then, for any $\delta\in(0,1)$ and*

$$C_0=\sum_{l=1}^{\mathcal{N}}C_{0,l}\ \le\ \mathcal{N}\max_{l=\overline{1,\mathcal{N}}}C_{0,l}\ =\ \mathcal{N}C_0^+$$

there exists a small-enough

$$\gamma_0=\gamma_0(\delta)<C:=\min\left\{\frac{1}{\sqrt{2}C_0^+\mathcal{N}},\frac{1+\sqrt{1+2\left(C_0^+\right)^2}}{2\left(C_0^+\right)^2\mathcal{N}}\right\}$$

where such that, for any $0 < \gamma \leq \gamma_0$*, sequence* $\{\tilde{v}_n\}$*, which generated by the equivalent extraproximal procedure, monotonically converges with exponential rate* $q \in (0,1)$ *to an equilibrium point* $\tilde{v}^*$*, i.e.,*

$$\|\tilde{v}_n - \tilde{v}^*\|^2 \leq e^{n \ln q} \|\tilde{v}_0 - \tilde{v}^*\|^2$$

where

$$q = 1 + \frac{4(\delta\gamma)^2}{1 + 2\delta\gamma - 2\gamma^2 C^2} - 2\delta\gamma < 1$$

and $q_{\min}$ *is given by*

$$q_{\min} = 1 - \frac{2\delta\gamma}{1 + 2\delta\gamma} = \frac{1}{1 + 2\delta\gamma}$$

Proof. Let $\tilde{w} = \tilde{v}_{n+1}$, then replacing in (C.5) we obtain

$$\frac{1}{2}\|\hat{v}_n - \tilde{v}_n\|^2 + \gamma L_\delta(\hat{v}_n, \tilde{v}_n) \leq \frac{1}{2}\|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_{n+1}, \tilde{v}_n) - \frac{1}{2}\|\tilde{v}_{n+1} - \hat{v}_n\|^2 \quad \text{(C.11)}$$

as well, let $\tilde{w} = \tilde{v}_\delta^* \in \tilde{U}^* \times \tilde{Z}^*$ then replacing in (C.7) we get

$$\frac{1}{2}\|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_{n+1}, \hat{v}_n) \leq \frac{1}{2}\|\tilde{v}_\delta^* - \tilde{v}_n\|^2 + \gamma L_\delta(\tilde{v}_\delta^*, \hat{v}_n) - \frac{1}{2}\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 \quad \text{(C.12)}$$

Adding Eq. (C.11) and Eq. (C.12) and multiplying by two yields

$$\begin{gathered}\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 - 2\gamma L_\delta(\tilde{v}_\delta^*, \hat{v}_n) + \\ 2\gamma[L_\delta(\tilde{v}_{n+1}, \hat{v}_n) + L_\delta(\hat{v}_n, \tilde{v}_n) - L_\delta(\tilde{v}_{n+1}, \tilde{v}_n)] \ \leq \ \|\tilde{v}_\delta^* - \tilde{v}_n\|^2\end{gathered} \quad \text{(C.13)}$$

Adding and subtracting $L_\delta(\hat{v}_n, \hat{v}_n)$ in Eq. (C.13) we have

$$\begin{gathered}\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 + 2\gamma\left[L_\delta(\hat{v}_n, \hat{v}_n) - L_\delta(\tilde{v}_\delta^*, \hat{v}_n)\right] + 2\gamma\left[L_\delta(\tilde{v}_{n+1}, \hat{v}_n) - \right. \\ \left. L_\delta(\hat{v}_n, \hat{v}_n) + L_\delta(\hat{v}_n, \tilde{v}_n) - L_\delta(\tilde{v}_{n+1}, \tilde{v}_n)\right] \ \leq \ \|\tilde{v}_\delta^* - \tilde{v}_n\|^2\end{gathered} \quad \text{(C.14)}$$

Let assign the following variables

$$\tilde{w} + h = \tilde{v}_{n+1}, \ \tilde{w} = \hat{v}_n, \ \tilde{v} + k = \tilde{v}_n, \ \tilde{v} = \hat{v}_n$$

having $h = \tilde{v}_{n+1} - \hat{v}_n$ and $k = \tilde{v}_n - \hat{v}_n$. Using (C.10) the inequality (C.14) becomes

$$\begin{gathered}\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 + 2\gamma\left[L_\delta(\hat{v}_n, \hat{v}_n) - L_\delta(\tilde{v}_\delta^*, \hat{v}_n)\right] - \\ 2\gamma C\|\tilde{v}_{n+1} - \hat{v}_n\|\|\tilde{v}_n - \hat{v}_n\| \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2\end{gathered}$$

Applying (C.3) to the last term in the left-hand side and in view of the strict convexity property of L_δ in Eq. (C.2) given by

$$L_\delta(\hat{v}_n, \hat{v}_n) - L_\delta(\tilde{v}_\delta^*, \hat{v}_n) \geq \delta \|\hat{v}_n - \tilde{v}_\delta^*\|^2$$

we get

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + 2\gamma\delta\|\hat{v}_n - \tilde{v}_\delta^*\|^2 + \left(1 - 2\gamma^2 C^2\right)\|\tilde{v}_n - \hat{v}_n\|^2 \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2$$

We know that

$$2\langle a - c, c - b\rangle = \|a - b\|^2 - \|a - c\|^2 - \|c - b\|^2$$

Then, replacing $a = \hat{v}_n$, $b = \tilde{v}_\delta^*$ and $c = \tilde{v}_n$, to the left-hand side of the last inequality we have

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + (1 - 2\gamma^2 C^2)\|\tilde{v}_n - \hat{v}_n\|^2 + 2\gamma\delta\left(2\langle \hat{v}_n - \tilde{v}_n, \tilde{v}_n - \tilde{v}_\delta^*\rangle\right) +$$

$$\|\tilde{v}_n - \hat{v}_n\|^2 + \|\tilde{v}_n - \tilde{v}_\delta^*\|^2 = \|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + (1 + 2\gamma\delta - 2\gamma^2 C^2)\|\tilde{v}_n - \hat{v}_n\|^2 +$$

$$4\gamma\delta\langle \hat{v}_n - \tilde{v}_n, \tilde{v}_n - \tilde{v}_\delta^*\rangle + 2\gamma\delta\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2$$

Completing the square form of the third and fourth terms yields

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + (1 + 2\gamma\delta - 2\gamma^2 C^2)\|\tilde{v}_n - \hat{v}_n\|^2 + 4\gamma\delta\langle \hat{v}_n - \tilde{v}_n, \tilde{v}_n - \tilde{v}_\delta^*\rangle +$$

$$\frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2}\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 - \frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2}\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 + 2\gamma\delta\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2$$

Then, we have that

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \left\|\sqrt{1 + 2\gamma\delta - 2\gamma^2 C^2}(\tilde{v}_n - \hat{v}_n) + \frac{2\gamma\delta}{\sqrt{1 + 2\gamma\delta - 2\gamma^2 C^2}}(\tilde{v}_n - \tilde{v}_\delta^*)\right\|^2 -$$

$$\left(\frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2}\right)\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 + 2\gamma\delta\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2$$

developing the terms we obtain that

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 \leq \|\tilde{v}_\delta^* - \tilde{v}_n\|^2 + \left(\frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2}\right)\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 - 2\gamma\delta\|\tilde{v}_n - \tilde{v}_\delta^*\|^2 -$$

$$\|\tilde{v}_{n+1} - \hat{v}_n\|^2 - \left\|\sqrt{1 + 2\gamma\delta - 2\gamma^2 C^2}(\tilde{v}_n - \hat{v}_n) + \frac{2\gamma\delta}{\sqrt{1 + 2\gamma\delta - 2\gamma^2 C^2}}(\tilde{v}_n - \tilde{v}_\delta^*)\right\|^2$$

as a result we have that

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 \leq \left(1 - 2\gamma\delta + \frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2}\right) \|\tilde{v}_\delta^* - \tilde{v}_n\|^2$$

where

$$q = 1 - 2\gamma\delta + \frac{(2\gamma\delta)^2}{1 + 2\gamma\delta - 2\gamma^2 C^2} < 1$$

Iterating over the previous inequality we have

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 \leq q\|\tilde{v}_\delta^* - \tilde{v}_n\|^2 \leq ... \leq e^{n+1\ln q}\|\tilde{v}_\delta^* - \tilde{v}_0\|^2 \tag{C.15}$$

That implies that the series converge and also that the trajectories are bounded. Then, by Eq. (C.15) we have that

$$\|\tilde{v}_\delta^* - \tilde{v}_{n+1}\|^2 \underset{n\to\infty}{\to} 0$$

Given that $\tilde{v}$ is a bounded sequence, by the Weierstrass Theorem there exists a point $\tilde{v}'$ such that any subsequence $\tilde{v}_{n_i}$ satisfies that $\tilde{v}_{n_i} \underset{n_i\to\infty}{\to} \tilde{v}'$. In addition, we have that $\|\tilde{v}_{n_i} - \tilde{v}_{n_i+1}\|^2 \to 0$. Fixing, $n = n_i$ in the equivalent proximal equation and computing the limit when $n_i \to \infty$ we have

$$\tilde{v}' = \arg\min_{\tilde{w}\in\tilde{U}\times\tilde{Z}} \left\{\frac{1}{2}\|\tilde{w} - \tilde{v}'\|^2 + \gamma L_\delta(\tilde{w}, \tilde{v}')\right\}$$

Then, we have that $\tilde{v}' = \tilde{v}_\delta^*$, i.e., any limit point of the sequence $\tilde{v}_n$ is a solution of the problem. Given that $\|\tilde{v}_n - \tilde{v}_\delta^*\|^2$ is monotonically decreasing then, there exists a unique limit point (equilibrium point). As a consequence, we have that the sequence $\tilde{v}_n$ satisfies that $\tilde{v}_n \underset{n\to\infty}{\to} \tilde{v}_\delta^*$ with a convergence velocity of $e^{n+1\ln q}$. ■

Remark C.5 *The exponential rate $q \in (0, 1)$ (see Figure C.1) satisfies*

$$q \simeq q_0\left(1 + \tfrac{1}{N^2}\right).$$

C.0.1 Convergence conditions of δ and γ

This section presents the convergence conditions and compute the estimate rate of convergence of the variables γ and δ [75]. The regularizing parameter δ and its asymptotic behavior when $\delta \to 0$ is analyzed. Also, the step size parameter γ and its asymptotic behavior when $\gamma \to 0$ is analyzed.

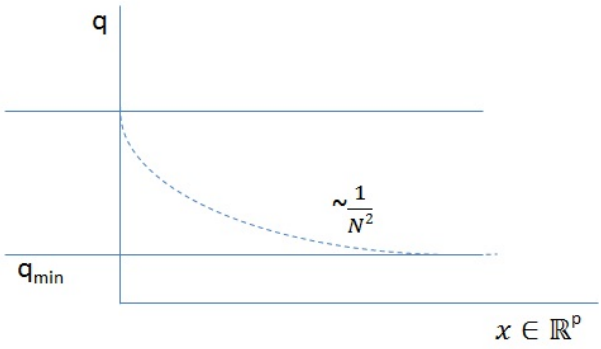


Figure C.1 Rate of convergence.

Theorem C.6 *Within the class of numerical sequences*

$$\gamma_n = \frac{\gamma_0}{(n+n_0)^\gamma} \qquad \gamma_0, n_0, \gamma > 0$$

$$\delta_n = \frac{\delta_0}{(n+n_0)^\delta} \qquad \delta_0, \delta > 0$$

the step size γ_n and the regularizing parameter δ_n satisfy the following conditions:

$$0 < \gamma_n \to 0,\ 0 < \delta_n \to 0 \qquad \textit{when } n \to \infty$$

$$\sum_{n=0}^{\infty} \gamma_n \delta_n = \infty$$

$$\frac{\gamma_n}{\delta_n} \to \varepsilon \textit{ which is small enough} \qquad \frac{|\delta_{n+1} - \delta_n|}{\gamma_n \delta_n} \to 0 \textit{ when } n \to \infty$$

for $\gamma+\delta \leq 1$, $\gamma \geq \delta$, $\gamma < 1$.

Proof. It follows from the estimates that

$$\gamma_n \delta_n = O\left(\frac{1}{n^{\gamma+\delta}}\right)$$

we have that

$$|\delta_{n+1} - \delta_n| = O\left(\frac{1}{n^\delta} - \frac{1}{(n+1)^\delta}\right) = O\left(\frac{1}{(n+1)^\delta}\left[\left(1+\frac{1}{n}\right)^\delta - 1\right]\right)$$
$$= O\left(\frac{1}{(n+1)^\delta}\left[\left(\frac{1}{n}\right)^\delta + o(1)\right]\right) = O\left(\frac{1}{n^\delta + 1}\right)$$

and

$$\frac{|\delta_{n+1} - \delta_n|}{\gamma_n \delta_n} = O\left(\frac{1}{n^{1-\gamma}}\right)$$

■

Theorem C.7 *Let u and x two variables with non-negative components for the players. Then, within the class of numerical sequences we have that*

$$\gamma_n = \frac{\gamma_0}{(n+n_0)^\gamma} \qquad \gamma_0, n_0, \gamma > 0$$

$$\delta_n = \frac{\delta_0}{(n+n_0)^\delta} \qquad \delta_0, \delta > 0$$

of the procedure given in proximal method, the rate of convergence for the players is given by the step size γ_n and the regularizing parameter δ_n

$$\|u_n - u^{**}\| + \|x_n - x^{**}\| = O\left(\frac{1}{n^{\varkappa}}\right)$$

where $\varkappa$ is equal to

$$\varkappa = \min\{\gamma - \delta; 1-\gamma; \delta\} \tag{C.16}$$

Then, the maximal rate $\varkappa^*$ of convergence is attained for

$$\gamma = \gamma^* = 2/3 \qquad \delta = \delta^* = 1/3 \tag{C.17}$$

Proof. It follows that for $\varkappa_0$ characterizing the rate of convergence is given by

$$r_n = \|u_n - u^*(\delta_n)\| + \|x_n - x^*(\delta_n)\| = O\left(\frac{1}{n^{\varkappa_0}}\right)$$

we have $\varkappa_0 = \min\{\gamma - \delta; 1 - \gamma; \delta\}$. It follows from the linear dependence of the regularized Lagrange function on δ that

$$\|u_n - u^{**}\| + \|x_n - x^{**}\| = r_n + O(\delta_n) = O\left(\frac{1}{n^{\varkappa_0}}\right) + O\left(\frac{1}{n^{\delta}}\right) = O\left(\frac{1}{n^{\min\{\varkappa_0;\delta\}}}\right)$$

which implies (C.16). The maximal value $\varkappa$ of $\varkappa^*$ is attained when $\gamma - \delta = 1 - \gamma = \delta$, i.e., when (C.17) holds. ■

Remark C.8 *In the case of a Stackelberg game, we have a similar rate of convergence for the followers given by*

$$\|v_n - v^{**}\| + \|w_n - w^{**}\| = O\left(\frac{1}{n^{\varkappa}}\right)$$

Appendix D
The Lagrange Method for Polylinear Programming Problems

D.1 Polylinear optimization problem formulation

Consider the following poly-linear programming problem

$$\begin{gathered} f(x)=\alpha_1\sum_{j_1=1}^{N}c_{j_1}x_{j_1}+\alpha_2\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}c_{j_1,j_2}x_{j_1}x_{j_2}+ \\ \alpha_3\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N}c_{j_1,j_2,j_3}x_{j_1}x_{j_2}x_{j_3}+\cdots+ \\ \alpha_{N-1}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\cdots\sum_{j_{N-1}=1}^{N}c_{j_1,\cdots,j_{N-1}}x_{j_1}\cdots x_{j_{N-1}}+ \\ \alpha_N\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\cdots\sum_{j_N=1}^{N}c_{j_1,\cdots,j_N}x_{j_1}\cdots x_{j_N}\to\min_{x\in X_{\text{adm}}} \end{gathered} \tag{D.1}$$

where $\alpha_j=\{0;1\}\ (j=1,...,N)$ are binary variables and X_{adm} is a bounded set defined as follows

$$X_{\text{adm}}:=\left\{x\in\mathbb{R}^N:x\geq 0,\ A_{\text{eq}}x=b_{\text{eq}}\in\mathbb{R}^{M_0},\ A_{\text{ineq}}x\leq b_{\text{ineq}}\in\mathbb{R}^{M_1}\right\}$$

Notice that this problem may have non-unique solution and $\det\left(A_{\text{eq}}^{\intercal}A_{\text{eq}}\right)=0$. Define by $X^*\subseteq X_{\text{adm}}$ the set of all solutions of the problem (D.1).

D.2 The Lagrange Method

Following [110] and [111] consider the *Regularized Lagrange Function (RLF)*

$$\begin{gathered} \mathcal{L}_{\theta,\delta}\left(x,\mu_0,\mu_1\right):=\theta f(x)+\mu_0^{\intercal}\left(A_{\text{eq}}x-b_{\text{eq}}\right)+\mu_1^{\intercal}\left(A_{\text{ineq}}x-b_{\text{ineq}}\right) \\ +\frac{\delta}{2}\left(\|x\|^2-\|\mu_0\|^2-\|\mu_1\|^2\right) \end{gathered} \tag{D.2}$$

where the parameters θ, δ are positive and the Lagrange vector-multipliers $\mu_1 \in \mathbb{R}^{M_1}$ are non-negative and the components of $\mu_0 \in \mathbb{R}^{M_0}$ may have any sign. Obviously, the optimization problem

$$\mathcal{L}_{\theta,\delta}(x, \mu_0, \mu_1) \to \min_{x \in X_{\text{adm}}} \max_{\mu_0, \mu_{1\geq 0}} \tag{D.3}$$

has a unique saddle-point on x since the optimized RLF (D.2) is *strongly convex* [74] if the parameters θ and $\delta > 0$ provide the condition

$$\frac{\partial^2}{\partial x \partial x^{\intercal}} \mathcal{L}_{\theta,\delta}(x, \mu_0, \mu_1) > 0 \quad \forall\, x \in X_{\text{adm}} \subset \mathbb{R}^N$$

and is strongly concave on the Lagrange multipliers μ_0, μ_1 for any $\delta > 0$. In view of these properties RLF has the unique saddle point $(x^*(\delta), \mu_0^*(\theta, \delta), \mu_1^*(\theta, \delta))$ (see The Kuhn-Tucker Theorem 21.13 in [74]) for which the following inequalities hold: for any μ_0, μ_1 with non-negative components and any $x \in \mathbb{R}^N$

$$\mathcal{L}_{\theta,\delta}(x, \mu_0^*(\theta, \delta), \mu_1^*(\theta, \delta)) \geq \mathcal{L}_{\theta,\delta}(x^*(\delta), \mu_0^*(\theta, \delta), \mu_1^*(\theta, \delta)) \geq \mathcal{L}_{\theta,\delta}(x^*(\delta), \mu_0, \mu_1)$$

As for the non-regularized function $\mathcal{L}_{1,0}(x, \mu_0, \mu_1)$, it may have several (not obligatory unique) saddle points $(x^*, \mu_0^*, \mu_1^*) \in X^* \otimes \Lambda^*$.

D.2.1 Property of Lagrange Method

Proposition D.1 *If the parameter θ and the regularizing parameter δ tend to zero by a particular manner, then we may expect that $x^*(\theta, \delta)$ and $\mu_0^*(\theta, \delta)$, $\mu_1^*(\theta, \delta)$ which are the solutions of the min-max optimization problem (*D.*3) tend to the set $X^* \otimes \Lambda^*$ of all saddle point of the original optimization problem (*D.*1), that is,*

$$\rho\{x^*(\theta, \delta), \mu_0^*(\theta, \delta), \mu_1^*(\theta, \delta); X^* \otimes \Lambda^*\} \underset{\theta,\delta \downarrow 0}{\to} 0 \tag{D.4}$$

where $\rho\{a; X^* \otimes \Lambda^*\}$ is the Hausdorff distance defined as

$$\rho\{a; X^* \otimes \Lambda^*\} = \min_{z^* \in X^* \otimes \Lambda^*} \|a - z^*\|^2$$

Below we define exactly how the parameters θ and δ should tend to zero to provide the property (D.4).

D.3 The extremal points of the regularized Lagrange function

The next lemma describes the dependence of the saddle-point $x^*(\theta,\delta)$ and $\mu_0^*(\theta,\delta)$, $\mu_1^*(\theta,\delta)$ of the RLF on the regularizing parameters δ,θ and analyses it asymptotic behavior when both of them tend to zero.

Theorem D.2 *Assume that*

*1) the bounded set X^* of all solutions of the original optimization problem (*D.*1) is not empty and the Slater's condition holds, that is, there exists a point $\mathring{x} \in X_{adm}$ such that*

$$A_{ineq}\mathring{x} < b_{ineq}$$

2) The parameters θ and δ are time-varying, i.e.,

$$\theta = \theta_n,\ \delta = \delta_n\ \ (n = 0,1,2,....)$$

such that

$$0 < \theta_n \downarrow 0,\ \frac{\theta_n}{\delta_n} \downarrow 0 \text{ when } n \to \infty$$

Then

$$x_n^* := x^*(\theta_n,\delta_n) \underset{n\to\infty}{\to} x^{**}$$

$$\mu_0^*(\theta_n,\delta_n) \underset{n\to\infty}{\to} \mu_0^{**}$$

$$\mu_1^*(\theta_n,\delta_n) \underset{n\to\infty}{\to} \mu_1^{**}$$

*where $x^{**} \in X^*$, $(\mu_0^{**},\mu_1^{**}) \in \Lambda^*$ define the solution of the original problem (*D.*1) with the minimal norm which is unique, i.e.,*

$$\|x^{**}\|^2 + \|\mu_0^{**}\|^2 + \|\mu_1^{**}\|^2 \le \|x^*\|^2 + \|\mu_0^*\|^2 + \|\mu_1^*\|^2$$

for all $x^* \in X^*$, and $(\mu_0^*,\mu_1^*) \in \Lambda^*$.

Proof.

a) First, prove that the Hessian matrix $H := \dfrac{\partial^2}{\partial x \partial x^\intercal}\mathcal{L}_{\theta,\delta}\left(x, \mu_0, \mu_1\right)$ is strictly positive definite for all $x \in \mathbb{R}^N$ and for some positive θ and δ, satisfying a special relation, namely, $H > 0$. We have

$$\frac{\partial^2}{\partial x^2}\mathcal{L}_{\theta,\delta}\left(x, \mu_0, \mu_1\right) = \theta\frac{\partial^2}{\partial x^2}f(x) + \delta I_{N\times N} \geq \delta\left(1 + \frac{\theta}{\delta}\lambda^-\right) I_{N\times N} > 0 \;\forall\; \delta > \theta\left|\lambda^-\right|$$

where

$$\lambda^- := \min_{x\in X_{\text{adm}}} \lambda_{\min}\left(\frac{\partial^2}{\partial x^2}f(x)\right)$$

fulfilling the property $H > 0$ if $\delta > \theta\left|\lambda^-\right|$. This means that RLF (D.2) is strongly convex on x and, hence, has a unique minimal point defined below as x^*.

b) In view of the properties

$$\left(\nabla f\left(x\right), \left(y - x\right)\right) \leq f\left(y\right) - f\left(x\right)$$

$$\left(\nabla f\left(x\right), \left(x - y\right)\right) \geq f\left(x\right) - f\left(y\right)$$

valid for any convex function $f\left(x\right)$ and any x, y, for RLF at any admissible points x, μ_0, μ_1 and $x_n^* = x^*\left(\theta_n, \delta_n\right)$, $\mu_{0,n}^* = \mu_0^*\left(\theta_n, \delta_n\right)$, $\mu_{1,n}^* = \mu_1^*\left(\theta_n, \delta_n\right)$ we have

$$\left(x - x_n^*, \frac{\partial}{\partial x}\mathcal{L}_{\theta_n,\delta_n}\left(x, \mu_0, \mu_1\right)\right) - \left(\mu_0 - \mu_{0,n}^*, \frac{\partial}{\partial \mu_0}\mathcal{L}_{\theta_n,\delta_n}\left(x, \mu_0, \mu_1\right)\right) -$$

$$\left(\mu_1 - \mu_{1,n}^*, \frac{\partial}{\partial \mu_1}\mathcal{L}_{\theta_n,\delta_n}\left(x, \mu_0, \mu_1\right)\right) = \theta_n\left(x - x_n^*\right)^\intercal \frac{\partial}{\partial x}f\left(x\right) +$$

$$\left(x - x_n^*\right)^\intercal\left[A_{\text{eq}}^\intercal\mu_0 + A_{\text{ineq}}^\intercal\mu_1 + \delta_n x\right] + \left(\mu_0 - \mu_{0,n}^*\right)^\intercal\left(\delta_n\mu_0 - A_{\text{eq}}x + b_{\text{eq}}\right)$$

$$\left(\mu_1 - \mu_{1,n}^*\right)^\intercal\left(\delta_n\mu_1 - A_{\text{ineq}}x + b_{\text{ineq}}\right) = \theta_n f(x) + \mu_0^\intercal\left(A_{\text{eq}}x - b_{\text{eq}}\right) + \mu_1^\intercal\left(A_{\text{ineq}}x - b_{\text{ineq}}\right)$$

$$+ \frac{\delta_n}{2}\left(\|x\|^2 - \|\mu_0\|^2 - \|\mu_1\|^2\right) - \theta_n f(x_n^*) - \left(\mu_{0,n}^*\right)^\intercal\left(A_{\text{eq}}x_n^* - b_{\text{eq}}\right)$$

$$- \left(\mu_{1,n}^*\right)^\intercal\left(A_{\text{ineq}}x_n^* - b_{\text{ineq}}\right) - \frac{\delta_n}{2}\left(\|x_n^*\|^2 - \left\|\mu_{0,n}^*\right\|^2 - \left\|\mu_{1,n}^*\right\|^2\right)$$

$$= \mathcal{L}_{\theta_n,\delta_n}\left(x, \mu_{0,n}^*, \mu_{1,n}^*\right) - \mathcal{L}_{\theta,\delta}\left(x_n^*, \mu_0, \mu_1\right)$$

$$+ \frac{\delta_n}{2}\left(\|x - x_n^*\|^2 + \left\|\mu_0 - \mu_{0,n}^*\right\|^2 + \left\|\mu_1 - \mu_{1,n}^*\right\|^2\right) \tag{D.5}$$

which by the the saddle-point condition (D.2) implies

$$\theta_n (x - x_n^*)^\intercal \frac{\partial}{\partial x} f(x) + (x - x_n^*)^\intercal \left[A_{\text{eq}}^\intercal \mu_0 + A_{\text{ineq}}^\intercal \mu_1 + \delta_n x\right] +$$

$$\left(\mu_0 - \mu_{0,n}^*\right)^\intercal (\delta_n - A_{\text{eq}} x + b_{\text{eq}}) + \left(\mu_1 - \mu_{1,n}^*\right)^\intercal (\delta_n - A_{\text{ineq}} x + b_{\text{ineq}}) \geq \tag{D.6}$$

$$\frac{\delta_n}{2}\left(\|x - x_n^*\|^2 + \left\|\mu_0 - \mu_{0,n}^*, \right\|^2 + \left\|\mu_1 - \mu_{1,n}^*\right\|^2\right)$$

c) Selecting in (D.6) $x := x^* \in X^*$ (x^* is one of admissible solutions such that $A_{\text{eq}} x^* = b_{\text{eq}}$ and $A_{\text{ineq}} x^* \leq b_{\text{ineq}}$) and $\mu_0 = \mu_0^*$, $\mu_1 = \mu_1^*$ in view of the complementary slackness conditions

$$(\mu_1^*)_i (A_{\text{ineq}} x^* - b_{\text{ineq}})_i = \left(\mu_{1,n}^*\right)_i (A_{\text{ineq}} x_n^* - b_{\text{ineq}})_i = 0$$

we obtain

$$\theta_n (x^* - x_n^*)^\intercal \frac{\partial}{\partial x} f(x^*) + (x^* - x_n^*)^\intercal \left[A_{\text{eq}}^\intercal \mu_0^* + A_{\text{ineq}}^\intercal \mu_1^* + \delta_n x^*\right] +$$

$$\left(\mu_0^* - \mu_{0,n}^*\right)^\intercal (\delta_n \mu_0^* - A_{\text{eq}} x^* + b_{\text{eq}}) + \left(\mu_1^* - \mu_{1,n}^*\right)^\intercal (\delta_n \mu_1^* - A_{\text{ineq}} x^* + b_{\text{ineq}})$$

$$= \theta_n (x^* - x_n^*)^\intercal \frac{\partial}{\partial x} f(x^*) + (\mu_0^*)^\intercal ([A_{\text{eq}} x^* - b_{\text{eq}}] - [A_{\text{eq}} x_n^* - b_{\text{eq}}])$$

$$+ (\mu_1^*)^\intercal ([A_{\text{ineq}} x^* - b_{\text{ineq}}] - [A_{\text{ineq}} x_n^* - b_{\text{ineq}}]) + \delta_n (x^* - x_n^*)^\intercal x^* +$$

$$\delta_n \left(\mu_0^* - \mu_{0,n}^*\right)^\intercal \mu_0^* + \left(\mu_1^* - \mu_{1,n}^*\right)^\intercal \delta_n \mu_1^* + \left(\mu_{1,n}^*\right)^\intercal (A_{\text{ineq}} x^* - b_{\text{ineq}})$$

$$\geq \frac{\delta_n}{2}\left(\|x^* - x_n^*\|^2 + \left\|\mu_0^* - \mu_{0,n}^*\right\|^2 + \left\|\mu_1^* - \mu_{1,n}^*\right\|^2\right) \geq 0$$

Simplifying the last inequality we have

$$\theta_n (x^* - x_n^*)^\intercal \frac{\partial}{\partial x} f(x^*) + \delta_n (x^* - x_n^*)^\intercal x^* + \delta_n \left(\mu_0^* - \mu_{0,n}^*\right)^\intercal \mu_0^* + \left(\mu_1^* - \mu_{1,n}^*\right)^\intercal \delta_n \mu_1^* \geq 0$$

Dividing both sides of this inequality by δ_n and taking $\dfrac{\theta_n}{\delta_n} \underset{n\to\infty}{\to} 0$ we get

$$0 \leq \limsup_{n\to\infty} \left[(x^* - x_n^*)^\intercal x^* + \left(\mu_0^* - \mu_{0,n}^*\right)^\intercal \mu_0^* + \left(\mu_1^* - \mu_{1,n}^*\right)^\intercal \mu_1^*\right]$$

This means that there obligatory exists subsequences δ_k and θ_k $(k \to \infty)$ on which there exist the limits

$$x_k^* = x^*(\theta_k, \delta_k) \to \tilde{x}^*, \ \mu_{0,k}^* = \mu_0^*(\theta_k, \delta_k) \to \tilde{\mu}_0^*$$

$$\mu_{1,k}^* = \mu_1^*(\theta_k, \delta_k) \to \tilde{\mu}_1^* \text{ as } k \to \infty$$

Suppose that there exist two limit points for two different convergent subsequences, i.e., there exist the limits

$$x^*_{k'} = x^*\left(\theta_{k'}, \delta_{k'}\right) \to \bar{x}^*,\ \mu^*_{0,k'} = \mu^*_0\left(\theta_{k'}, \delta_{k'}\right) \to \bar{\mu}^*_0$$

$$\mu^*_{1,k'} = \mu^*_1\left(\theta_{k'}, \delta_{k'}\right) \to \bar{\mu}^*_1 \text{ as } k \to \infty$$

Then on these subsequences one has

$$0 \leq \left(x^* - \tilde{x}^*\right)^\intercal x^* + \left(\mu^*_0 - \tilde{\mu}^*_0\right)^\intercal \mu^*_0 + \left(\mu^*_1 - \tilde{\mu}^*_1\right)^\intercal \mu^*_1$$

$$0 \leq \left(x^* - \bar{x}^*\right)^\intercal x^* + \left(\mu^*_0 - \bar{\mu}^*_0\right)^\intercal \mu^*_0 + \left(\mu^*_1 - \bar{\mu}^*_1\right)^\intercal \mu^*_1$$

From this inequalities it follows that points $\left(\tilde{x}^*, \tilde{\mu}^*_0, \tilde{\mu}^*_1\right)$ and $\left(\bar{x}^*, \bar{\mu}^*_0, \bar{\mu}^*_1\right)$ correspond to the minimum point of the function

$$s\left(x^*, \mu^*_0, \mu^*_1\right) := \frac{1}{2}\left(\left\|x^*\right\|^2 + \left\|\mu^*_0\right\|^2 + \left\|\mu^*_1\right\|^2\right)$$

defined on $X^* \otimes \Lambda^*$ for all possible saddle-points of the non-regularized Lagrange function. But the function $s\left(x^*, \mu^*_0, \mu^*_1\right)$ is strictly convex, and, hence, its minimum is unique that gives $\tilde{x}^* = \bar{x}^*$, $\tilde{\mu}^*_0 = \bar{\mu}^*_0$, $\tilde{\mu}^*_0 = \bar{\mu}^*_0$. Proposition is proven.

■

Lemma D.3 *Under the assumptions of the Theorem D.2 there exist positive constants C_μ and C_δ such that*

$$\left\|x^*_n - x^*_m\right\| + \left\|\mu^*_{0,n} - \mu^*_{0,m}\right\| + \left\|\mu^*_{1,n} - \mu^*_{1,m}\right\| \leq C_\theta \left|\theta_n - \theta_m\right| + C_\delta \left|\delta_n - \delta_m\right|$$

Proof. It follows also from the necessary and sufficient conditions (D.5) for the points $x^*_n = x^*\left(\theta_n, \delta_n\right)$, $\mu^*_{0,n} = \mu^*_0\left(\theta_n, \delta_n\right)$, $\mu^*_{1,n} = \mu^*_1\left(\theta_n, \delta_n\right)$ to be the extremal points of the function $\mathcal{L}_{\theta_n, \delta_n}\left(x, \mu_0, \mu_1\right)$. ■

Bibliography

[1] Aiyoshi, E., Shimizu, K.: Hierarchical decentralized systems and its new solution by abarrier method. IEEE Transactions on Systems, Man, and Cybernetics **11**, 444–449 (1981)

[2] An, B., Pita, J., Shieh, E., Tambe, M., Kiekintveld, C., Marecki, J.: GUARDS and PROTECT: Next generation applications of security games. SIGECOM **10**, 31–34 (2011)

[3] Anant, T.C.A., Mukherji, B., Basu, K.: Bargaining without convexity: Generalizing the kalai-smorodinsky solution. Economics Letters **33**(2), 115–119 (1990)

[4] Antipin, A.S.: The convergence of proximal methods to fixed points of extremal mappings and estimates of their rate of convergence. Computational Mathematics and Mathematical Physics **35**(5), 539–551 (1995)

[5] Antipin, A.S.: An extraproximal method for solving equilibrium programming problems and games. Computational Mathematics and Mathematical Physics **45**(11), 1893–1914 (2005)

[6] Attouch, H., Soubeyran, A.: Local search proximal algorithms as decision dynamics with costs to move. Set Valued Analysis **19**, 157–177 (2011)

[7] Aumann, R.: Contributions to the Theory of Games, *Annals of Mathematics Study*, vol. IV, chap. Acceptable points in general cooperative n-person games, pp. 287–324 (1959)

[8] Axelrod, R., Dion, D.: More on the evolution of cooperation. Science **242**, 1385–1390 (1988)

[9] Bao, T.Q., Mordukhovich, B.S., Soubeyran, A.: Variational analysis in psychological modeling. Journal of Optimization Theory **164**(1), 290–315 (2015)

[10] Bard, J.: Practical bilevel optimization:algorithms and applications. The Netherlands:Kluwer (1998)

[11] Bard, J., Falk, J.: An explicit solution to the multi-level programming problem. Computers and Operations Research **9**, 77–100 (1982)

[12] Beyer, A., Gutman, I., Marinovic, I.: Optimal contracts with performance manipulation. Journal of Accounting Research **52**(4), 817–847 (2014)

[13] Bianco, L., Caramia, M., Giordani, S.: A bilevel flow model for hazmat transportation network design. Transportation Research Part C: Emerging Technologies **17**(2), 175–196 (2009)

[14] Birkeland, S., Tungodden, B.: Fairness motivation in bargaining: a matter of principle. Theory and Decision **77**(1), 125–151 (2014)

[15] Bos, D.: Privatization: A theoretical treatment. Clarendon Press, Oxford (1991)

[16] Christie, R., Geis, F.: Studies in Machiavellianism. Academic Press (1970)

[17] Clempner, J.B.: A continuous-time markov stackelberg security game approach for reasoning about real patrol strategies. International Journal of Control (2017). DOI 10.1080/00207179.2017.1371853. To be published

[18] Clempner, J.B.: A game theory model of manipulation based on the machiavellian social interaction theory: Moral and ethical behavior. J. Artif. Soc. Soc. Simulat. (2017). To be published

[19] Clempner, J.B.: Computing multiobjective markov chains handled by the extraproximal method. Annals of Operations Research (2018). DOI 10.1007/s10479-018-2755-9. To be published

[20] Clempner, J.B., Poznyak, A.S.: Computing the strong Nash equilibrium for Markov chains games. Applied Mathematics and Computation **265**, 911–927 (2015)

[21] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forwardbackward splitting. Multiscale Modeling and Simulation **4**(4), 1168–1200 (2006)

[22] Cournot, A.A.: Recherches sur les principes mathematiques de la theorie des richesses. Hachette, Paris (1838)

[23] Dawkins, R.: The selfish gene (1st ed.). Oxford, England: Oxford University Press (1976)

[24] Demange, G.: Intermediate preferences and stable coalition structures. Journal of Mathematical Economics **23**, 45–58 (1994)

[25] Demange, G., Henriet, D.: Sustainable oligopolies. Journal of Economic Theory **54,**, 417–428 (1991)

[26] Dev, N.K., Shankar, R., Choudhary, A.: Strategic design for inventory and production planning in closed-loop hybrid systems. Int. J. Production Economics **183**, 345–353 (2017)

[27] Eckstein, J.: Nonlinear proximal point algorithms using bregman functions, with applications to convex programming. Mathematics of Operations Research **18**(1), 202–226 (1993)

[28] Fave, F., Jiang, A., Yin, Z., Zhang, C., Tambe, M., Kraus, S., Sullivan, J.: Game-theoretic security patrolling with dynamic execution uncertainty and a case study on a real transit system. Journal of Artificial Intelligence Research **50**, 321–367 (2014)

[29] Forgó, F., Szép, J., F., S.: Introduction to the Theory of Games: concepts, methods, applications. Kluwer Academic Publishers (1999)

[30] Gatti, N., Rocco, M., Sandholm, T.: Algorithms for strong nash equilibrium with more than two agents. In: The Twenty-Seventh AAAI Conference on Artificial Intelligence, pp. 342–349. Bellevue, Washington, USA (2013)

[31] Gatti, N., Rocco, M., Sandholm, T.: Strong nash equilibrium is in smoothed p. In: The Twenty-Seventh AAAI Conference on Artificial Intelligence, pp. 29–31. Bellevue, Washington, USA (2013)

[32] Germeyer, Y.: Introduction to the theory of operations research. Nauka, Moscow (1971)

[33] Germeyer, Y.: Games with nonantagonistic Interests. Nauka, Moscow (1976)

[34] Greenberg, J., Weber, S.: Strong tiebout equilibrium under restricted preferences domain. Journal of Economic Theory **38**, 101–117 (1986)

[35] Guo, X., Hernández-Lerma, O.: Continuos–Time Markov Decision Processes: Theory and Applications. Springer-Verlag Berlin Heidelberg (2009)

[36] Han, Z., Ji, Z., Liu, K.: Fair multiuser channel allocation for ofdma networks using nash bargaining solutions and coalitions. IEEE Transactions on Communications **53**(8), 1366–1376 (2005)

[37] Herskovits, J., Leontiev, A., Dias, G., Santos, G.: Contact shape optimization: A bilevel programming approach. Structural and Multidisciplinary Optimization **20**, 214–221 (2000)

[38] Hotzman, R., Law-Yone, N.: Strong equilibrium in congestion games. Games and Economic Bahavior **21**, 85–101 (1997)

[39] Howard, N.: Paradoxes of rationality: Theory of metagames and political behaviour. MIT Press (1971)

[40] Hulse, E.O., Camponogara, E.: Robust formulations for production optimization of satellite oil wells. Engineering Optimization **49**(5), 846–863 (2017)

[41] Ichiishi, T.: A social coalitional equilibrium existence lemma. Econometrica **49**, 369–377 (1981)

[42] Jahn, J.: Multicriteria decision making, *International Series in Operations Research & Management Science*, vol. 21, chap. Theory of vector maximization: various concepts of efficient solutions, pp. 37–68. Springer US (1999)

[43] Jain, M., Kardes, E., Kiekintveld, C., Ordoñez, F., Tambe, M.: Security games with arbitrary schedules: A branch and price approach. In: Proceedings of the National Conference on Artificial Intelligence (AAAI). Atlanta, GA, USA (2010)

[44] Kaelbling, L., Littman, M., Moore, A.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research **4**, 237–285 (1996)

[45] Kalai, E.: Social Goals and Social Organization,, chap. Solutions to the bargaining problem, pp. 75–105. Cambridge, University Press (1985)

[46] Kalai, E., Smorodinsky, M.: Other solutions to nash's bargaining problem. Econometrica **43**(3), 513–518 (1975)

[47] Ke, S., Guo, D., Niu, Q., Huang, D.: Optimized production planning model for a multiplant cultivation system under uncertainty. Engineering Optimization **47**(2), 204–220 (2015)

[48] Kim, T., Glock, C.H.: Production planning for a two-stage production system with multiple parallel machines and variable production rates. Int. J. Production Economics **196**, 284–292 (2018)

[49] Konishi, H., Le Breton, M., Weber, S.: Equilibria in a model with partial rivalry. Journal of Economic Theory **72**, 225–237 (1997)

[50] Konishi, H., Le Breton, M., Weber, S.: Equivalence of strong and coalition-proof nash equilibria in games without spillovers. Economic Theory **9**, 97–113 (1997)

[51] Kubica, B.J., Wozniak, A.: Interval methods for computing strong nash equilibria of continuous games. Decision Making in Manufacturing and Services **9**(1), 63–78 (2015)

[52] Kumral, M.: Robust stochastic mine production scheduling. Engineering Optimization **42**(6), 567–579 (2010)

[53] Lemaire, B.: The proximal algorithm. International series of numerical mathematics **87**, 73–87 (1989)

[54] Machiavelli, N.: Discourses in the first ten books of Titus Livius. Duke University Press (1965)

[55] Machiavelli, N.: The Art of War. Da Capo Press (2001)

[56] Madani, K., Hipel, K.: Non-cooperative stability definitions for strategic analysis of generic water resources conflicts. Water Resources Managment **25**(8), 1949–1977 (2011)

[57] Martinet, B.: Breve communication. regularisation d'inequations variationnelles par approximations successives. ESAIM: Mathematical Modelling and Numerical Analysis **4**(3), 154–158 (1970)

[58] Merril, W., Schneider, N.: Government firms in oligopoly industries: a short-run analysis. Quaterly Journal of Economics **80**(3), 400–412 (1966)

[59] Moreno, F.G., Oliveira, P.R., Soubeyran, A.: A proximal algorithm with quasidistance. application to habit's formation. Optimization **61**, 1383–1403 (2011)

[60] Muthoo, A.: Bargaining theory with applications. Cambridge University Press (2002)

[61] Nash, J.F.: The bargaining problem. Econometrica **18**(2), 155–162 (1950)

[62] Nash, J.F.: Non-cooperative games. Annals of Mathematics **54**, 286–295 (1951)

[63] Nash, J.F.: Two person cooperative games. Econometrica **21**, 128–140 (1953)

[64] Nessah, R., Tian, G.: On the existence of strong nash equilibria. Journal of Mathematical Analysis and Applications **414**(2), 871–885 (2014)

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

[66] Osborne, M., Rubinstein, A.: Bargaining and Markets. Academic Press, Inc. (1990)

[67] Ostrom, E.: Governing the commons: The evolution of institutions for collective action. Cambridge University Press (1990)

[68] Ostrom, E.: A behavioral approach to the rational choice theory of collective action. The American Political Science Review **92**(1), 1–22 (1998)

[69] Ostrom, E., Gardner, R., Walker, J.: Rules, games, and common-pool resources. The University of Michigan Press (1994)

[70] Parikh, N., Boyd, S.: Proximal algorithms. Foundations and Trends in Optimization **1**(3), 123–231 (2014)

[71] Peters, H., Tijs, S.: Individually monotonic bargaining solutions for n-person bargaining games. Methods of Operations Research **51**, 377–384 (1984)

[72] Pita, J., Jain, M., Ordoñez, F., Portwa, C., Tambe, M., Western, C.: Using game theory for Los Angeles airport security. AI Magazine **30**(1), 43–57 (2009)

[73] Pita, J., Tambe, M., Kiekintveld, C., Cullen, S., Steigerwald, E.: GUARDS: game theoretic security allocation on a national scale. In: Proceedings of the The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), vol. 1, pp. 37–44. Taipei, Taiwan (2011)

[74] Poznyak, A.S.: Advance Mathematical Tools for Automatic Control Engineers. Deterministic Techniques, vol. 1. Elsevier, Amsterdam (2008)

[75] Poznyak, A.S.: Advance Mathematical Tools for Automatic Control Engineers. Stochastic Techniques, vol. 2. Elsevier, Amsterdam (2009)

[76] Poznyak, A.S., Najim, K., Gomez-Ramirez, E.: Self-learning control of finite Markov chains. Marcel Dekker, New York (2000)

[77] Raiffa, H.: Arbitration schemes for generalized two-person games. Annals of Mathematics Studies **28**, 361–387 (1953)

[78] Ribeiro, C.: Reinforcement learning agents. Artificial Intelligence Review **17**(3), 223–250 (2002)

[79] Rockafellar, R.: Monotone operators and the proximal point algorithm. SIAM journal on control and optimization **14**(5), 877–898 (1976)

[80] Roth, A.E.: An impossibility result converning n-person bargaining games. Int. Journal of Game Theory **8**(3), 129–132 (1979)

[81] Rozenfeld, O., Tennenholtz, M.: Strong and correlated strong equilibria in monotone congestion games. In: The 2nd Workshop on Internet & Network Economics (WINE 06), pp. 74–86 (2006)

[82] Rubinstein, A.: Strong perfect equilibrium in supergames. International Journal of Game Theory **9**(1), 1–12 (1980)

[83] Rubinstein, A.: Perfect equilibrium in a bargaining model. Econometrica **50**(1), 97–109 (1982)

[84] Rubinstein, A.: Finite automata play the repeated prisoner's dilemma. Journal of Economic Theory **39**(1), 83–96 (1986)

[85] Sakalaki, M., Kyriakopoulos, G., Kanellaki, S.: Are social representations consistent with social stategies? machiavellianism, opportunism and aspects of lay thinking. Hell. J. Psychol. **7**, 141–158 (2010)

[86] Salmeron, J., Wood, K., Baldick, R.: Analysis of electric grid security under terrorist threat. IEEE Transactions on Power Systems **19**(2), 905–912 (2004)

[87] Sánchez, E.M., Clempner, J.B., Poznyak, A.S.: A priori-knowledge/actor-critic reinforcement learning architecture for computing the mean-variance customer portfolio:The case of bank marmarket campaigns. Engineering Applications of Artificial Intelligence **46**, 82–92 (2015)

[88] Selbirak, T.: Some concepts of non-myopic equilibria in games with finite strategy sets and their properties. Annals of Operations Research **51**(2), 73–82 (1994)

[89] von Stackelberg, H.: Marktform und Gleichgewicht. Springer, Vienna (1934)

[90] von Stengel, B., Zamir, S.: Leadership games with convex strategy sets. Games and Economic Behavior **69**, 446–457 (2010)

[91] Tanaka, K.: The closest solution to the shadow minimum of a cooperative dynamic game. Computers & Mathematics with Applications **18**(1-3), 181–188 (1989)

[92] Tanaka, K., Yokoyama, K.: On ϵ-equilibrium point in a noncooperative n-person game. Journal of Mathematical Analysis and Applications **160**, 413–423 (1991)

[93] Tikhonov, A.N., Arsenin, V.Y.: Solution of Ill-posed Problems. Washington: Winston & Sons (1977)

[94] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: Numerical Methods for the Solution of Ill-Posed Problems. Kluwer Academic Publishers (1995)

[95] Trejo, K.K., Clempner, J.B.: New Perspectives and Applications of Modern Control Theory, chap. Setting Nash vs. Kalai-Smorodinsky bargaining approach: Computing the continuous-time controllable Markov game, pp. 335–369. Springer International Publishing (2018)

[96] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Computing the L_p-strong Nash equilibrium looking for cooperative stability in multiple agents Markov games. In: 12th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 309–314. Mexico City, Mexico (2015)

[97] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Computing the stackelberg/nash equilibria using the extraproximal method: convergence analysis and implementation details for markov chains games. International Journal of Applied Mathematics and Computer Science **25**(2), 337–351 (2015)

[98] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: A Stackelberg security game with random strategies based on the extraproximal theoretic approach. Engineering Applications of Artificial Intelligence **37**, 145–153 (2015)

[99] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Adapting strategies to dynamic environments in controllable Stackelberg security games. In: 55th IEEE Conference on Decision and Control (CDC), pp. 5484–5489 (2016)

[100] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: An optimal strong equilibrium solution for cooperative multi-leader-follower Stackelberg Markov chains games. Kybernetika **52**(2), 258–279 (2016)

[101] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Computing the strong l_p-nash equilibrium for markov chains games: convergence and uniqueness. Applied Mathematical Modelling **41**, 399–418 (2017)

[102] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Nash bargaining equilibria for controllable markov chains games. In: The 20th World Congress of the International Federation of Automatic Control (IFAC), vol. 50, pp. 12,261–12,266 (2017)

[103] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Adapting attackers and defenders patrolling strategies: A reinforcement learning approach for stackelberg security games. Journal of Computer and System Sciences **95**, 35–54 (2018)

[104] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Computing the bargaining approach for equalizing the ratios of maximal gains in continuous-time markov chains games. Computational Economics (2018). DOI 10.1007/s10614-018-9859-9. DOI: 10.1007/s10614-018-9859-9

[105] Trejo, K.K., Clempner, J.B., Poznyak, A.S.: Proximal constrained optimization approach with time penalization. Engineering Optimization (2018). DOI 10.1080/0305215X.2018.1519072. DOI: 10.1080/0305215X.2018.1519072

[106] Tsai, J., Rathi, S., Kiekintveld, C., Ordoñez, F., Tambe, M.: IRIS - a tool for strategic security allocation in transportation networks. In: Eighth International Conference on Autonomous Agents and Multiagent Systems - Industry Track, pp. 37–44 (2009)

[107] Wilson, D., Near, D., Miller, R.: Machiavellianism: A synthesis of the evolutionary and psychological literatures. Psychol. Bull. **119**(2), 285–299 (1996)

[108] Yin, G.G., Zhang, Q.: Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach. Springer (2013)

[109] Yin, K.K., Yin, G.G., Liu, H.: Stochastic modeling for inventory and production planning in the paper industry. AIChE Journal **50**(11), 2877–2890 (2004)

[110] Zangwill, W.I.: Nonlinear programming: A unified approach. Prentice-Hall, Englewood Cliffs (1969)

[111] Zangwill, W.I., Garcia, C.B.: Pathways to solutions, fixed points and equilibria. Prentice-Hall, Englewood Cliffs (1981)

[112] Zhang, Z., Shi, J., Chen, H., Guizani, M., P., Q.: A cooperation strategy based on nash bargaining solution in cooperative relay networks. IEEE Transactions on Vehicular Technology **57**(4), 2570–2577 (2008)