Article

Analytical Method for Mechanism Design in Partially Observable Markov Games

Julio B. Clempner ^{1,*} and Alexander S. Poznyak ²

¹ Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional, School of Physics and Mathematics, National Polytechnic Institute, Edificio 9 U.P. Adolfo Lopez Mateos, Col. San Pedro Zacatenco, 07730 Mexico City, Mexico

² Center for Research and Advanced Studies Av. IPN 2508, Department of Control Automatics, Col. San Pedro Zacatenco, 07360 Mexico City, Mexico; apozyak@ctrl.cinvestav.mx

* Correspondence: julio@clempner.name

Abstract: A theme that become common knowledge of the literature is the difficulty of developing a mechanism that is compatible with individual incentives that simultaneously result in efficient decisions that maximize the total reward. In this paper, we suggest an analytical method for computing a mechanism design. This problem is explored in the context of a framework, in which the players follow an average utility in a non-cooperative Markov game with incomplete state information. All of the Nash equilibria are approximated in a sequential process. We describe a method for the derivative of the player's equilibrium that instruments the design of the mechanism. In addition, it showed the convergence and rate of convergence of the proposed method. For computing the mechanism, we consider an extension of the Markov model for which it is introduced a new variable that represents the product of the mechanism design and the joint strategy. We derive formulas to recover the variables of interest: mechanisms, strategy, and distribution vector. The mechanism design and equilibrium strategies computation differ from those in previous literature. A numerical example presents the usefulness and effectiveness of the proposed method.

Keywords: dynamic mechanism design; incentive-compatible mechanisms; Markov games with private information; partially observable Markov chains; incomplete state information

Citation: Clempner, J.B.; Poznyak, A.S. Analytical Method for Mechanism Design in Partially Observable Markov Games. *Mathematics* **2021**, *9*, 321. <https://doi.org/10.3390/math9040321>

Received: 29 November 2020

Accepted: 11 January 2021

Published: 6 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Brief Review

Hurwicz [1] published his seminal work on mechanism design that has emerged as a practical framework for tackling game theory problems with an engineering viewpoint when considering players that interact rationally [2–4]. For a survey see [5]. This theory is based on games with incomplete information for modeling mechanisms (that implements a social choice function) compatible with individual incentives that result in efficient decisions for maximizing the total reward. The primary aim consists of establishing games that consider independent private values and quasilinear payoffs [6,7], in which players receive messages containing information that is relevant to payoffs [8]. In the evolutions of the game, players commit to a mechanism that presents a result in terms of a function of the possibly untruthfully reported type. It should be pointed out that the mechanism is unknown. The mechanism designer determines a social choice function that is a mapping of the true type profile directly to the alternatives. However, a mechanism maps the reported type profile to the alternatives. The main task in computational mechanism design is to find a mechanism that both maintains the game-theoretic original futures and is computationally “efficient” and “feasible”.

This approach makes it possible managing the restrictions and controlling the information of the players that are engaged in a game. From this perspective, Arrow [9] presented a framework to claim revelation that realizes efficiency and avoids the spend of resources

in the incentive payments. d'Aspremont and Gerard-Varet [10] suggested two separate methods to design a mechanism with incomplete information: a) the former consist of the fact that players beliefs are not considered, and the second where they are. Saari [11] presented a mechanism design, which involves types of information. Rogerson [12] proposed a general approach of the hold-up problem, in which several players make relation-specific investments and then decide on some cooperative action proving that first-best solutions exist under a variety of different assumptions regarding the nature of information asymmetries. Mailath and Postlewaite [13] established an approach for the bargaining problems with asymmetric information while considering multiple agents. Miyakawa [14] provided a necessary and sufficient condition for the existence of a stationary perfect Bayesian equilibrium. Athey and Bagwell [15] and Hörner et al. [16] have relevant results on equilibria in repeated games, which consider communication. Clempner and Poznyak [17] suggested a Bayesian partially observable Markov games model supported by an AI approach. Different approaches are presented in the literature, for instance, see [18–20].

1.2. Main Results

We contribute to this literature by proposing original outcomes, presenting an analytical method for developing a mechanism that considers incomplete state information whose preferences evolve following a Markov process, and characterizing an approximately equilibrium behavior in game theory models [17]. The foundation of the proposed method is the derivation of formulas for computing the mechanism μ . Subsequently, given the mechanism, compute the equilibrium strategy. The derivation of these formulas rely on a direct mechanism design. We propose an extension of the Markov model, suggesting a new variable z that represents the product of the mechanism μ and the joint strategy c . Additionally, the joint strategy c is defined by the product of the strategy π , the observer q , and the distribution vector P . We derive formulas to recover the variables of interest: mechanism μ , the strategies π , and the distribution vectors P . We describe a method for the derivative of the player's equilibrium that instruments the design of the mechanism and we also showed the convergence of the proposed method.

1.3. Organization of the Paper

For ease of exposition, in the next section, we describe the Markov game model. In Section 3, we introduce the variables c and z and suggest the derivation of the formulas. The ergodicity condition expressed in z variables is proven in Section 4. The convergence to a Nash equilibrium is presented in Section 5. Section 6 concludes with some remarks.

2. Markov Games with Incomplete Information

Let us introduce a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is a finite set of elementary events, $\mathcal{F}$ is the discrete σ -algebra of the subsets of Ω , and $\mathcal{P}$ is a given probability measure defined on $\mathcal{F}$. Let us also consider the natural sequence $t = 1, 2, \dots$ as a time argument. Let S be a finite set that consists of states $\{s_1, \dots, s_N\}$, $N \in \mathbb{N}$, called the state space. A Stationary Markov chain [21,22] is a sequence of S -valued random variables $s(t)$, $t \in \mathbb{N}$, satisfying the Markov condition:

$$\begin{aligned} P(s(t+1) = s_j | s(t) = s_i, s(t-1) = s_{i_{t-1}}, \dots, s(1) = s_{i_1}) = \\ P(s(t+1) = s_j | s(t) = s_i) =: p_{ji}. \end{aligned} \quad (1)$$

The random variables $s(t)$ are defined on the sample space Ω and they take values in S . The stochastic process $\{s(t), t \in \mathbb{N}\}$ is assumed to be a Markov chain. 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 in Equation (1). The matrix $P = (p_{ji})_{(s(i), s(j)) \in S} \in [0, 1]^{N \times N}$ determines the evolution of the chain: for each $n \in \mathbb{N}$, the power P^n has in each entry (s_i, s_j) the probability of going from state s_i to state s_j in exactly n steps.

Let $MC = (S, A, \{A(s)\}_{s \in S}, \mathbb{K}, P)$ be a Markov chain [21,22], where S is a finite set of states, $S \subset \mathbb{N}$ and A is a 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 $A = \cup_{s \in S} A(s)$. Whereas, $\mathbb{K} = \{(s, a) | s \in S, a \in A(s)\}$ is the set of admissible state-action pairs. The variable $p_{j|ik}$ is a stationary controlled transition matrix, where $p_{j|ik} := P(X_{t+1} = s_j | X_t = s_i, A_t = a_k) \forall t \in \mathbb{N}$ represents the probability that is associated with the transition from state s_i to state s_j , $i = \overline{1, N}$ ($i = 1, \dots, N$) and $j = \overline{1, N}$ ($j = 1, \dots, N$), under an action $a_k \in A(s_i)$, $k = \overline{1, K}$ ($k = 1, \dots, K$). The distribution vector is given by $P(X_t = s_i) = P_i$, such that $P_i \in \mathcal{S}^N$, where $\mathcal{S}^N = \{s \in \mathbb{R}^N : \sum_{i=1}^N P(s_i) = 1, P(s_i) \geq 0\}$.

We consider the case where the process is not directly observable [23]. Let us associate with S the observation set Y , which takes values in a finite space $\{1, \dots, M\}$, $M \in \mathbb{N}$. The stochastic process $\{Y_t, t \in \mathbb{N}\}$ is called the observation process. By observing Y_t at time t information regarding the true value of X_t is obtained. If $X_t = s_i$ and $A_t = a_k$ an observation $Y_t = y_m$ will have a probability $q_{m|ik} := P(Y_t = y_m | X_t = s_i, A_t = a_k)$, that denotes the relationship between the state and observation when an action $a_k \in A(s_i)$ is chosen at time t . The observation kernel is a stochastic kernel on Y , as given by $Q = [q_{m|ik}]$. We restrict ourselves to consider $Q = [q_{m|i}]$.

Definition 1. A controllable Partially Observable Markov Decision Process (POMDP) is a tuple

$$POMDP = \{MC, Y, Q, Q_0, P, V\}$$

where: i) MC is a Markov chain; ii) Y is the observation set, which takes values in a finite space $\{1, \dots, M\}$, $M \in \mathbb{N}$; iii) $Q = [q_{m|i}]_{m=\overline{1, M}, i=\overline{1, N}}$ denotes the observation kernel is a stochastic kernel on Y , such that $\sum_m q_{m|i} = 1$; iv) $Q_0 = [q_{m|i}]_{m=\overline{1, M}, i=\overline{1, N}}$ denotes the initial observation kernel; v) P is the (a priori) initial distribution; and, vi) V_{ijmk} , is the reward function at time t , given the state s_i , the observable state y_m , when the action $a_k \in A(s_i, y_m)$ is taken.

A realization of the partially observable system at time t is given by the sequence $(s_0, y_0, a_0, s_1, y_1, a_1, \dots) \in \Omega := (SYA)^\infty$, where s_0 has a given by the distribution $P(X_0 = s_0)$ and $\{A_t\}$ is a control sequence in A that is determined by a control policy. To define a policy we cannot use the (unobservable) states $s_0, s_1, \dots$. Then, we introduce the observable histories $h_0 := (p, Y_0) \in H_0$ and $h_t := (s_0, y_0, a_0, \dots, y_{t-1}, a_{t-1}, y_t) \in H_t$ for all $t \geq 1$ and $H_t := H_{t-1}(AY)$, if $t \geq 1$. Now, a policy is defined as a sequence $\{\pi_{k|m}(t)\}$, such that, for each t , $\pi_{k|m}(t)$ is a stochastic kernel on A given H_t . The set of all policies is denoted by Π . A policy $\pi_{k|m}(t) \in \Pi$ and an initial distribution $P(X_0 = s_0)$, also denoted as P_0 , determine all possible realizations of the POMDP. A control strategy satisfies that $\sum_k \pi_{k|m}(t) = 1$ and $\pi_{k|m}(t) \geq 0$, $m = 1, \dots, M$.

A game consists of a set $\mathcal{N} = \{1, \dots, n\}$ of players (indexed by $l = \overline{1, n}$). We employ l in order to emphasize the l -th player's variables and $-l$ subsumes all the other players' variables. The dynamics is described, as follows. At time $t = 0$, the initial state s_0 has a given a priori distribution P^l , and the initial observation y_0 is generated according to the initial observation kernel $Q_0^l(y_0 | s_0)$. If, at time t , the state of the system is X_t and the control $A_t^l \in A^l$ is applied, then each of strategy is allowed to randomize, with distribution $\pi_{k|m}^l(t)$, over the pure action choices $A_t^l \in A^l(X_t)$. These choices induce immediate utilities V_{ijmk}^l . The system tries to maximize the corresponding one-step utility. Next, the system moves to new state $X_{t+1} = s_j$, according to the transition probabilities $P^l(\pi_{k|m}^l(t))$. Subsequently, the observation Y_t is generated by the observation kernel $Q^l(Y_t | X_t)$. Based on the obtained utility, the systems adapt a mixed strategy computing $\pi_{k|m}^l(t+1)$ for the

next selection of the control actions. For any stationary strategies $\pi_{k|m}^l(t) = \pi_{k|m}^l$, we have $P_j^l = \sum_{i=1}^N (\sum_{k=1}^M p_{j|ik}^l \pi_{k|m}^l q_{m|i}^l) P_i^l$. Then,

$$U^l(\pi) := \sum_{m=1}^M \sum_{i=1}^N \sum_{k=1}^K W_{imk}^l \prod_{l=1}^n \pi_{k|m}^l q_{m|i}^l P_i^l,$$

where $W_{imk}^l = \sum_j^N V_{ijmk}^l p_{j|ik}^l$. Each player maximizes the individual payoff function $U^l(\pi_{k|m})$, realizing the rule that is given by

$$(\pi_{k|m}^{l*}) \in \text{Arg max}_{\pi^l \in \Pi^l} \sum_{l \in \mathcal{N}} U^l(\pi_{k|m}^{l*}). \quad (2)$$

where for a given strategies $\pi_{k|m}^*$ satisfy the Nash equilibrium [24,25] fulfilling, for all admissible $\pi_{k|m}$, the condition

$$U^l(\pi_{k|m}^{l*}) \geq U^l(\pi_{k|m}^l, \pi_{k|m}^{-l*}). \quad (3)$$

3. Main Relations

Following [21,26] and [27], let us introduce a matrix of elements $c = [c_{imk}]$, as follows

$$c_{imk}^l = d_{km}^l P_i^l = \pi_{k|m}^l q_{m|i}^l P_i^l.$$

Let us define $\Xi_{i|m}^l = (Q^l)^{-1}$, $Q^l = [q_{m|i}^l]$. Formally, a mechanism is any function $\mu_{k|m}$, such that given c_{imk}^l represents the nonlinear programming problem

$$\tilde{U}^l(c) = \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K W_{imk}^l \prod_{l=1}^n (\mu_{k'|m}^l c_{imk}^l) \rightarrow \max_{c^l \in C_{adm}^l, \mu_{k'|m}^l \in \mathcal{M}_{adm}^l}$$

and defining $\mu_{k'|m}^l = \mu_{k'|m} \forall l = 1, \dots, n$, we have that

$$\tilde{U}^l(c) = \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K W_{imk}^l \prod_{l=1}^n (\mu_{k'|m} c_{imk}^l) \rightarrow \max_{c^l \in C_{adm}^l, \mu_{k'|m} \in \mathcal{M}_{adm}} \quad (4)$$

such that

$$C_{adm}^l := \left\{ c_{imk}^l \geq 0 \mid \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K c_{imk}^l = 1, \sum_{m=1}^M \sum_{k=1}^K c_{imk}^l = P_i^l > 0, \right. \\ \sum_{m=1}^M \sum_{k=1}^K \sum_{i=1}^N [\delta_{ij} - p_{j|ik}^l] c_{imk}^l = 0, j = \overline{1, N}; \\ \sum_{h=1}^M \sum_{k=1}^K \sum_{i=1}^N [\delta_{hm} - q_{m|i}^l] c_{imk}^l = 0, m = \overline{1, M}, \\ \left. \sum_{m=1}^M \sum_{k=1}^K \sum_{i=1}^N \Xi_{h|m}^l c_{imk}^l \geq 0, h = \overline{1, N} \right\}, \quad (5)$$

$$\mathcal{M}_{adm} = \left\{ \mu_{k'|m} \geq 0 \mid \sum_{k'=1}^K \mu_{k'|m} = 1, m = \overline{1, M} \right\}. \quad (6)$$

Now, let us introduce the z-variable, as follows

$$z_{imkk'}^l := \mu_{k'|m} c_{imk}^l = \mu_{k'|m} \pi_{k|m}^l q_{m|i}^l P_i^l$$

$$\tilde{U}^l(z) = \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K W_{imk}^l \prod_{l=1}^n z_{imkk'}^l \rightarrow \max_{z \in Z_{adm}} \quad (7)$$

where

$$Z_{adm}^l := \left\{ z_{imkk'}^l \geq 0 \mid \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^l = 1, \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^l = P_i^l > 0, \right. \\ \sum_{m=1}^M \sum_{k=1}^K \sum_{i=1}^N [\delta_{ij} - p_{j|i}^l] \sum_{k'=1}^K z_{imkk'}^l = 0, j = \overline{1, N}; \sum_{h=1}^M \sum_{k=1}^K \sum_{i=1}^N [\delta_{hm} - q_{m|i}^l] \\ \left. \sum_{k'=1}^K z_{ihkk'}^l = 0, m = \overline{1, M}, \sum_{m=1}^M \sum_{k=1}^K \sum_{i=1}^N \Xi_{h|m}^l \sum_{k'=1}^K z_{imkk'}^l \geq 0, h = \overline{1, N}, \right\} \quad (8)$$

Notice that by the relations

$$\sum_{k'=1}^K \mu_{k'|m} = 1, \quad \sum_{k=1}^K \pi_{k|m}^l = 1, \quad \sum_{m=1}^M q_{m|i}^l = 1, \quad \sum_{i=1}^N P_i^l = 1$$

$\mu_{k'|m} \geq 0$, it is easy to check that $z^l \in \mathcal{S}^l$, where

$$\mathcal{S}^l := \left\{ z_{imkk'}^l \geq 0 \mid \sum_{i=1}^N \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^l = 1, \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^l = P_i^l > 0 \right\} \quad (9)$$

We define the solution of the problem (7) as z^{l*} . The next lemma clarifies how we may recover $\mu_{k'|m}^*$ and c_{imk}^{l*} .

Lemma 1. Variables $\mu_{k'|m}^*$ and c_{imk}^{l*} can be recovered from $z_{imkk'}^{l*}$, as follows:

$$(a) \quad \mu_{k'|m}^* = \frac{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{imkk'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{imk\kappa}^{l*}} \quad (b) \quad c_{\alpha\beta\gamma}^{l*} = \frac{1}{K} \sum_{k'=1}^K \frac{z_{\alpha\beta\gamma k'}^{l*}}{\mu_{k'|m}^*} \quad (10)$$

Proof. See Appendix A. $\square$

Corollary 1. In addition, $c_{\alpha\beta\gamma}^{l*} = \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^{l*}$.

Now, in order to derive $\pi_{k|m}^{l*}$ and $\bar{P}_m^{l*}$ we have that

$$\pi_{kim}^{l*} = \begin{cases} \frac{\sum_{k'=1}^K z_{imkk'}^{l*}}{\sum_{h=1}^M \sum_{k'=1}^K z_{imhk'}^{l*}} & \text{if } \sum_{h=1}^M \sum_{k'=1}^K z_{imhk'}^{l*} > 0 \\ 0 & \text{if } \sum_{h=1}^M \sum_{k'=1}^K z_{imhk'}^{l*} = 0, \end{cases} \quad (11)$$

$$P_i^{l*} = \sum_{m=1}^M \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^{l*} \quad (12)$$

Corollary 2. The strategy $\pi_{k|m}^{l*}$ constructed from π_{kim}^{l*} (11), and the distribution $\bar{P}_m^{l*}$ are given by

$$\pi_{k|m}^{l*} = \frac{1}{N} \sum_{i=1}^N \pi_{kim}^{l*}, \quad \bar{P}_m^{l*} = \sum_{i=1}^N \sum_{k=1}^K \sum_{k'=1}^K z_{imkk'}^{l*}. \quad (13)$$

4. Ergodicity Conditions Expressed in z Variables

We have derived the formulas, which maximize Equation (7) that is based on the variables $z_{imkk'}^{l*}$ and the formulas to recover the policy $\pi_{k|m}^{l*}$, the mechanism $\mu_{k'|m}^*$ and $\bar{P}_m^{l*}$. Accordingly, we focus our attention on the ergodicity restrictions.

Theorem 1. The strategy $\pi_{k|m}^{l*}$ and the mechanism $\mu_{k'|m}^{l*}$ are in Nash equilibrium, where every agent maximizes its expected utility, for every $l = \overline{1, n}$,

$$\tilde{U}^l(\mu_{k'|m}^{l*} \pi_{k|m}^{l*} q_{m|i}^{l*} P_i^{l*}) \geq \tilde{U}^l(\mu_{k'|m}^l \pi_{k|m}^l q_{m|i}^l P_i^l)$$

if the quantities of $z_{imkk'}^l$ satisfies the following restrictions

$$\sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N [\delta_{\alpha j} - p_{j|\alpha\gamma}^l] \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l = 0, j = \overline{1, N}. \quad (14)$$

$$\sum_{h=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l = 0, \beta = \overline{1, M}. \quad (15)$$

$$\sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N \Xi_{h|\beta}^l \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l \geq 0, h = \overline{1, N}. \quad (16)$$

Proof. See Appendix B. $\square$

5. Convergence Analysis

The Nash Equilibrium is a game theory concept that involves several players that determines the solution in a non-cooperative game in which each player lacks any incentive to only change his/her own strategy. A practical notion in deriving Nash equilibria is a player's best reply. The best reply is the strategy (or set of strategies) that maximizes/minimizes his/her payoff taking other players' strategies as given. Then, a player has not just one best-reply strategy, however he/she has a best-reply strategy for each arrangement of strategies for the other players. All of the Nash's equilibrium can be approximated in a (best reply) sequential process. We want to compute the solution of the problem (7), defined as $z^l = \mu_{k'|m}^{l*} \pi_{k|m}^{l*} q_{m|i}^{l*} P_i^{l*}$, when considering the best reply approach. For solving the problem (7), let us consider a game whose strategies are denoted by $x^l \in X^l$, where X is a convex and compact set, where $x^l := \text{col}(z_{imkk'}^l)$ and $X^l := Z_{adm}^l$. Let $x = (x^1, \dots, x^n)^\top \in X$ be the joint strategy of the players and $x^{\hat{l}} := (x^1, \dots, x^{l-1}, x^{l+1}, \dots, x^n)^\top \in X^{\hat{l}}$ be a strategy of the rest of the players adjoint to $x^l \in X_{adm}^l$. We consider a Nash equilibrium problem with n players and denote, by $x = (x^l, x^{\hat{l}}) \in \mathbb{R}^n$, the vector representing the x -th player's strategy $X_{adm} = X_{adm}^l \times X_{adm}^{\hat{l}}$. The method of Lagrange multipliers is an optimization approach for finding the local minimum (maximum) of a function subject to equality constraints (A_{eq}), as given in Equation (8). Let us consider the Lagrange function that is given by

$$\mathbb{L}(x, \hat{x}(x), \lambda) := U(x, \hat{x}(x)) + \lambda^\top A_{eq} x$$

where the Lagrange vector-multipliers $\lambda \in \Lambda$ may have any sign. The optimization problem

$$\mathbb{L}(x, \hat{x}(x), \lambda) \rightarrow \min_{x \in X_{adm}, \hat{x}(x) \in X_{adm}} \max_{\lambda \in \Lambda}$$

for which we propose the following iteration algorithm:

1. Proximal prediction step:

$$\begin{aligned} \bar{\lambda}_n &= \arg \min_{\lambda \geq 0} \left\{ -\frac{1}{2} \|\lambda - \lambda_n\|^2 + \theta \mathbb{L}(x_n, \hat{x}_n, \lambda) \right\} \\ \bar{x}_n &= \arg \min_{x \in X} \left\{ \frac{1}{2} \|x - x_n\|^2 + \theta \mathbb{L}(x, \hat{x}_n, \bar{\lambda}_n) \right\} \\ \bar{\hat{x}}_n &= \arg \min_{\hat{x} \in \hat{X}} \left\{ \frac{1}{2} \|\hat{x} - \hat{x}_n\|^2 + \theta \mathbb{L}(x_n, \hat{x}, \bar{\lambda}_n) \right\} \end{aligned} \quad (17)$$

2. Gradient approximation step:

$$\begin{aligned}\lambda_{n+1} &= \arg \min_{\lambda \geq 0} \left\{ -\frac{1}{2} \|\lambda - \lambda_n\|^2 + \theta \mathbb{L}(\bar{x}_n, \bar{x}_n, \lambda) \right\} \\ x_{n+1} &= \arg \min_{x \in X} \left\{ \frac{1}{2} \|x - x_n\|^2 + \theta \mathbb{L}(x, \bar{x}_n, \bar{\lambda}_n) \right\} \\ \hat{x}_{n+1} &= \arg \min_{\hat{x} \in \hat{X}} \left\{ \frac{1}{2} \|\hat{x} - \hat{x}_n\|^2 + \theta \mathbb{L}(\bar{x}_n, \hat{x}, \bar{\lambda}_n) \right\}\end{aligned}\quad (18)$$

Let us define the following variables

$$\tilde{x} := \begin{pmatrix} x \\ \hat{x} \end{pmatrix} \in \tilde{X} := X \times \hat{X}, \quad \tilde{y} := (\lambda) \in \tilde{Y} := \mathbb{R}^+$$

Subsequently, the Lagrange function can be expressed as

$$\tilde{\mathcal{L}}(\tilde{x}, \tilde{y}) := \mathbb{L}(x, \hat{x}, \lambda)$$

The equilibrium point that satisfies Equations (17) and (18) can be represented by

$$\begin{aligned}\tilde{x}^* &= \arg \min_{\tilde{x} \in \tilde{X}} \left\{ \frac{1}{2} \|\tilde{x} - \tilde{x}^*\|^2 + \gamma \tilde{\mathcal{L}}(\tilde{x}, \tilde{y}^*) \right\} \\ \tilde{y}^* &= \arg \max_{\tilde{y} \in \tilde{Y}} \left\{ \frac{1}{2} \|\tilde{y} - \tilde{y}^*\|^2 + \gamma \tilde{\mathcal{L}}(\tilde{x}^*, \tilde{y}) \right\}\end{aligned}$$

In addition, let us introduce the following variables

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

and let us define the Lagrangian in terms of the previous variables

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

For $\tilde{w}_1 = \tilde{x}$, $\tilde{w}_2 = \tilde{y}$, $\tilde{v}_1 = \tilde{v}_1^* = \tilde{x}^*$ and $\tilde{v}_2 = \tilde{v}_2^* = \tilde{y}^*$, we have

$$L(\tilde{w}, \tilde{v}^*) := \tilde{\mathcal{L}}(\tilde{x}, \tilde{y}^*) - \tilde{\mathcal{L}}_\delta(\tilde{x}^*, \tilde{y})$$

In these variables, the relation Equations (17) and (18) can be represented by

$$\tilde{v}^* = \arg \min_{\tilde{v} \in \tilde{X} \times \tilde{Y}} \left\{ \frac{1}{2} \|\tilde{w} - \tilde{v}^*\|^2 + \gamma L(\tilde{w}, \tilde{v}^*) \right\} \quad (19)$$

We provide the convergence analysis of the sequence $\{v_n\}_{n \in \mathbb{N}}$ in the following theorem [28].

Theorem 2. Let $\mathbb{L}(w, v)$ be a convex and differentiable function with the gradient satisfying the Lipschitz condition, i.e., $\|\nabla \mathbb{L}(v) - \nabla \mathbb{L}(w)\| \leq L\|v - w\|$ for all $v, w \in V_{adm}$, where V_{adm} is a convex and compact set. Let $\{v_n\}_{n \in \mathbb{N}}$ be a sequence defined by the local search and proximal iteration algorithm that is given by

$$\begin{aligned}\bar{v}_n &= \arg \min_{v \in V_{adm}} \left\{ \frac{1}{2} \|v - v_n\|^2 + \theta_n L(w, v) \right\} \\ v_{n+1} &= \arg \min_{v \in V_{adm}} \left\{ \frac{1}{2} \|v - v_n\|^2 + \theta_n L(\bar{w}_n, \bar{v}_n) \right\}\end{aligned}\quad (20)$$

then, the sequence $\{v_n\}_{n \in \mathbb{N}}$ converges to a Nash equilibrium point $v^* \in V_{adm}$.

Proof. See Appendix C. $\square$

6. Political Numerical Example

The theory that is related to electoral competition originates in the original contributions of Hotelling [29] and Downs [30]. The proposed framework suggests a majority rule election, where political candidates compete for a position by simultaneously and independently proposing a model from a unidimensional policy space. It is common knowledge that the equilibrium of this model is fundamentally determined on the candidates' incentives for running for such a position. This example considers a three-player game ($l = \overline{1,3}$) that is engaged in a political contest, in which the player with the highest performance wins. A question arises as to what is the design of a mechanism to select a candidate? The goal of each candidate is to end up on top. The next time a political position rolls around, pay attention to the campaigning. Candidates who are behind will talk about not only what a good choice they are for such a position, but also what a bad choice the front-runner is.

The assumption that is involved in this example considers the incomplete information version of the game, in which candidates have the same relative weight to their preference strategies versus their desire to win the position. This case is relevant from a theoretical point of view, and it is empirically important. The dynamics are modeled when considering $N = 4$, $M = 4$, and $K = 2$ with transition matrices for describing the evolution of the partially observed Markov game. The initial transition matrices are defined, as follows:

$$p_{j|i1}^{1*} = \begin{bmatrix} 0.2307 & 0.2130 & 0.3120 & 0.2443 \\ 0.4989 & 0.1075 & 0.2250 & 0.1686 \\ 0.1857 & 0.0801 & 0.3938 & 0.3404 \\ 0.3235 & 0.2372 & 0.1065 & 0.3327 \end{bmatrix} \quad p_{j|i2}^{1*} = \begin{bmatrix} 0.4344 & 0.0728 & 0.3970 & 0.0959 \\ 0.5605 & 0.1505 & 0.1423 & 0.1467 \\ 0.1866 & 0.2866 & 0.3168 & 0.2100 \\ 0.1140 & 0.2091 & 0.2878 & 0.3892 \end{bmatrix}$$

$$p_{j|i1}^{2*} = \begin{bmatrix} 0.1613 & 0.4207 & 0.1745 & 0.2435 \\ 0.3372 & 0.1160 & 0.2305 & 0.3163 \\ 0.2488 & 0.3219 & 0.0323 & 0.3970 \\ 0.3696 & 0.5068 & 0.0363 & 0.0873 \end{bmatrix} \quad p_{j|i2}^{2*} = \begin{bmatrix} 0.3587 & 0.1023 & 0.1044 & 0.4346 \\ 0.1796 & 0.3039 & 0.2303 & 0.2862 \\ 0.0115 & 0.3002 & 0.2537 & 0.4346 \\ 0.2102 & 0.3296 & 0.4079 & 0.0523 \end{bmatrix}$$

$$p_{j|i1}^{3*} = \begin{bmatrix} 0.1354 & 0.3183 & 0.0631 & 0.4832 \\ 0.2442 & 0.2663 & 0.2572 & 0.2323 \\ 0.4768 & 0.2447 & 0.0144 & 0.2641 \\ 0.3380 & 0.1812 & 0.3172 & 0.1636 \end{bmatrix} \quad p_{j|i2}^{3*} = \begin{bmatrix} 0.1201 & 0.0840 & 0.4017 & 0.3942 \\ 0.3531 & 0.1164 & 0.2559 & 0.2746 \\ 0.2919 & 0.0985 & 0.3721 & 0.2375 \\ 0.5341 & 0.0798 & 0.0850 & 0.3010 \end{bmatrix}$$

As well as, the initial observation matrices are defined, as follows:

$$q_{m|i}^{1*} = \begin{bmatrix} 0.2423 & 0.3692 & 0.0931 & 0.3550 \\ 0.1881 & 0.0397 & 0.2043 & 0.2126 \\ 0.2848 & 0.3933 & 0.1156 & 0.3773 \\ 0.2847 & 0.1978 & 0.5870 & 0.0552 \end{bmatrix} \quad q_{m|i}^{2*} = \begin{bmatrix} 0.1749 & 0.4870 & 0.2563 & 0.4140 \\ 0.2719 & 0.0190 & 0.2748 & 0.4030 \\ 0.2814 & 0.4030 & 0.2405 & 0.1692 \\ 0.2718 & 0.0910 & 0.2284 & 0.0137 \end{bmatrix}$$

$$q_{m|i}^{3*} = \begin{bmatrix} 0.1734 & 0.0374 & 0.0430 & 0.3581 \\ 0.4350 & 0.2568 & 0.4207 & 0.0150 \\ 0.2747 & 0.3091 & 0.3390 & 0.3458 \\ 0.1170 & 0.3967 & 0.1973 & 0.2811 \end{bmatrix}$$

Fixing $\theta = 0.055$ in the extraproximal method that is given in Equations (17) and (18), we have that the Nash equilibrium results from computing the strategies and the mechanism design applying Equations (10) and (13), which are given, as follows:

$$\pi_{k|m}^{1*} = \begin{bmatrix} 0.4633 & 0.5367 \\ 0.4543 & 0.5457 \\ 0.4850 & 0.5150 \\ 0.3859 & 0.6141 \end{bmatrix} \quad \pi_{k|m}^{2*} = \begin{bmatrix} 0.3092 & 0.6908 \\ 0.3707 & 0.6293 \\ 0.3879 & 0.6121 \\ 0.3863 & 0.6137 \end{bmatrix}$$

$$\pi_{k|m}^{3*} = \begin{bmatrix} 0.3937 & 0.6063 \\ 0.4165 & 0.5835 \\ 0.3850 & 0.6150 \\ 0.4990 & 0.5010 \end{bmatrix} \quad \mu_{k|m}^* = \begin{bmatrix} 0.2929 & 0.7071 \\ 0.3321 & 0.6679 \\ 0.2410 & 0.7590 \\ 0.2187 & 0.7813 \end{bmatrix}$$

We present a full characterization of the Nash equilibrium for the case of partially observable Markov games. Figures 1–3 show the convergence of the strategies $z_{imkk'}^{l*}$.

Figure 1. Convergence of strategies $z_{imkk'}^{l*}$ for player 1.

Figure 2. Convergence of strategies $z_{imkk'}$ for player 2.

Figure 3. Convergence of strategies $z_{imkk'}$ for player 3.

7. Conclusions

This paper contributed to the literature on mechanism design for Markov games with incomplete state information (partially observable). We suggested an analytical method for the design of a mechanism. The main result of this work is based on the introduction of the new variable z , which makes the game problem computationally tractable and allow for obtaining the mechanism solution μ and the strategies π for all of the players in the game. The variable z allows for the introduction of new natural additional linear restrictions for computing the Nash equilibrium of the game. A no feasible solution can be detected with a simple test on the variable z , i.e., it is possible to detect unusual conditions in the solver of the game given the information available for the simplex. A major advantage of introducing this variable relies on the fact it can be efficiently implemented for real settings, which is consistent with the engineering approach for designing economic mechanisms or incentives, toward desired objectives, where players act rationally. We applied these results to a numerical example that is related to political promotion.

In relation to future work, there are several challenges that are left to address. One interesting technical challenge is that of addressing extremum seeking in the context of mechanism design [31–33]. Another interesting challenge would be to consider the observer design approach in order to extend the mechanism design theory [23].

Author Contributions: Authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof Lemma 3.1

Proof. Since $\mu_{k'|m}^*$ does not depend on indices l, i, k' it may be obtained from Equations (8) and (9):

$$\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{imkk'}^{l*} := \mu_{k'|m}^* \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K c_{imk}^{l*} = \mu_{k'|m}^* \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \left(\pi_{k|m}^{l*} q_{m|i}^l p_i^{l*} \right) = \mu_{k'|m}^* \sum_{l=1}^n \sum_{i=1}^N \left(q_{m|i}^l p_i^{l*} \right)$$

Hence,

$$\mu_{k'|m}^* = \frac{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{imkk'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \left(q_{m|i}^l p_i^{l*} \right)} = \frac{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{imkk'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{imk\kappa}^{l*}}.$$

Let us define $z_{\alpha\beta\gamma k'}^{l*}$ as follows

$$c_{\alpha\beta\gamma}^{l*} = \frac{1}{K} \sum_{k'=1}^K \frac{z_{\alpha\beta\gamma k'}^{l*}}{\mu_{k'|m}^*} = \frac{1}{K} \sum_{k'=1}^K \left(\frac{z_{\alpha\beta\gamma k'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{l*}} \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{l*} \right) \quad (A1)$$

To verify that the definitions of $\mu_{k'|m}^*$ and c_{imk}^{l*} (10) are correct we need to check the fulfilling of Equations (5) and (6), i.e., $\mu_{k'|m}^* \in \mathcal{M}_{adm}$ and $c_{imk}^{l*} \in C_{adm}^l$.

(a) As for variables $\mu_{k'|m}^*$, these properties follow directly:

$$\mu_{k'|m}^* = \frac{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{imkk'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{imk\kappa}^{l*}} \geq 0 \quad (\text{A2})$$

since $z_{imkk'}^{l*} \geq 0$. Summing (A2) by k' directly leads to the property $\sum_{k'=1}^K \mu_{k'|m}^* = 1$.

(b) To prove that $c_{imk}^{l*} \in C_{adm}^l$ defined by (A1) notice that

$$\begin{aligned} \sum_{\gamma=1}^K c_{\alpha\beta\gamma}^{l*} &= \frac{1}{K} \sum_{\gamma=1}^K \sum_{k'=1}^K \left(\frac{z_{\alpha\beta\gamma k'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{l*}} \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{l*} \right) = \\ &= \frac{1}{K} \sum_{k'=1}^K \left(\frac{\sum_{\gamma=1}^K z_{\alpha\beta\gamma k'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{l*}} \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{l*} \right) \end{aligned}$$

and

$$\sum_{\beta=1}^M \sum_{\gamma=1}^K c_{\alpha\beta\gamma}^{l*} = \frac{1}{K} \sum_{k'=1}^K \sum_{\beta=1}^M \left(\frac{\sum_{\gamma=1}^K z_{\alpha\beta\gamma k'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{l*}} \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{l*} \right),$$

which leads to following relation

$$\begin{aligned} \sum_{\alpha=1}^N \sum_{\beta=1}^M \sum_{\gamma=1}^K c_{\alpha\beta\gamma}^{l*} &= \frac{1}{K} \sum_{k'=1}^K \sum_{\beta=1}^M \sum_{\alpha=1}^N \left(\frac{\sum_{\gamma=1}^K z_{\alpha\beta\gamma k'}^{l*}}{\sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{l*}} \sum_{h=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{h*} \right) = \\ &= \frac{1}{K} \sum_{\beta=1}^M \sum_{k'=1}^K \left(\frac{\sum_{\alpha=1}^N \sum_{\gamma=1}^K z_{\alpha\beta\gamma k'}^{l*}}{\sum_{h=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{h*}} \sum_{h=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{h*} \right) = \\ &= \frac{1}{K} \sum_{\beta=1}^M \left[\left(\sum_{h=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{i\beta k \kappa}^{h*} \right) \sum_{k'=1}^K \frac{\sum_{\alpha=1}^N \sum_{\gamma=1}^K z_{\alpha\beta\gamma k'}^{l*}}{\sum_{h=1}^n \sum_{i=1}^N \sum_{k=1}^K z_{i\beta k k'}^{h*}} \right] = \sum_{\beta=1}^M q_{i|\beta}^l P_i^l \frac{\sum_{\alpha=1}^N q_{\alpha|\beta}^h P_{\alpha}^l}{\sum_{h=1}^n \sum_{i=1}^N \sum_{\alpha=1}^N q_{i|\beta}^l P_i^l} = \\ &= \sum_{\alpha=1}^N \left[\sum_{\beta=1}^M q_{\alpha|\beta}^h \right] P_{\alpha}^h = \sum_{\alpha=1}^N P_{\alpha}^h = 1 \end{aligned}$$

Then, $z_{\alpha\beta\gamma k'}^{l*} \in \mathcal{S}$, see Equation (9). The Lemma is proved.

Appendix B. Proof of Theorem 4.1

Proof. This means that new variables $z_{\alpha\beta\theta k'}^l$ should satisfy the following linear ergodicity constraints:

$$\begin{aligned} P_j^l &= \sum_{\alpha=1}^N \left(\sum_{\beta=1}^M \sum_{\gamma=1}^K p_{j|\alpha\gamma}^l \mu_{k'|\beta}^l \pi_{\gamma|\beta}^l q_{\beta|\alpha}^l \right) P_{\alpha}^l = \sum_{\alpha=1}^N \left(\sum_{\beta=1}^M \sum_{\gamma=1}^K p_{j|\alpha\gamma}^l z_{\alpha\beta\gamma k'}^l \right), \\ &= \sum_{j=1}^N \left(\sum_{\beta=1}^M \sum_{\gamma=1}^K p_{j|\alpha\gamma}^l z_{j\beta\gamma k'}^l \right) = \sum_{\alpha=1}^N \left(\sum_{\beta=1}^M \sum_{\gamma=1}^K p_{j|\alpha\gamma}^l z_{\alpha\beta\gamma k'}^l \right), \end{aligned}$$

which implies

$$\sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N [\delta_{\alpha j} - p_{j|\alpha\gamma}^l] \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l = 0, \quad j = \overline{1, N}$$

Then,

$$z \in \mathcal{E} := \left\{ z_{\alpha\beta\gamma k'}^l \mid \sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N [\delta_{\alpha j} - p_{j|\alpha\gamma}^l] \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l = 0, j = \overline{1, N}, \right\} \quad (\text{A3})$$

The Equation (15) is fulfilled automatically since

$$\begin{aligned} & \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{ihk\kappa}^l \sum_{k'=1}^K \frac{\mu_{k'|h} \sum_{\gamma=1}^K \pi_{\gamma|h}^l q_{\alpha|h}^l P_{\alpha}^l}{\mu_{k'|h} \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \pi_{k|h}^l q_{i|h}^l P_i^l} = \\ & \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{ihk\kappa}^l \sum_{k'=1}^K \frac{\left(\sum_{\gamma=1}^K \pi_{\gamma|h}^l \right) q_{\alpha|h}^l P_{\alpha}^l}{\sum_{l=1}^n \sum_{i=1}^N \left(\sum_{k=1}^K \pi_{k|h}^l \right) q_{i|h}^l P_i^l} = \\ & K \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{l=1}^n \sum_{i=1}^N \sum_{k=1}^K \sum_{\kappa=1}^K z_{ihk\kappa}^l \frac{q_{\alpha|h}^l P_{\alpha}^l}{\sum_{l=1}^n \sum_{i=1}^N q_{i|h}^l P_i^l} = \\ & K \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{l=1}^n \sum_{i=1}^N \left(\sum_{k=1}^K \left(\sum_{\kappa=1}^K \mu_{\kappa|h} \right) \pi_{k|h}^l \right) q_{i|h}^l P_i^l \frac{q_{\alpha|h}^l P_{\alpha}^l}{\sum_{l=1}^n \sum_{i=1}^N q_{i|h}^l P_i^l} = \\ & = K \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \sum_{l=1}^n \sum_{i=1}^N q_{i|h}^l P_i^l \frac{q_{\alpha|h}^l P_{\alpha}^l}{\sum_{l=1}^n \sum_{i=1}^N q_{i|h}^l P_i^l} = K \sum_{h=1}^M \sum_{\alpha=1}^N [\delta_{h\beta} - q_{\alpha|\beta}^l] \left(q_{\alpha|h}^l P_{\alpha}^l \right) = \\ & K \left[\sum_{h=1}^M \sum_{\alpha=1}^N \delta_{h\beta} \left(q_{\alpha|h}^l P_{\alpha}^l \right) - \sum_{h=1}^M \sum_{\alpha=1}^N q_{\alpha|\beta}^l \left(q_{\alpha|h}^l P_{\alpha}^l \right) \right] = \\ & K \left[\sum_{h=1}^M \sum_{\alpha=1}^N \delta_{h\beta} q_{\alpha|h}^l P_{\alpha}^l - \sum_{\alpha=1}^N q_{\alpha|\beta}^l \left(\sum_{h=1}^M q_{\alpha|h}^l \right) P_{\alpha}^l \right] = K \left[\sum_{\alpha=1}^N \left(q_{\alpha|\beta}^l P_{\alpha}^l \right) - \sum_{\alpha=1}^N q_{\alpha|\beta}^l P_{\alpha}^l \right] = 0 \end{aligned}$$

Now, we prove the relation given in Equation (16) as follows:

$$\begin{aligned} \sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N \Xi_{h|\beta}^l \sum_{k'=1}^K z_{\alpha\beta\gamma k'}^l &= \sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N \Xi_{h|\beta}^l c_{\alpha\beta\gamma}^l = \sum_{\beta=1}^M \sum_{\gamma=1}^K \sum_{\alpha=1}^N \Xi_{h|\beta}^l \pi_{\gamma|\beta}^l q_{\alpha|\beta}^l P_{\alpha}^l = \\ & \sum_{\alpha=1}^N P_{\alpha}^l \sum_{\beta=1}^M q_{\alpha|\beta}^l \Xi_{h|\beta}^l \sum_{\gamma=1}^K \pi_{\gamma|\beta}^l = \sum_{\alpha=1}^N P_{\alpha}^l \delta_{\alpha,h} = P_h^l \geq 0. \end{aligned}$$

The Theorem is proved.

Appendix C. Proof of Theorem 5.1

Let us define $\omega_0 = \sum_{l=1}^{|\mathcal{N}|} \omega_{0,l} \leq |\mathcal{N}| \max_{l=1, \dots, |\mathcal{N}|} \omega_{0,l} = |\mathcal{N}| \omega_0^+$ and $\tilde{w} = \tilde{v}_{n+1}$, then

$$\frac{1}{2} \|\hat{v}_n - \tilde{v}_n\|^2 + \theta L(\hat{v}_n, \tilde{v}_n) \leq \frac{1}{2} \|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \theta L(\tilde{v}_{n+1}, \tilde{v}_n) - \frac{1}{2} \|\tilde{v}_{n+1} - \hat{v}_n\|^2 \quad (\text{A4})$$

Let also $\tilde{w} = \tilde{v}^* \in \tilde{X}^* \times \tilde{Y}^*$ then

$$\frac{1}{2} \|\tilde{v}_{n+1} - \tilde{v}_n\|^2 + \theta L(\tilde{v}_{n+1}, \hat{v}_n) \leq \frac{1}{2} \|\tilde{v}^* - \tilde{v}_n\|^2 + \theta L(\tilde{v}^*, \hat{v}_n) - \frac{1}{2} \|\tilde{v}^* - \tilde{v}_{n+1}\|^2 \quad (\text{A5})$$

Adding Equations (A4) and (A5) and multiplying by two yields

$$\|\tilde{v}^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 - 2\theta L(\tilde{v}^*, \hat{v}_n) + 2\theta [L(\tilde{v}_{n+1}, \hat{v}_n) + L(\hat{v}_n, \tilde{v}_n) - L(\tilde{v}_{n+1}, \tilde{v}_n)] \leq \|\tilde{v}^* - \tilde{v}_n\|^2 \quad (\text{A6})$$

Adding and subtracting $L(\hat{v}_n, \hat{v}_n)$ in Equation (A6) we have

$$\begin{aligned} & \|\tilde{v}^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 + 2\theta[L(\hat{v}_n, \hat{v}_n) - L(\tilde{v}^*, \hat{v}_n)] + 2\theta[L(\tilde{v}_{n+1}, \hat{v}_n) - \\ & L(\hat{v}_n, \hat{v}_n) + L(\hat{v}_n, \tilde{v}_n) - L(\tilde{v}_{n+1}, \tilde{v}_n)] \leq \|\tilde{v}^* - \tilde{v}_n\|^2 \end{aligned} \quad (A7)$$

Let $\tilde{w} + h = \tilde{v}_{n+1}$, $\tilde{w} = \hat{v}_n$, $\tilde{v} + k = \tilde{v}_n$, $\tilde{v} = \hat{v}_n$ having $h = \tilde{v}_{n+1} - \hat{v}_n$ and $k = \tilde{v}_n - \hat{v}_n$. The inequality (A7) becomes

$$\begin{aligned} & \|\tilde{v}^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \|\hat{v}_n - \tilde{v}_n\|^2 + 2\theta[L(\hat{v}_n, \hat{v}_n) - L(\tilde{v}^*, \hat{v}_n)] - \\ & 2\theta\omega\|\tilde{v}_{n+1} - \hat{v}_n\|\|\tilde{v}_n - \hat{v}_n\| \leq \|\tilde{v}^* - \tilde{v}_n\|^2 \end{aligned}$$

Using $L(\hat{v}_n, \hat{v}_n) - L(\tilde{v}^*, \hat{v}_n) \geq \|\hat{v}_n - \tilde{v}^*\|^2$ we obtain $\|\tilde{v}^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + 2\theta\|\hat{v}_n - \tilde{v}^*\|^2 + \left(1 - 2\theta^2\omega^2\right)\|\tilde{v}_n - \hat{v}_n\|^2 \leq \|\tilde{v}^* - \tilde{v}_n\|^2$. Now, by the fact

$$2\langle a - c, c - b \rangle = \|a - b\|^2 - \|a - c\|^2 - \|c - b\|^2$$

and replacing $a = \hat{v}_n$, $b = \tilde{v}^*$ and $c = \tilde{v}_n$, to the left-hand side of the last inequality we have

$$\begin{aligned} & \|\tilde{v}^* - \tilde{v}_{n+1}\|^2 + \|\tilde{v}_{n+1} - \hat{v}_n\|^2 + \left(1 - 2\theta^2\omega^2\right)\|\tilde{v}_n - \hat{v}_n\|^2 + 2\theta[2\langle \hat{v}_n - \tilde{v}_n, \tilde{v}_n - \tilde{v}^* \rangle + \|\tilde{v}_n - \hat{v}_n\|^2 + \\ & \|\tilde{v}_n - \tilde{v}^*\|^2] \leq \|\tilde{v}^* - \tilde{v}_n\|^2 \end{aligned}$$

Computing the square form of the third and fourth terms we have that

$$\|\tilde{v}^* - \tilde{v}_{n+1}\|^2 \leq \left(1 - 2\theta + \frac{(2\theta)^2}{1 + 2\theta - 2\theta^2\omega^2}\right)\|\tilde{v}^* - \tilde{v}_n\|^2$$

Let $\xi = 1 - 2\theta + \frac{(2\theta)^2}{1 + 2\theta - 2\theta^2\omega^2} < 1$. then iterating over the previous inequality, we have

$$\|\tilde{v}^* - \tilde{v}_{n+1}\|^2 \leq q\|\tilde{v}^* - \tilde{v}_n\|^2 \leq \dots \leq e^{n+1 \ln \xi} \|\tilde{v}^* - \tilde{v}_0\|^2 \quad (A8)$$

By Equation (A8) we have that $\|\tilde{v}^* - \tilde{v}_{n+1}\| \xrightarrow{n \rightarrow \infty} 0$. Taking into account that $\tilde{v}$ is a sequence that is bounded, we have that there is a point $\tilde{v}'$ such that any subsequence $\tilde{v}_{n_i}$ fulfills that $\tilde{v}_{n_i} \xrightarrow{n_i \rightarrow \infty} \tilde{v}'$ (Weierstrass theorem). Now, we have that $\|\tilde{v}_{n_i} - \tilde{v}_{n_i+1}\|^2 \rightarrow 0$. Letting $n = n_i$ in Equation (19) and taking the limit when $n_i \rightarrow \infty$ we obtain

$$\tilde{v}' = \arg \min_{\tilde{w} \in \tilde{X} \times \tilde{Y}} \left\{ \frac{1}{2} \|\tilde{w} - \tilde{v}'\|^2 + \theta L(\tilde{w}, \tilde{v}') \right\}$$

As a result, we have that $\tilde{v}' = \tilde{v}^*$. Provided that $\|\tilde{v}_n - \tilde{v}^*\|^2$ is monotonically decreasing then there exists a unique limit point (equilibrium point). As a result, the sequence $\tilde{v}_n$ satisfies that $\tilde{v}_n \xrightarrow{n \rightarrow \infty} \tilde{v}^*$ with a rate given by $e^{n+1 \ln \xi}$.

References

1. Hurwicz, L. Optimality and informational efficiency in resource allocation processes. In *Mathematical Methods in the Social Sciences: Proceedings of the First Stanford Symposium*; Arrow, K.J., Karlin, S., Suppes, P., Eds.; Stanford University Press: Palo Alto, CA, USA, 1960; pp. 27–46.
2. Nobel. *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2007: Scientific Background*; Technical Report; The Nobel Foundation: Stockholm, Sweden, 2007.
3. Myerson, R.B. *Allocation, Information and Markets*; Chapter Mechanism Design; The New Palgrave; Palgrave Macmillan: London, UK, 1989; pp. 191–206.
4. Vickrey, W. Counterspeculation, auctions, and competitive sealed tenders. *J. Financ.* **1961**, *16*, 8–37. [CrossRef]
5. Bergemann, D.; Välimäki, J. Dynamic mechanism design: An introduction. *J. Econ. Perspect* **2019**, *52*, 235–274. [CrossRef]
6. Clarke, E. Multi-part pricing of public goods. *Public Choice* **1971**, *11*, 17–23. [CrossRef]

7. Groves, T. Incentives in teams. *Econometrica* **1973**, *41*, 617–631. [\[CrossRef\]](#)
8. Harsanyi, J.C. Games with incomplete information played by bayesian players. part i: The basic model. *Manag. Sci.* **1967**, *14*, 159–182. [\[CrossRef\]](#)
9. Arrow, K. *Economics and Human Welfare*; Chapter The Property Rights Doctrine and Demand Revelation under Incomplete Information; Academic Press: New York, NY, USA, 1979; pp. 23–39.
10. d’Aspremont, C.; Gerard-Varet, L. Incentives and incomplete information. *J. Public Econ.* **1979**, *11*, 25–45.
11. Saari, D.G. On the types of information and mechanism design. *J. Comput. Appl. Math.* **1988**, *22*, 231–242. [\[CrossRef\]](#)
12. Rogerson, W. Contractual solutions to the hold-up problem. *Rev. Econ. Stud.* **1992**, *59*, 777–793. [\[CrossRef\]](#)
13. Mailath, G.; Postlewaite, A. Asymmetric information bargaining problems with many agents. *Rev. Econ. Stud.* **1990**, *57*, 351–360. [\[CrossRef\]](#)
14. Miyakawa, T. *Non-Cooperative Foundation of Nash Bargaining Solution under Incomplete Information*; Osaka University of Economics Working Paper Serier No. 2012-2.; Osaka University: Suita, Japan, 2012.
15. Athey, S.; Bagwell, K. Collusion with persistent cost shocks. *Econometrica* **2008**, *76*, 493–540. [\[CrossRef\]](#)
16. Hörner, J.; Takahashi, S.; Vieille, N. Truthful equilibria in dynamic bayesian games. *Econometrica* **2015**, *83*, 1795–1848. [\[CrossRef\]](#)
17. Clempner, J.B.; Poznyak, A.S. A nucleus for bayesian partially observable markov games: Joint observer and mechanism design. *Eng. Appl. Artif. Intell.* **2020**, *95*, 103876. [\[CrossRef\]](#)
18. Rahman, B. The power of communication. *Am. Econ. Rev.* **2014**, *104*, 3737–3751. [\[CrossRef\]](#)
19. Bernheim, B.; Madsen, E. Price cutting and business stealing in imperfect cartels. *Am. Econ. Rev.* **2017**, *107*, 387–424. [\[CrossRef\]](#)
20. Escobar, J.F.; Llanes, G. Cooperation dynamics in repeated games of adverse selection. *J. Econ. Theory* **2018**, *176*, 408–443. [\[CrossRef\]](#)
21. Poznyak, A.S.; Najim, K.; Gómez-Ramírez, E. *Self-Learning Control of Finite Markov Chains*; Marcel Dekker, Inc.: New York, NY, USA, 2000.
22. Clempner, J.B.; Poznyak, A.S. Simple computing of the customer lifetime value: A fixed local-optimal policy approach. *J. Syst. Sci. Syst. Eng.* **2014**, *23*, 439–459. [\[CrossRef\]](#)
23. Clempner, J.B.; Poznyak, A.S. Observer and control design in partially observable finite markov chains. *Automatica* **2019**, *110*, 108587. [\[CrossRef\]](#)
24. Clempner, J.B. On lyapunov game theory equilibrium: Static and dynamic approaches. *Int. Game Theory Rev.* **2018**, *20*, 1750033. [\[CrossRef\]](#)
25. Clempner, J.B.; Poznyak, A.S. Finding the strong nash equilibrium: Computation, existence and characterization for markov games. *J. Optim. Theory Appl.* **2020**, *186*, 1029–1052. [\[CrossRef\]](#)
26. Sragovich, V.G. *Mathematical Theory of Adaptive Control*; World Scientific Publishing Company: Singapore, 2006.
27. Asiain, E.; Clempner, J.B.; Poznyak, A.S. A reinforcement learning approach for solving the mean variance customer portfolio for partially observable models. *Int. J. Artif. Intell. Tools* **2018**, *27*, 1850034–1–1850034–30. [\[CrossRef\]](#)
28. Trejo, K.K.; Clempner, J.B.; Poznyak, A.S. Computing the lp-strong nash equilibrium for markov chains games. *Appl. Math. Model.* **2017**, *41*, 399–418. [\[CrossRef\]](#)
29. Hotelling, H. Stability in competition. *Econ. J.* **1929**, *39*, 41–57. [\[CrossRef\]](#)
30. Downs, A. *An Economic Theory of Democracy*; Harper & Brothers: New York, NY, USA, 1957.
31. Solis, C.; Clempner, J.B.; Poznyak, A.S. Robust extremum seeking for a second order uncertain plant using a sliding mode controller. *Int. J. Appl. Math. Comput. Sci.* **2019**, *29*, 703–712. [\[CrossRef\]](#)
32. Solis, C.; Clempner, J.B.; Poznyak, A.S. Robust integral sliding mode controller for optimisation of measurable cost functions with constraints. *Int. J. Control* **2019**, 1–13; To be published. [\[CrossRef\]](#)
33. Solis, C.; Clempner, J.B.; Poznyak, A.S. Continuous-time gradient-like descent algorithm for constrained convex unknown functions: Penalty method application. *J. Comput. Appl. Math.* **2019**, *355*, 268–282. [\[CrossRef\]](#)