Operations Research Letters 42 (2014) 7–11

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

On discounted stochastic games with incomplete information on payoffs and a security application

Erim Kardeş *

Nisantepe Mah Orman Sk No 13, Ozyegin University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 5 October 2013
Accepted 14 October 2013
Available online 26 October 2013

Keywords:
Stochastic games
Robust optimization
Robust game theory
Security application

ABSTRACT

This paper presents a robust optimization model for n-person finite state/action stochastic games with incomplete information on payoffs. For polytopic uncertainty sets, we propose an explicit mathematical programming formulation for an equilibrium calculation. It turns out that a global optimal of this mathematical program yields an equilibrium point and epsilon-equilibria can be calculated based on this result. We briefly describe an incomplete information version of a security application that can benefit from robust game theory.

1. Introduction

Noncooperative stochastic games model situations where antagonistic decision-makers, called players, choose actions over time to individually optimize their respective objectives. Objectives, however, do not only depend on players' actions in the current state of the game, but also on states possibly to be visited in the future. Given actions in a state, each player receives a payoff, known by all players. The process then moves onto another state with some probability and continues thereon.

In this paper, we consider n-person discounted finite state/action stochastic games in discrete time, where players do not know the payoffs incurred exactly. That is, we study incomplete information games where players only do know that payoffs corresponding to possible actions in a state belong to a set. We do not assume a probability distribution on this set; hence, the proposed model is a distribution-free incomplete information game. It is assumed that players use a robust optimization approach to address payoff uncertainty, and each player knows that others adopt this approach as well. Our main result is that when payoffs belong to a polytopic uncertainty set for each state, the problem of calculating an equilibrium point can be formulated as a nonlinear programming problem. We prove that a global optimal point of this mathematical program yields an optimal objective function value of 0 and corresponds to an equilibrium point. Further, we show that a feasible point of this program that yields an objective value close to 0 forms an approximate equilibrium point. This conclusion is important since the nonlinear program we provide is nonconvex, and this result implies that descent algorithms applied to our formulation that yield an objective function value close to 0 would indeed provide an ϵ-equilibrium. A similar result holds for the complete information case and we observe that it is preserved when the uncertainty sets are polytopic in the incomplete information case. Robust versions of stochastic games pose a difficulty, since the formulations for complete information games are nonconvex and one has to deal with the computation of the value of a game to players, along with equilibrium strategies. We make use of duality theorems to overcome this hurdle.

Our motivation for this work is that in various applications it can be difficult to quantify payoff functions exactly. This is especially the case if there is a lack of data on payoffs, or if it is difficult to elicit probability distributions on payoffs. Application areas include dynamic security games, service and admission rate control in service systems, strategic customer behavior in queueing systems, inventory control for substitutable products, communication systems to name a few, which are all characterized by dynamic interactions among several competing decision-makers over time.

The next subsections provide the related literature and robust optimization basics. Section 2 presents the notation and problem setup. Section 3 presents results and proofs. We briefly describe an incomplete information security application in Section 4.

1.1. Related literature

[3] introduces robust optimization for convex problems. Since then the robust approach is adopted in many areas of operations research. The merge of robust optimization and game theory is relatively recent. Robust game theory is introduced by [1] to

* Tel.: +90 2165649363.
E-mail address: erim.kardes@ozyegin.edu.tr.

0167-6377/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.orl.2013.10.005

address payoff uncertainty in one-shot games. In that paper, authors present a distribution-free model of incomplete information one-shot games, in which players use a robust optimization approach to contend with payoff uncertainty. They prove the existence of an equilibrium point when the payoff uncertainty set is bounded. Their formulation considers that payoffs belong to a polytope, yielding a method to compute an equilibrium point. Our work extends the approach by [1] to stochastic games with incomplete information on payoffs. [8] proves the existence of an equilibrium point in discounted robust stochastic games and provides a characterization when there is uncertainty only in the transition data. The current paper extends the characterization result by [8] to address payoff uncertainty in stochastic games, and further addresses ϵ-equilibrium.

The first paper on stochastic games was written by Shapley in 1953 [13]. Many extensions are proposed since then including the incomplete information case. [7] provides a comprehensive introduction to finite state/action stochastic games in discrete time. Nonlinear programming formulations for complete information stochastic games are presented by [5,7]. [14,15] consider a class of stochastic games with incomplete information on one side and with a single nonabsorbing state. In that work, the payoff matrix of a game is chosen according to a probability distribution. It is then shown that these games have a min–max and a max–min value. [16] studies stochastic games with incomplete information on states. [12] considers two-person zero-sum games, where the incomplete information is described by a finite set of stochastic games. A game is to be played out of this finite set, over which a probability distribution is specified. That paper focuses on stochastic games in which one player controls the transitions. We note that the approach proposed by [2] requires a probability distribution over a set of games. [6] presents the original complete information traveling inspector model. [9] studies the computation of Stackelberg strategies in stochastic games. [11] presents a stochastic game model of security and intrusion detection in computer networks. [17,10] present further models of network security. In particular, [10] uses the nonlinear programming formulation for the complete information stochastic games. The model presented by [10] can be extended to account for incomplete information and can then be solved using our results.

1.2. Robust optimization

This section briefly reviews robust optimization, as introduced by [3]. Consider the following problem P_γ : $\min_{\mathbf{x}\in\mathfrak{R}^n} f(\mathbf{x},\boldsymbol{\gamma})$ s.t. $F(\mathbf{x},\boldsymbol{\gamma}) \in K \subset \mathfrak{R}^m$, where $\boldsymbol{\gamma} \in \mathfrak{R}^M$ is the data vector, $\mathbf{x} \in \mathfrak{R}^n$ is the decision vector, and K is a convex cone. It is assumed that the data of P_γ is not known exactly but it is known that it belongs to an uncertainty set $U \in \mathfrak{R}^M$. The robust optimization approach considers the problem $P = \{P_\gamma\}_{\gamma\in U}$, where the constraints $F(\mathbf{x},\boldsymbol{\gamma}) \in K$ must be satisfied no matter what the actual realization of $\boldsymbol{\gamma} \in U$ is. An optimal solution to the uncertain problem P is defined as a solution that gives the best possible guaranteed value under all possible realizations of constraints. Formally, it is an optimal solution of the following program P_R: $\min_{\mathbf{x}\in\mathfrak{R}^n}\{\sup_{\gamma\in U} f(\mathbf{x},\boldsymbol{\gamma})\ s.t.\ F(\mathbf{x},\boldsymbol{\gamma}) \in K,\ \forall \boldsymbol{\gamma} \in U\}$. Prior work [4,3] has shown that for many function types and uncertainty sets, the robust counterpart problem P_R can be solved as a single optimization problem of size comparable to a deterministic version of the problem.

2. Notation summary and problem setup

The set of states is denoted by $\mathcal{S} = \{1,\ldots,S\}$, and the set of players by $\mathcal{I} = \{1,\ldots,I\}$. Given the game is in state s, player i can choose an action a^i from a finite set of actions. We assume without loss of generality that each player has J actions in every state. The extension to different numbers of actions is straightforward and requires only more complex notation. Suppose each player makes a choice in a state, i.e., we have an action tuple $a = (a^1,\ldots,a^I) \in A$, where A is the set of all possible action tuples in any state. Then the game moves into state k with probability $P_{sak} \geq 0$, $\sum_{k=1}^S P_{sak} = 1$. The strategy of player i in state s is $\mathbf{x}_s^i = (x_{s1}^i,\ldots,x_{sJ}^i)$, which belongs to the J-dimensional probability simplex, $\Delta = \{\mathbf{x}_s^i \in \mathfrak{R}_+^J \mid \sum_{j=1}^J x_{sj}^i = 1\}$. For mathematical tractability and ease of implementation, we consider stationary strategies in this work, which prescribe a player the same probabilities for his choices each time the player visits a certain state. Stationary strategies of a player i are represented by $\mathbf{x}^i = (\mathbf{x}_1^i,\ldots,\mathbf{x}_S^i)$, and the set of mixed strategies of all players in all states by $\mathbf{x} = (\mathbf{x}^1,\ldots,\mathbf{x}^I)$. We denote, for all states, the mixed strategies of all players except player i by $\mathbf{x}^{-i} = (\mathbf{x}^1,\ldots,\mathbf{x}^{i-1},\mathbf{x}^{i+1},\ldots,\mathbf{x}^I)$. The following notation is used to distinguish a mixed strategy of player i from those of others: $(\mathbf{x}^{-i},\mathbf{u}^i) = (\mathbf{x}^1,\ldots,\mathbf{x}^{i-1},\mathbf{u}^i,\mathbf{x}^{i+1},\ldots,\mathbf{x}^I)$. Players choose their actions independently at a given state. Hence, the probability that an action tuple a is chosen by the players in state s is denoted by

$$\pi_s^a(\mathbf{x}^{-i},\mathbf{u}^i) = \prod_{\substack{m=1\\ m\neq i}}^{I} x_{sa^m}^m u_{sa^i}^i.$$

For each $s \in \mathcal{S}$, elements of the set C_s are vectors $\tilde{\mathbf{c}}_s = [\tilde{C}_{sa}^i]$, $i \in \mathcal{I}$, $a \in A$. That is, for each player i and action tuple a in state s, there is an immediate cost $\tilde{C}_{sa}^i$ that is not known exactly. We denote the β-discounted robust value of an arbitrary strategy $\mathbf{y}$ to player i in state s by $\omega_s^i(\beta,\mathbf{y})$. $\boldsymbol{\omega}^i(\beta,\mathbf{y})$ represents the $S\times 1$ robust value vector achieved by strategy $\mathbf{y}$.

Given $\mathbf{x}^{-i}$, the β-discounted robust values to player i, resulting from i's best response, satisfy the following Bellman-type equation:

$$\omega_s^i = \min_{\mathbf{u}_s^i\in\Delta}\max_{\tilde{\mathbf{c}}_s\in C_s}\sum_{a\in A}\pi_s^a(\mathbf{x}^{-i},\mathbf{u}^i)\left\{\tilde{C}_{sa}^i + \beta\sum_{k=1}^S P_{sak}\omega_k^i\right\}. \tag{1}$$

Here, ω_s^i, $\forall s, i$ is the variable representing the β-discounted robust value to player i starting the process in state s, and the inner maximization problem is with respect to the uncertain immediate costs.

For simplicity of notation, we define

$$\psi_s^i(\tilde{\mathbf{c}}_s,\mathbf{p}_s;\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i) = \sum_{a\in A}\pi_s^a(\mathbf{x}^{-i},\mathbf{u}^i)\left\{\tilde{C}_{sa}^i + \beta\sum_{k=1}^S P_{sak}\omega_k^i\right\},$$

where $\mathbf{p}_s$ is the vector of probabilities out of state s under all possible action tuples a.

Definition. A point $\mathbf{x}$ is a *robust Markov perfect equilibrium* point in a discounted robust stochastic game if and only if, $\exists\boldsymbol{\omega} = [\boldsymbol{\omega}^1,\ldots,\boldsymbol{\omega}^I]$ satisfying Eq. (1), such that, $\forall i \in \mathcal{I}$, $\forall s \in \mathcal{S}$,

$$\mathbf{x}_s^i \in \operatorname*{argmin}_{\mathbf{u}_s^i\in\Delta}\max_{\tilde{\mathbf{c}}_s\in C_s}\psi_s^i(\tilde{\mathbf{c}}_s,\mathbf{p}_s;\mathbf{x}_s^{-i},\mathbf{u}_s^i;\boldsymbol{\omega}^i). \tag{2}$$

3. Characterization of equilibria

We will show that when the uncertainty set of payoffs of a game is a polytope, the problem of finding an equilibrium point could be formulated as a nonlinear optimization problem. The characterization result we present here generalizes a previous result for normal form one-shot games by [1] to discounted stochastic games with incomplete information on payoffs.

The definition of a robust Markov perfect equilibrium and conditions (2) are equivalent to the requirement that $\forall i \in \mathcal{I}$, $s \in \mathcal{S}$, $\exists q_s^{*i} \in \Re$ such that $(\mathbf{x}_s^i, q_s^{*i})$ is an optimizer of the following robust mathematical program P_R with the objective value at optimality being equal to the robust value ω_s^i, i.e. $q_s^{*i} = \omega_s^i$:

$$(P_R): \min_{u_s^i, q_s^i} q_s^i$$
$$q_s^i \geq \max_{\tilde{\mathbf{c}}_s \in C_s} \psi_s^i(\tilde{\mathbf{c}}_s, \mathbf{p}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i)$$
$$\mathbf{u}_s^i \in \Delta.$$

Here, $(\mathbf{x}^{-i}, \boldsymbol{\omega}^i)$ is treated as data. The following denotes the transition matrix induced by a strategy $(\mathbf{x}^{-i}, \mathbf{u}^i)$:

$$\mathbf{P}(\mathbf{x}^{-i}, \mathbf{u}^i) = \left[\sum_{a \in A} \prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m u_{sa^i}^i P_{sak} \right]_{s=1,k=1}^{S,\,S}.$$

The sth row of the induced transition matrix is denoted by the column vector $\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{u}^i)$.

$\mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s) \in \Re^{(J^{I-1}) \times J}$ denotes the matrix with uncertain entries, whose rows are given by the vectors

$$\left[\prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m \tilde{C}_{s(a^{-i}, a^i)}^i \right]_{a^i \in \{1,\dots,J\}}. \tag{3}$$

Note that we denote an action tuple a by (a^{-i}, a^i). a^{-i} denotes an action tuple comprised of all players' actions except player i, i.e., $a^{-i} = (a^1, \dots, a^{i-1}, a^{i+1}, \dots, a^I) \in A^{-i}$, where A^{-i} is the set of all such tuples. The number of possible action tuples for all players except player i is J^{I-1}, and hence the matrix $\mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)$ has J^{I-1} rows. Using this notation, the constraints in problem P_R can be rewritten as follows:

$$\begin{aligned} q_s^i &\geq \max_{\tilde{\mathbf{c}}_s \in C_s} \psi_s^i(\tilde{\mathbf{c}}_s, \mathbf{p}_s; \mathbf{x}_s^{-i}, \mathbf{u}_s^i; \boldsymbol{\omega}^i) \\ &= \max_{\tilde{\mathbf{c}}_s} \beta[\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]^T \boldsymbol{\omega}^i + \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i, \end{aligned} \tag{4}$$

where $\mathbf{1}$ is a column vector of ones of appropriate dimension.

Let

$$C_s = \{\tilde{\mathbf{c}}_s | \mathbf{A}_s \tilde{\mathbf{c}}_s \geq \mathbf{b}_s\} \neq \emptyset.$$

Note here that $\tilde{\mathbf{c}}_s \in \Re^{IJ^I}$. Given $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i)$, the maximization problem in (4) is a linear program and is as follows:

$$\left\{ \max_{\tilde{\mathbf{c}}_s \in C_s} \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i \right\}. \tag{5}$$

Lemma 1. *Condition (4), that is, $q_s^i \geq \max_{\tilde{\mathbf{c}}_s} \beta[\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]^T \boldsymbol{\omega}^i + \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i$ is equivalent to the condition that, $\forall s$, $\exists \mathbf{m}_s \in \Re^l$ such that*

$$\begin{aligned} &q_s^i - \beta[\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]^T \boldsymbol{\omega}^i \geq \mathbf{b}_s^T \mathbf{m}_s \\ &\mathbf{A}_s^T \mathbf{m}_s = \mathbf{Y}_s^i(\mathbf{x}_s^{-i})\mathbf{u}_s^i \\ &\mathbf{m}_s \leq 0, \end{aligned} \tag{6}$$

where $\mathbf{Y}_s^i(\mathbf{x}_s^{-i}) \in \Re^{IJ^I \times J}$ is the matrix such that

$$[\tilde{\mathbf{c}}_s]^T \mathbf{Y}_s^i(\mathbf{x}_s^{-i})\mathbf{u}_s^i = \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i.$$

Proof. The dual of the maximization problem in (5) is as follows:

$$\left\{ \min_{\mathbf{m}_s} \mathbf{b}_s^T \mathbf{m}_s \ s.t.\ \mathbf{A}_s^T \mathbf{m}_s = \mathbf{Y}_s^i(\mathbf{x}_s^{-i})\mathbf{u}_s^i,\ \mathbf{m}_s \leq 0 \right\}, \tag{7}$$

where $\mathbf{m}_s \in \Re^l$ is the vector of dual variables.

By the definition of the uncertainty set C_s for each state, problem (5) is feasible and bounded. Therefore, by strong duality, problem (7) is bounded, feasible, and its optimal objective value is equal to that of problem (5). Hence, given $(\mathbf{x}_s^{-i}, \mathbf{u}_s^i, \boldsymbol{\omega}^i)$, if condition (4) is satisfied, then condition (4) is equivalent to the condition that $\exists \mathbf{m}_s \in \Re^l$ such that condition (6) is satisfied.

On the other hand, if condition (6) is satisfied, then problem (7) is feasible. By weak duality, any feasible solution $\mathbf{b}_s^T \mathbf{m}_s$ of problem (7) is greater than or equal to any solution $\mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i$ of problem (5). Hence,

$$q_s^i - \beta[\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{u}^i)]^T \boldsymbol{\omega}^i \geq \mathbf{b}_s^T \mathbf{m}_s \geq \max_{\tilde{\mathbf{c}}_s \in C_s} \mathbf{1}^T \mathbf{E}_s^i(\mathbf{x}_s^{-i}, \tilde{\mathbf{c}}_s)\mathbf{u}_s^i,$$

showing that condition (4) is recovered. □

To present our main result, let

$$\mathbf{P}_s^i(\boldsymbol{\omega}^i) = \left[\sum_{k \in \mathcal{S}} P_{s(a^i, a^{-i})k} \omega_k^i \right]_{a^i \in \{1,\dots,J\},\ a^{-i} \in A^{-i}}.$$

$\mathbf{P}_s^i(\boldsymbol{\omega}^i)$ is a $J \times J^{(I-1)}$ matrix. $\boldsymbol{\pi}(\mathbf{x}_s^{-i}) \in \Re^{J^{(I-1)}}$ is a column vector each entry of which denotes the joint probability of an action tuple chosen by all players except player i. That is, each entry is of the form $\prod_{\substack{m=1 \\ m \neq i}}^{I} x_{sa^m}^m$, where $(a^1, \dots, a^{i-1}, a^{i+1}, \dots, a^m)$ forms an action tuple of all but player i. The order of these action tuples matches the order of columns of $\mathbf{P}_s^i(\boldsymbol{\omega}^i)$.

Let $\mathbf{t}_s \in \Re^{IJ^I}$, $\mathbf{t} = [\mathbf{t}_s]_{s \in \mathcal{S}}$, $\mathbf{m} = [\mathbf{m}_s]_{s \in \mathcal{S}}$ be variable vectors.

Theorem 1. *Consider a point $\mathbf{z}^{T*} = (\boldsymbol{\omega}^*, \mathbf{x}^*, \mathbf{t}^*, \mathbf{m}^*)$. The strategy part $\mathbf{x}^*$ of $\mathbf{z}^*$ forms an equilibrium point of a discounted robust stochastic game with payoff uncertainty if and only if $\mathbf{z}^*$ is the global minimum of the following nonlinear program, yielding an objective value of 0:*

$$\begin{aligned} &\min \sum_{i \in \mathcal{I}} \sum_{s \in \mathcal{S}} \left([\mathbf{x}_s^i]^T [\mathbf{Y}_s^i(\mathbf{x}_s^{-i})]^T \mathbf{t}_s + \beta [\mathbf{x}_s^i]^T \mathbf{P}_s^i(\boldsymbol{\omega}^i)\boldsymbol{\pi}(\mathbf{x}_s^{-i}) - \omega_s^i \right) \\ &\omega_s^i - \beta[\mathbf{p}_s(\mathbf{x}^{-i}, \mathbf{x}^i)]^T \boldsymbol{\omega}^i \geq \mathbf{b}_s^T \mathbf{m}_s \\ &\mathbf{A}_s^T \mathbf{m}_s - \mathbf{Y}_s^i(\mathbf{x}_s^{-i})\mathbf{x}_s^i = 0 \\ &\mathbf{A}_s \mathbf{t}_s \geq \mathbf{b}_s \\ &\mathbf{1}^T \mathbf{x}_s^i = 1 \\ &[\mathbf{Y}_s^i(\mathbf{x}_s^{-i})]^T \mathbf{t}_s + \beta \mathbf{P}_s^i(\boldsymbol{\omega}^i)\boldsymbol{\pi}(\mathbf{x}_s^{-i}) \geq \omega_s^i \mathbf{1} \\ &\mathbf{m}_s \leq 0, \qquad \mathbf{x}_s^i \geq 0. \end{aligned} \tag{8}$$

Proof. Recall problem P_R. By equilibrium conditions (2) and Lemma 1, if $\mathbf{x}^*$ is an equilibrium point, given $(\mathbf{x}_s^{*-i}, \boldsymbol{\omega}^{*i})$, $\forall i \in \mathcal{I}$, $s \in \mathcal{S}$, $(\mathbf{x}_s^{*i}, \omega_s^{*i}, \mathbf{m}_s^*)$ is an optimizer of

$$\begin{aligned} &\min_{\mathbf{u}_s^i, q_s^i, \mathbf{m}_s} \quad q_s^i \\ &q_s^i - \beta[\mathbf{p}_s(\mathbf{x}^{*-i}, \mathbf{u}^i)]^T \boldsymbol{\omega}^{*i} \geq \mathbf{b}_s^T \mathbf{m}_s \\ &\mathbf{A}_s^T \mathbf{m}_s = \mathbf{Y}_s^i(\mathbf{x}_s^{*-i})\mathbf{u}_s^i \\ &\mathbf{m}_s \leq 0 \\ &\mathbf{1}^T \mathbf{u}_s^i = 1 \\ &\mathbf{u}_s^i \geq 0. \end{aligned} \tag{9}$$

The dual of the above is

$$\max_{\nu_s^i, \mathbf{t}_s} \quad \nu_s^i \tag{10}$$

$$\nu_s^i \mathbf{1} \leq [\mathbf{Y}_s^i(\mathbf{x}_s^{*-i})]^T \mathbf{t}_s + \beta \mathbf{P}_s^i(\boldsymbol{\omega}^{*i})\boldsymbol{\pi}(\mathbf{x}_s^{*-i})$$
$$\mathbf{A}_s \mathbf{t}_s \geq \mathbf{b}_s.$$

The constraints in Theorem 1 therefore hold by strong duality. Note that $\boldsymbol{\omega}^*$ is the robust value vector achieved by the equilibrium strategy $\mathbf{x}^*$ and the worst-case payoff perspectives of the players in state s captured in $\mathbf{t}_s^*$, which denotes the vector attained at optimality by the dual vector $\mathbf{t}_s$ in problem (10). By Theorem 2 in [8], there exists a corresponding unique robust value vector for any strategy $\mathbf{x}$. Therefore, by construction, the equality $[\mathbf{x}_s^{*i}]^T[\mathbf{Y}_s^i(\mathbf{x}_s^{*-i})]^T\mathbf{t}_s^* + \beta[\mathbf{x}_s^{*i}]^T\mathbf{P}_s^i(\boldsymbol{\omega}^{*i})\boldsymbol{\pi}(\mathbf{x}_s^{*-i}) = \omega_s^{*i}$ holds, yielding an objective value of 0 at point $\mathbf{z}^*$. To see why $\mathbf{z}^*$ is a global optimal point, suppose $\bar{\mathbf{z}}^T = (\bar{\omega}, \bar{\mathbf{x}}, \bar{\mathbf{t}}, \bar{\mathbf{m}})$ is feasible for the constraints in the theorem.

Multiplying the second to last constraint in the theorem from the left by $[\bar{\mathbf{x}}_s^i]^T$ yields

$$[\bar{\mathbf{x}}_s^i]^T[\mathbf{Y}_s^i(\bar{\mathbf{x}}_s^{-i})]^T\bar{\mathbf{t}}_s + \beta[\bar{\mathbf{x}}_s^i]^T\mathbf{P}_s^i(\bar{\boldsymbol{\omega}}^i)\boldsymbol{\pi}(\bar{\mathbf{x}}_s^{-i}) \geq \bar{\omega}_s^i.$$

Therefore, every term $[\mathbf{x}_s^i]^T[\mathbf{Y}_s^i(\mathbf{x}_s^{-i})]^T\mathbf{t}_s + \beta[\mathbf{x}_s^i]^T\mathbf{P}_s^i(\boldsymbol{\omega}^i)\boldsymbol{\pi}(\mathbf{x}_s^{-i}) - \omega_s^i$ in the objective function is nonnegative. Thus, the objective function value at a global minimum is 0, which is attained at an equilibrium.

For the other direction, suppose that $\mathbf{z}^*$ is a global minimum of the nonlinear program stated in the theorem, yielding the objective value of 0. Then, $\boldsymbol{\omega}^*$ is the robust value vector corresponding to $\mathbf{x}^*$. Let $\mathbf{e}_j$ denote the j-th unit column vector of dimension J. Let

$$\nu_s^{*i} = \min_{j\in\{1,\ldots,J\}} \mathbf{e}_j^T[\mathbf{Y}_s^i(\mathbf{x}_s^{*-i})]^T\mathbf{t}_s^* + \beta\mathbf{e}_j^T\mathbf{P}_s^i(\boldsymbol{\omega}^{*i})\boldsymbol{\pi}(\mathbf{x}_s^{*-i}),$$
$$q_s^{*i} = \beta[\mathbf{p}_s(\mathbf{x}^{*-i}, \mathbf{x}^{*i})]^T\boldsymbol{\omega}^{*i} + \mathbf{b}_s^T\mathbf{m}_s^*.$$

Then, given $(\mathbf{x}_s^{*-i}, \boldsymbol{\omega}^{*i}), \forall i \in \mathcal{I}, s \in \mathcal{S}$, $(\mathbf{x}_s^{*i}, q_s^{*i}, \mathbf{m}_s^*)$ is feasible for problem (9) and $(\nu_s^{*i}, \mathbf{t}_s^*)$ is feasible for problem (10) with $\nu_s^{*i} \geq q_s^{*i}$ due to the first and second to last constraints in the theorem. By weak duality, $\nu_s^{*i} \leq q_s^{*i}$; hence, $\nu_s^{*i} = q_s^{*i}$. Therefore, $(\mathbf{x}_s^{*i}, q_s^{*i}, \mathbf{m}_s^*)$ is optimal for problem (9). Equivalently $(\mathbf{x}_s^{*i}, q_s^{*i})$ is optimal in problem P_R, which means that $\mathbf{x}_s^{*i}$ is a best response to $\mathbf{x}_s^{*-i}$, $\forall i, s$. Thus, the strategy part $\mathbf{x}^*$ of $\mathbf{z}^*$ forms an equilibrium point of the game with $\boldsymbol{\omega}^*$ being the corresponding robust value vector. □

Given $\epsilon > 0$, strategy $\mathbf{x}$ forms an ϵ-equilibrium if $\boldsymbol{\omega}^i(\beta, (\mathbf{x}^i, \mathbf{x}^{-i})) \leq \boldsymbol{\omega}^i(\beta, (\mathbf{y}^i, \mathbf{x}^{-i})) + \mathbf{1}\epsilon$, for any $\mathbf{y}^i$, $\forall i$.

Corollary 1. *Suppose $\bar{\mathbf{z}}^T = (\bar{\omega}, \bar{\mathbf{x}}, \bar{\mathbf{t}}, \bar{\mathbf{m}})$ is feasible for the nonlinear program stated in Theorem 1 yielding an objective function value of $\delta > 0$. Then, the strategy part $\bar{\mathbf{x}}$ of $\bar{\mathbf{z}}$ forms an ϵ-equilibrium with $\epsilon = \delta/(1-\beta)$.*

Proof. The first term in the objective function evaluated at $\bar{\mathbf{z}}$, i.e. $[\bar{\mathbf{x}}_s^i]^T[\mathbf{Y}_s^i(\bar{\mathbf{x}}_s^{-i})]^T\bar{\mathbf{t}}_s$, represents the immediate expected payoff to player i in state s, resulting from strategy $\bar{\mathbf{x}}$ and vector $\bar{\mathbf{t}}_s \in C_s$, which accommodates players' perspectives on the uncertainty. To ease the exposition for the current proof, let us represent the $S \times 1$ vector of immediate expected payoffs to player i by $\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}})$, which has its s-th entry given by $r_s^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) = [\bar{\mathbf{x}}_s^i]^T[\mathbf{Y}_s^i(\bar{\mathbf{x}}_s^{-i})]^T\bar{\mathbf{t}}_s$. Further, let $\beta\mathbf{P}(\bar{\mathbf{x}})\bar{\boldsymbol{\omega}}^i$ represent the vector of discounted expected amount incurred by visiting future states. Its s-th entry represents the discounted expected amount incurred by visiting states in the future, given the initial state is s, i.e. the second term in the objective evaluated at $\bar{\mathbf{z}}$, $\beta[\bar{\mathbf{x}}_s^i]^T\mathbf{P}_s^i(\bar{\boldsymbol{\omega}}^i)\boldsymbol{\pi}(\bar{\mathbf{x}}_s^{-i})$.

As seen in the proof of Theorem 1, every term in the objective function of problem (8) is nonnegative, and by assumption, the objective function value is δ. Therefore, we have

$$\bar{\boldsymbol{\omega}}^i \geq \mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) + \beta\mathbf{P}(\bar{\mathbf{x}})\bar{\boldsymbol{\omega}}^i - \mathbf{1}\delta. \tag{11}$$

Substituting this inequality into itself yields $\bar{\boldsymbol{\omega}}^i \geq \mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) + \beta\mathbf{P}(\bar{\mathbf{x}})\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) + \beta^2\mathbf{P}^2(\bar{\mathbf{x}})\bar{\boldsymbol{\omega}}^i - \beta\mathbf{P}(\bar{\mathbf{x}})\mathbf{1}\delta - \mathbf{1}\delta$, where $\mathbf{P}^2(\bar{\mathbf{x}})$ is the two-step transition matrix induced by $\bar{\mathbf{x}}$. Substituting inequality (11) into itself k times yields

$$\bar{\boldsymbol{\omega}}^i \geq \mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) + \cdots + \beta^k\mathbf{P}^k(\bar{\mathbf{x}})\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) + \beta^{k+1}\mathbf{P}^{k+1}(\bar{\mathbf{x}})\bar{\boldsymbol{\omega}}^i - \left(\sum_{j=0}^{k}\beta^k\right)\mathbf{1}\delta.$$

Letting $k \to \infty$ yields

$$\bar{\boldsymbol{\omega}}^i \geq [\mathbf{I} - \beta\mathbf{P}^k(\bar{\mathbf{x}})]^{-1}\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) - \frac{1}{1-\beta}\mathbf{1}\delta.$$

The first term on the right-hand-side of the above inequality yields the robust value vector corresponding to strategy $\bar{\mathbf{x}}$. That is, $[\mathbf{I} - \beta\mathbf{P}^k(\bar{\mathbf{x}})]^{-1}\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}}) = \boldsymbol{\omega}^i(\beta, \bar{\mathbf{x}})$. This is so since $\mathbf{r}^i(\bar{\mathbf{x}}, \bar{\mathbf{t}})$ is the vector of immediate expected payoffs to player i given by the payoff perspectives of the players captured by the variable $\bar{\mathbf{t}}$. In the complete information case, a similar equality holds, where players' perspectives do not take place, since payoffs are known exactly (see page 13 in [7] for the 1-player case). Thus,

$$\bar{\boldsymbol{\omega}}^i \geq \boldsymbol{\omega}^i(\beta, \bar{\mathbf{x}}) - \frac{1}{1-\beta}\mathbf{1}\delta. \tag{12}$$

Now, multiplying the second to last constraint in the theorem from the left by an arbitrary strategy $[\mathbf{x}_s^i]^T$ yields

$$[\mathbf{x}_s^i]^T[\mathbf{Y}_s^i(\bar{\mathbf{x}}_s^{-i})]^T\bar{\mathbf{t}}_s + \beta[\mathbf{x}_s^i]^T\mathbf{P}_s^i(\bar{\boldsymbol{\omega}}^i)\boldsymbol{\pi}(\bar{\mathbf{x}}_s^{-i}) \geq \bar{\omega}_s^i, \quad \forall s$$

which we rewrite as

$$\mathbf{r}^i((\mathbf{x}^i, \bar{\mathbf{x}}^{-i}), \bar{\mathbf{t}}) + \beta\mathbf{P}(\mathbf{x}^i, \bar{\mathbf{x}}^{-i})\bar{\boldsymbol{\omega}}^i \geq \bar{\boldsymbol{\omega}}^i.$$

Similar to the above derivation, substituting this inequality into itself infinitely many times yields for any $\mathbf{x}^i$,

$$\boldsymbol{\omega}^i(\beta, (\mathbf{x}^i, \bar{\mathbf{x}}^{-i})) \geq \bar{\boldsymbol{\omega}}^i. \tag{13}$$

Inequalities (12) together with (13) show that $\bar{\mathbf{x}}$ is an ϵ-equilibrium strategy with $\epsilon = \delta/(1-\beta)$. □

4. A security application

We briefly describe an incomplete information version of the traveling inspector model where the results of this paper can be used. The original model is described in detail by [7,6]. Suppose an inspector is to inspect a given number of nodes for compliance with a regulation by moving from one node to the other at each time period. The inspector can inspect one node at a time. An antagonistic decision-maker can choose to violate the regulation at any node at certain levels. The inspector's objective is to minimize the losses generated by undetected violations and costs of inspection. For example, a node can be a computer on a network, a factory site, or an airport security checkpoint. Inspector's alternatives are the nodes to inspect in the next time period, whereas his/her opponent's alternatives are the levels of violation at each node. The node that is inspected last is the state of the game. For example, if the inspector chooses to inspect node j next (which means choosing alternative j), then the process will be in state j with probability 1 in the next time period. Therefore, the transition probabilities are either 1 or 0 in this model and the transitions are controlled only by the inspector. It is difficult, however, to quantify the payoffs to each player. Given the last node visited, the violation level chosen by the violator for each node, and the node to be inspected next chosen by the inspector, each player receives a payoff, which can be represented by the vector $\tilde{\mathbf{c}}_s$ introduced in Section 2. Depending on the application, a polytopic uncertainty set can be constructed representing the payoff information. Note that what we obtain then is a two-person stochastic game with incomplete information on payoffs, where transitions are known exactly and are controlled by one player only. If a player is assumed not to use robust optimization, then constraints of the uncertainty set corresponding to that player can be modified to account for this assumption. Therefore, the results of this paper can be used for this model. Another model where our results can be used is the incomplete information extension of the model presented by [10].

References

[1] M. Aghassi, D. Bertsimas, Robust game theory, Mathematical Programming, Series B 107 (2006) 231–273.

[2] R.J. Aumann, M.B. Maschler, Repeated Games of Incomplete Information: The Zero-Sum Extensive Case, Tech. Rep., US Arms Control and Disarmament Agency, sT-143, Washington DC, 1968, p. 37116. Chapter III.

[3] A. Ben-Tal, A. Nemirovski, Robust convex optimization, Mathematics of Operations Research 23 (4) (1998).

[4] D. Bertsimas, M. Sim, The price of robustness, Operations Research 52 (2004) 35–53.

[5] J.A. Filar, T.A. Schultz, Nonlinear programming and stationary strategies in stochastic games, Mathematical Programming 35 (1986) 243–247.

[6] J.A. Filar, T.A. Schultz, The traveling inspector model, OR Spectrum 8 (1986) 33–36.

[7] J. Filar, K. Vrieze, Competitive Markov Decision Processes, Springer-Verlag, 1997.

[8] E. Kardeş, F. Ordóñez, R.W. Hall, Discounted robust stochastic games and an application to queueing control, Operations Research 59 (2011) 365–382.

[9] J. Letchford, L. MacDermed, V. Conitzer, R. Parr, C. Isbell, Computing optimal strategies to commit to in stochastic games, AAAI (2012).

[10] K. Lye, J.M. Wing, Game strategies in network security, International Journal of Information Security 4 (2005) 71–86.

[11] K. C. Nguyen, T. Alpcan, T. Basar, Stochastic games for security in networks with interdependent nodes, in: Proceedings of the International Conference on Game Theory for Networks (GameNets 2009), 2009.

[12] D. Rosenberg, E. Solan, N. Vieille, Stochastic games with a single controller and incomplete information, SIAM Journal on Control and Optimization 43 (1) (2013).

[13] L.S. Shapley, Stochastic games, in: Proceedings of the National Academy of Sciences of the United States of America, Vol. 39, 1953, pp. 1095–1100.

[14] S. Sorin, Big match with lack of information on one side I, International Journal of Game Theory 13 (1984) 201–255.

[15] S. Sorin, Big match with lack of information on one side II, International Journal of Game Theory 14 (1985) 173–204.

[16] S. Sorin, Stochastic games with incomplete information, NATO Science Series 570 (2003) 375–395.

[17] Q. Zhu, H. Tembine, T. Başar, Network security configuration: a nonzero-sum stochastic game approach, in: Proceedings of the 2010 American Control Conference (ACC 2010), 2010, Baltimore, Maryland, June 30–July 2, 2010, pp. 1059–1064.