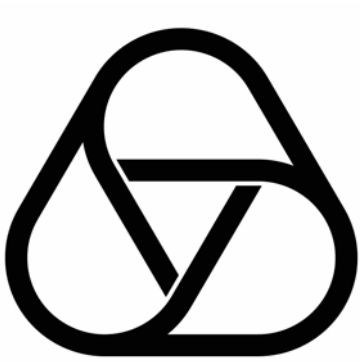
Oberwolfach Preprints



OWP 2011 - 35

ENDRE BOROS; KHALED ELBASSIONI;
VLADIMIR GURVICH; KAZUHISA MAKINO

On Canonical Forms for Two-person Zero-sum Limit Average Payoff Stochastic Games

Mathematisches Forschungsinstitut Oberwolfach gGmbH
Oberwolfach Preprints (OWP) ISSN 1864-7596

Oberwolfach Preprints (OWP)

Starting in 2007, the MFO publishes a preprint series which mainly contains research results related to a longer stay in Oberwolfach. In particular, this concerns the Research in Pairs-Programme (RiP) and the Oberwolfach-Leibniz-Fellows (OWLF), but this can also include an Oberwolfach Lecture, for example.

A preprint can have a size from 1 - 200 pages, and the MFO will publish it on its website as well as by hard copy. Every RiP group or Oberwolfach-Leibniz-Fellow may receive on request 30 free hard copies (DIN A4, black and white copy) by surface mail.

Of course, the full copy right is left to the authors. The MFO only needs the right to publish it on its website *www.mfo.de* as a documentation of the research work done at the MFO, which you are accepting by sending us your file.

In case of interest, please send a **pdf file** of your preprint by email to *rip@mfo.de* or *owlf@mfo.de*, respectively. The file should be sent to the MFO within 12 months after your stay as RiP or OWLF at the MFO.

There are no requirements for the format of the preprint, except that the introduction should contain a short appreciation and that the paper size (respectively format) should be DIN A4, "letter" or "article".

On the front page of the hard copies, which contains the logo of the MFO, title and authors, we shall add a running number (20XX - XX).

We cordially invite the researchers within the RiP or OWLF programme to make use of this offer and would like to thank you in advance for your cooperation.

Imprint:

Mathematisches Forschungsinstitut Oberwolfach gGmbH (MFO)
Schwarzwaldstrasse 9-11
77709 Oberwolfach-Walke
Germany

Tel +49 7834 979 50
Fax +49 7834 979 55
Email admin@mfo.de
URL www.mfo.de

The Oberwolfach Preprints (OWP, ISSN 1864-7596) are published by the MFO.

On Canonical Forms for Two-person Zero-sum Limit Average Payoff Stochastic Games *

Endre Boros† Khaled Elbassioni‡ Vladimir Gurvich§ Kazuhisa Makino¶

Abstract

We consider two-person zero-sum mean payoff undiscounted stochastic games. We give a sufficient condition for the existence of a saddle point in uniformly optimal stationary strategies. Namely, we obtain sufficient conditions that enable us to to bring the game, by applying *potential transformations* to a *canonical form* in which *locally* optimal strategies are *globally* optimal, and hence the value for every initial position and the optimal strategies of both players can be obtained by playing the local game at each state. We show that this condition is satisfied by the class of *additive transition games*, that is, the special case when the transitions at each state can be decomposed into two parts, each controlled completely by one of the two players.

An important special case of additive games is the so-called *BWR-games* which are played by two players on a directed graph with positions of three types: Black, White and Random. We given an independent proof for the existence of canonical form in such games, and use this to derive the existence of canonical form (and hence of a saddle point in uniformly optimal stationary strategies) in a wide class of games, which includes *stochastic games with perfect information*, *switching controller games* and *additive rewards, additive transition games*.

1 Definitions and Notations

1.1 Matrix Games

Given a real matrix $M \in \mathbb{R}^{I \times J}$, where I and J are sets labeling the rows and columns, respectively, we consider the corresponding zero-sum matrix game in which M is the payoff matrix. We view the sets I and J as sets of possible actions (and conveniently also, vectors of pure strategies) of the players. We consider the row player (player 1, or WHITE) as the maximizer and the column player (player 2, or BLACK) as the minimizer, and the matrix entry $(M)_{i,j}$ represents the payoff what player 1 receives from player 2 if $i \in I$ and $j \in J$ are the chosen actions. Of course, players can randomize, and use mixed strategies. Let us denote by

$$\Delta(I) \;=\; \{\alpha \in \mathbb{R}^I \mid \sum_{i \in I} \alpha_i = 1,\ \alpha_i \geq 0 \ \text{ for } \ i \in I\}$$

the set of mixed strategies corresponding to the action set I, and denote by

$$\text{Val}\,(M) \;=\; \max_{\alpha \in \Delta(I)} \min_{\beta \in \Delta(J)} \alpha M \beta \;=\; \min_{\beta \in \Delta(J)} \max_{\alpha \in \Delta(I)} \alpha M \beta$$

*This research was partially supported by DIMACS, Center for Discrete Mathematics and Theoretical Computer Science, Rutgers University, and by the Scientific Grant-in-Aid from Ministry of Education, Science, Sports and Culture of Japan. Part of this research was done at the Mathematisches Forschungsinstitut Oberwolfach during a stay within the Research in Pairs Program from March 7 to March 18, 2011.

†RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (boros@rutcor.rutgers.edu)

‡Max-Planck-Institute for Informatics; Stuhlsatzenhausweg 85, 66123, Saarbruecken, Germany (elbassio@mpi-sb.mpg.de)

§RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; (gurvich@rutcor.rutgers.edu)

¶Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 113-8656, Japan; (makino@mist.i.u-tokyo.ac.jp)

the value of the game (which always exists and can be computed via linear programming, see von Neumann [vN28]).

For subsets $Q \subseteq \Delta(I)$ and $R \subseteq \Delta(J)$ we define the "restricted" matrix $M[Q,R] \in \mathbb{R}^{Q\times R}$ by

$$(M[Q,R])_{\alpha,\beta} \;=\; \alpha M \beta \qquad \text{for all} \quad \alpha \in Q \quad \text{and} \quad \beta \in R,$$

and introduce

$$\underset{Q\times R}{\text{Val}}\,(M) \;=\; \text{Val}\,(M[Q,R])\,.$$

We call $M[Q,R]$ a *restriction* of M because we have $\Delta(Q)_I \subseteq \Delta(I)$ and $\Delta(R)_J \subseteq \Delta(J)$, and

$$\max_{\lambda\in\Delta(Q)} \min_{\delta\in\Delta(R)} \lambda M[Q,R]\delta \;=\; \max_{\alpha\in\Delta(Q)_I} \min_{\beta\in\Delta(R)_J} \alpha M \beta,$$

where, for sets $S \subseteq \Delta(Q)$ and $T \subseteq \Delta(R)$, we use the notation $S_I \;=\; \{\sum_{\alpha\in Q}\lambda_\alpha \alpha \mid \lambda \in S\}$ and $T_J \;=\; \{\sum_{\beta\in R}\delta_\beta\beta \mid \delta \in T\}$.

Let us note that this operation is additive, that is if $M, M' \in \mathbb{R}^{I\times J}$ and $\lambda, \mu \in \mathbb{R}$, then we have

$$(\lambda M + \mu M')\,[Q,R] \;=\; \lambda\,(M[Q,R]) + \mu\,(M'[Q,R])\,. \tag{1}$$

Let us denote by $\Omega = \Omega(M) \subseteq \Delta(I)$ and by $\Lambda = \Lambda(M) \subseteq \Delta(J)$ the extremal optimal mixed strategies of the row player and column player, respectively, in the matrix game with payoff matrix M. By Shapley and Snow [SS50] we know that both sets Ω and Λ are vertices of bounded polyhedra, and hence both sets are finite.

Remark 1 *The following equalities are well known, and easy to derive:*

$$\text{Val}\,(M) \;=\; \underset{\Omega\times\Lambda}{\text{Val}}\,(M) \;=\; \underset{\Omega\times J}{\text{Val}}\,(M) \;=\; \underset{I\times\Lambda}{\text{Val}}\,(M). \tag{2}$$

In fact, $M[\Omega,\Lambda]$ is a constant matrix having Val (M) *in all its entries, and we have the equalities*

$$\begin{aligned} \Omega(M) \;&=\; \Omega(M[\Omega,\Lambda])_I \;=\; \Omega(M[\Omega,J])_I \qquad \textit{and} \\ \Lambda(M) \;&=\; \Lambda(M[\Omega,\Lambda])_J \;=\; \Lambda(M[I,\Lambda])_J. \end{aligned}$$

Remark 2 *Let us finally note a few basic properties of matrix games (see []).*

- *If $M \le M' \in \mathbb{R}^{I\times J}$ are matrices, $\lambda, \delta \in \mathbb{R}$ reals, $\lambda \ge 0$, and $\mathbb{E} \in \mathbb{R}^{I\times J}$ is the matrix in which all entries are one, then we have*

$$\text{Val}\,(M) \le \text{Val}\,(M') \qquad \textit{and} \qquad \text{Val}\,(\lambda M + \delta\mathbb{E}) \;=\; \lambda\ \text{Val}\,(M) + \delta. \tag{3}$$

- Val (M) *is a Lipschitz-continuous function in its coefficients.*

1.2 Stochastic Games

Stochastic games were introduced in 1953 by Shapley [Sha53] for the discounted case, and extended to the undiscounted case by Gillette [Gil57]. Each such game $\Gamma = (p^{vu}_{k\ell},\ r^{vu}_{k\ell} \;\mid\; k \in K^v,\ \ell \in L^v,\ u, v \in V)$ is played by two players on a finite set of vertices (states) V; K^v and L^v for $v \in V$ are finite sets of actions (pure strategies) of the players, $r^{vu}_{k\ell}$ is the reward player 1 (White) receives from player 2 (Black) if k and ℓ are the chosen actions and the game moves from state v to state u, and $p^{vu}_{k\ell}$ is the transition probability from state v to state u if players chose actions $k \in K^v$ and $\ell \in L^v$ at state $v \in V$. We assume that the game is non-stopping[1], that is

$$\sum_{u\in V} p^{vu}_{k\ell} = 1 \tag{4}$$

[1] Shapley's original stochastic games were assumed to have positive *stopping probabilities*, i.e., at each state v, $\sum_{u\in V} p^{k\ell}_{vu} < 1$, and with probability $1 - \sum_{u\in V} p^{k\ell}_{vu}$, the game stops at state v if actions k and ℓ are selected by the players.

for all states $v \in V$ and for all choices of actions $k \in K^v$ and $\ell \in L^v$. To simplify later expressions, let us denote by $P^{vu} \in \mathbb{R}^{K^v \times L^v}$ the transition matrix, the elements of which are the $p^{vu}_{k\ell}$ probabilities.

We associate in Γ a *reward matrix* A^v to every state $v \in V$ defined by

$$(A^v)_{k\ell} \;=\; r^v_{k\ell} := \sum_{u \in V} p^{vu}_{k\ell} r^{vu}_{k\ell}. \tag{5}$$

When the game Γ is not clear from the context, we shall write $r^{vu}_{k\ell}(\Gamma)$, $p^{vu}_{k\ell}(\Gamma)$, $A^v(\Gamma)$, etc.

In the game Γ players first agree on an initial vertex $v_0 = w \in V$ to start. Then, in a general step $j = 0, 1, ...$, when the game arrives to state $v = v_j \in V$, they choose strategies $\alpha^{v_j} \in \Delta(K^v)$ and $\beta^{v_j} \in \Delta(L^v)$, player 1 receives the amount of

$$b_j = \alpha^{v_j} A^v \bar{\beta}^{v_j}$$

from player 2, and the game moves to the next state $u = v_{j+1}$ chosen according to the transition probabilities

$$p^{vu}(\alpha, \beta) \;=\; \alpha^{v_j} P^{vu} \beta^{v_j}. \tag{6}$$

The *undiscounted limit average effective payoff* (for player 1), when players play according to strategies α and β, is the Cesáro average

$$g^w(\Gamma, \alpha, \beta) \;=\; \liminf_{N \to \infty} \frac{1}{N+1} \sum_{j=0}^{\infty} \mathbb{E}[b_j], \tag{7}$$

where the expectation is taken over the random choices made (according to mixed strategies and transition probabilities) up to step j of the play. (When Γ is clear from the context, we shall simply write $g^w(\alpha, \beta)$.)

Let $\delta \in [0, 1)$ be a constant called the *discount factor*. In the *discounted* version, the payoff is discounted by a factor δ^i at step i of the play, and the *δ-discounted payoff* is given by the so-called *Abel-sum*

$$g^w_\delta(\Gamma, \alpha, \beta) = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \mathbb{E}[b_j]. \tag{8}$$

It is well-known that $g^w(\Gamma, \alpha, \beta) = \lim_{\delta \to 1} g^w_\delta(\Gamma, \alpha, \beta)$ (since the sequence $\{\mathbb{E}[b_j]\}_{j=0}^{\infty}$ is bounded; see, e.g., [HL31]) .

In both cases, the purpose of player 1 is to maximize $g^w(\Gamma, \alpha, \beta)$ (respectively, $g^w_\delta(\Gamma, \alpha, \beta)$), while player 2 would like to minimize it. For brevity in what follows, we will sometimes assume that $\delta \in [0.1]$, in which case we use $\delta = 1$ to denote the undiscounted case (for instance, $g_1(\alpha, \beta) = g(\alpha, \beta)$, etc). Unless stated explicitly otherwise, we will be considering undsicounted games.

1.3 Stationary Strategies

The strategy of the players is a sequence of mixed strategies, the selection of which may depend on the history, that is, all previous steps. More precisely, denoting by α^{v_j} and β^{v_j} the (mixed) strategies chosen by players 1 and 2 respectively, for $j = 0, 1, \ldots, i-1$, the strategies α^{v_i} and β^{v_i} at the ith step are functions of the sequence $v_0, \alpha^{v_0}, \beta^{v_0}, v_1, \alpha^{v_1}, \beta^{v_1}, v_{i-1}, \alpha^{v_{i-1}}, \beta^{v_{i-1}}, v_i$.

In 1981, Mertens and Neymann in their seminal paper [MN81] proved that every undiscounted stochastic game has value from any initial position in terms of history-dependent strategies. More precisely, they showed that for any $\epsilon > 0$, there exists a pair of (history-dependent) strategies $\alpha(\epsilon), \beta(\epsilon)$, such that $\alpha(\epsilon)$ guarantees player 1 a value of at least $g^v - \epsilon$ from the starting position v, while $\alpha(\epsilon)$ guarantees player 2 a value of at most $g^v + \epsilon$ from v. This common value g^v is called the value of the game starting from position v. If all the values are equal $g^v = c$ for all $v \in V$ and some constant c, we will say that the game is *ergodic*.

When α^{v_i} and β^{v_i} depend only on v_i, but not on the time or on the preceding positions or moves, they are called *stationary* strategies (if they depend only on the time and position, but not on the

history, they are called *Markovian*). Let us denote by $\mathcal{K}(\Gamma)$ and $\mathcal{L}(\Gamma)$ the sets of stationary strategies of WHITE and BLACK, respectively, that is

$$\mathcal{K}(\Gamma) = \bigotimes_{v \in V} \Delta(K^v) \quad \text{and} \quad \mathcal{L}(\Gamma) = \bigotimes_{v \in V} \Delta(L^v). \tag{9}$$

In this paper we will be concerned mainly with stationary strategies. For this case the mechanism of the game can be described in a simpler way. To a pair of stationary strategies $\alpha = (\alpha^v | v \in V) \in \mathcal{K}(\Gamma)$ and $\beta = (\beta^v | v \in V) \in \mathcal{L}(\Gamma)$ we associate a Markov chain $\mathcal{M}(\Gamma, \alpha, \beta)$ on states in V, defined by the transition probability matrix $P(\alpha, \beta) = (p^{vu}(\alpha, \beta))_{v,u \in V} = (\alpha^v P^{vu} \beta^v))_{v,u \in V}$. Then, this Markov chain has unique limiting probability distributions $Q(\alpha, \beta) = (q^{vu}(\alpha, \beta) |\ u \in V)$, where $q^{vu}(\alpha, \beta)$ is the probability of staying in state $u \in V$ when the initial vertex is $v \in V$. Then the undiscounted and discounted rewards of player 1 starting from vertex $v \in V$ can be computed, repsectively, as

$$g(\alpha, \beta) = Q(\alpha, \beta) a(\alpha, \beta), \tag{10}$$

$$g_\delta(\alpha, \beta) = (1 - \delta)(I - \delta P(\alpha, \beta))^{-1} a(\alpha, \beta), \tag{11}$$

where $g_\delta(\alpha, \beta) = (g^v_\delta(\alpha, \beta))_{v \in V}$, $a(\alpha, \beta) = (\alpha^u A^u \beta^u)_{u \in V}$, and I is the $|V| \times |V|$ identity matrix.

It is well-known that

$$\lim_{\delta \to 1^-} (1 - \delta)(I - \delta P(\alpha, \beta))^{-1} = Q(\alpha, \beta) \tag{12}$$

see, e.g., [How60, Bla62, MO70]. Thus, it follows that $\lim_{\delta \to 1^-} g_\delta(\alpha, \beta) = g(\alpha, \beta)$.

The purpose of the (discounted/undiscounted) game now can be formulated as to find stationary strategies $\alpha(v) \in \mathcal{K}(\Gamma)$ and $\beta(v) \in \mathcal{L}(\Gamma)$ for all initial states $v \in V$ such that

$$g^v_\delta(\alpha(v), \beta(v)) = \max_{\alpha \in \mathcal{K}(\Gamma)} g^v_\delta(\alpha, \beta(v)) = \min_{\beta \in \mathcal{L}(\Gamma)} g^v_\delta(\alpha(v), \beta). \tag{13}$$

We shall denote by $g^v_\delta(\Gamma)$, $v \in V$, the above optimum values, when exist. The existence of such values for the discounted case, $\delta \in [0, 1)$, was shown by Shapley [Sha53].

Let us note that the values g^v_δ, $v \in V$ depend on the entire game, not only on the local parameters. Typically $g^v_\delta(\Gamma) \neq$ Val (A^v) for all states $v \in V$.

Let us remark that for some (undiscounted) stochastic games we may not have the second equality in (13), in other words, max min and min max may be different. Furthermore, the function $g^v(\alpha, \beta)$ as a function of stationary strategies may not be continuous, and and even if the value $\sup \inf = \inf \sup$ exists, we may not have stationary strategies to realize it.

We include here two well known examples to demonstrate some of these difficulties.

Example 1.1 *Gilette [Gil57] introduced the so called* BIG MATCH *game, see Figure 1, to illustrate that the value in stationary strategies from some initial state may not exist in a stochastic game. In this game both* $(1, 0)$ *and* $(0, 1)$ *in state 1 are optimal stationary strategies for player 1, guaranteeing* 0 *for him, i.e.,* $0 = \max \min$. *However player 2 can only guarantee* $1/2$ *by choosing* $(1/2, 1/2)$ *as his strategy in state 1, implying* $1/2 = \min \max$.

Example 1.2 *Vrieze [Vri80, Chapter 8] showed an example, see Figure 2, for a stochastic game which has values, but in which only one of the players has optimal stationary strategies. It is easy to verify that this game has values* $g^1 = 0$ *and* $g^2 = 0$, *and the maximizer has optimal stationary strategies* $\{(1, 0), (0, 1)\}$, *while the minimizer (player 2) has no optimal stationary strategies. In fact for every* $0 < \epsilon \leq 1$ *the strategy* $(1 - \epsilon, \epsilon) \in \Delta(L^1)$ *guarantees at most* ϵ *for the minimizer. In the limit, if* $\epsilon \to 0$ *however white can respond by* $(1, 0) \in \Delta(K^1)$ *providing a value of* 1. *Thus, the* $\min \max$ *value does not exist in this game, and the minimizer can only guarantee with an appropriate stationary strategy at most* ϵ *for any* $\epsilon > 0$, *but not* 0.

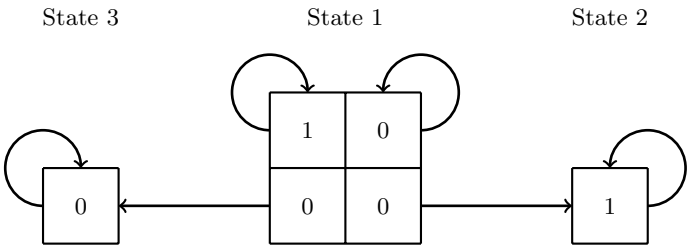


Figure 1: This game (Big Match) has three states $V = \{1, 2, 3\}$, and zero-one transition probabilities. State 1 has a 2×2 payoff matrix, the other two states are absorbing with payoffs equal to 1 and 0, respectively. Arrows in the picture indicate the nonzero transitions, e.g., if player 1 chooses the first row in state 1, and player 2 chooses the second column, then player 1 earns a local payoff of 0 and then the game returns to state 1; while if player 1 chooses the second row and player 2 chooses the first column then player 1 earns 0 and then the game moves to state 3, where it gets stuck forever, providing an effective payoff of 0 for player 1.

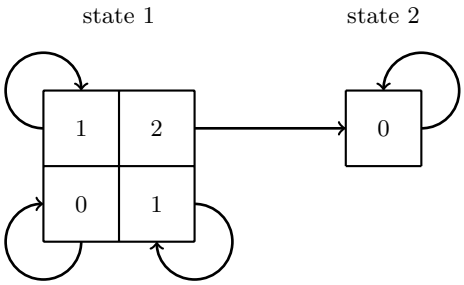


Figure 2: In this example we have $|K^1| = |L^1| = 2$, and $|K^2| = |L^2| = 1$. As before, arcs in the picture indicate the nonzero transitions, and all transition probabilities are zero or one. This game has values $g = (0, 0)$ and the maximizer has uniformly optimal stationary strategies. However the minimizer cannot guarantee 0 with a stationary strategy.

1.4 Potential Transformations

We shall consider a special family of transformations of stochastic games, called *potential transformations*. Such a transformed game is looking quite different from the original game in the sense that the reward matrices of the games at each state are different. However each of these transformed games remain completely equivalent with the original game. Namely, the transformed game has values if an only if the original one has; it has the same set of stationary (and non-stationary) strategies, and the values, as defined in (10), remain the same for all pairs of strategies (α, β).

To every vector $x \in \mathbb{R}^V$, and a discount factor $\delta \in (0, 1]$, we can associate a potential transformation by defining new local rewards as

$$r^{vu}_{k\ell}(x) \;=\; r^{vu}_{k\ell} + x^v - \delta x^u \tag{14}$$

for all $v, u \in V$, $k \in K^v$, and $\ell \in L^v$. This transformation is called discounted if $\delta < 1$ and undiscounted if $\delta = 1$. Undiscounted trasnsforms were first introduced in 1958 by Gallai [Gal58], then applied to stochastic games in 1966 by Hoffman and Karp [HK66] and to minimum cycle means in digraphs in 1978 by Karp [Kar78]. Discounted transforms were first mentioned in [GKK88] (page 87) and then considered in more detail in [ZP96].

Unless otherwise stated, we will be mainly dealing with undiscounted transformations. To indicate this role of the real vector x, we shall call it sometimes a *potential vector*. We denote by $\Gamma_\delta(x)$ the stochastic game which the same transition probabilities as Γ and which has its local rewards

defined by (14). We can also define the local payoff matrix for $\Gamma_\delta(x)$ by

$$(A^v(x))_{k\ell} \;=\; \sum_{u\in V} p^{vu}_{k\ell} r^{vu}_{k\ell}(x) \;=\; \sum_{u\in V} p^{vu}_{k\ell}(r^{vu}_{k\ell} + x^v - \delta x^u).$$

In other words, we have

$$A^v(x) \;=\; A^v + x^v J - \delta \sum_{u\in V} x^u P^{vu}, \tag{15}$$

where J stands for an all-ones matrix of size $|K^v| \times |L^v|$. Then, given any pair of strategies for the players, the one step expected payoff amount changes to $\mathbb{E}[b_j(x)] = \mathbb{E}[b_j] + \mathbb{E}[x^{v_j}] - \delta\mathbb{E}[x^{v_{j+1}}]$, where $v_j \in V$ is the (random) position at step j. However, the limit average payoff remains the same for all finite potentials:

$$g^{v_0}(\Gamma(x)) \;=\; g^{v_0}(\Gamma) + \lim_{N\to\infty} \frac{1}{N+1}\mathbb{E}[x^{v_0} - x^{v_{N+1}}] = g^{v_0}(\Gamma) \tag{16}$$

$$g^{v_0}_\delta(\Gamma_\delta(x)) \;=\; g^{v_0}_\delta(\Gamma) + (1-\delta)\sum_{j=0}^{\infty}(\delta^j\mathbb{E}[x^{v_j}] - \delta^{j+1}\mathbb{E}[x^{v_{j+1}}]) = g^{v_0}_\delta(\Gamma) + (1-\delta)x^{v_0}. \tag{17}$$

Thus, the transformed game $\Gamma_\delta(x)$ remains equivalent with the original one.

Remark 3 *Note that the equalities (16) and (17) are valid for any pair of history-dependent strategies. In the case of stationary strategies, we can also see directly that potential transformations do not change the game. Indeed, let us observe that $\mathcal{K}(\Gamma_\delta(x)) = \mathcal{K}(\Gamma)$, $\mathcal{L}(\Gamma_\delta(x)) = \mathcal{L}(\Gamma)$, and for any pair of stationary strategies (α,β) we get the same Markov chain $\mathcal{M}(\alpha,\beta) = \mathcal{M}(\Gamma,\alpha,\beta) = \mathcal{M}(\Gamma_\delta(x),\alpha,\beta)$ in both games. Consequently, we get the same limiting probability distributions $q^{vu}(\alpha,\beta) = q^{vu}(\Gamma,\alpha,\beta) = q^{vu}(\Gamma(x),\alpha,\beta)$. Thus, writing $a(x) = (\alpha^v A^v(x)\beta^v)_{v\in V}$, we get $a(x) = a + (I - \delta P(\alpha,\beta))x$, and according to (10), the vector of values $g(\alpha,\beta)$ in $\Gamma(x)$ can be written as*

$$\begin{aligned} g(\Gamma(x),\alpha,\beta) \;&=\; Q(\alpha,\beta)a(x) = Q(\alpha,\beta)\left(a + (I - P(\alpha,\beta))x\right) \\ &=\; Q(\alpha,\beta)a + Q(\alpha,\beta)x - QP(\alpha,\beta)x = Q(\alpha,\beta)a = g(\Gamma,\alpha,\beta). \end{aligned}$$

The fourth equality follows by the fact that $Q(\alpha,\beta)$ is the limiting distribution. Similarly, for the discounted case, we have by (11)

$$\begin{aligned} g_\delta(\Gamma(x),\alpha,\beta) \;&=\; (1-\delta)(I - P(\alpha,\beta))^{-1}a(x) \\ &=\; (1-\delta)(I - P(\alpha,\beta))^{-1}\left(a + (I - P(\alpha,\beta))x\right) \\ &=\; g_\delta(\Gamma,\alpha,\beta) + (1-\delta)x. \end{aligned}$$

Thus, the above shows that, indeed, in Γ and $\Gamma(x)$ we have the same undiscounted values associated to any pair of strategies, while the discounted values in Γ and $\Gamma_\delta(x)$ differ by a constant that depends only on the potential vector but is independent of the strategies.

Given a stochastic game Γ and subsets of strategies $\mathcal{Q} = \{Q^v \subseteq \Delta(K^v) \mid v \in V\}$ and $\mathcal{R} = \{R^v \subseteq \Delta(L^v) \mid v \in V\}$, we define the *restriction* $\Gamma[\mathcal{Q},\mathcal{R}]$ as the stochastic game with reward matrices $A^v[Q^v,R^v]$ and transition probabilities $P^{vu}[Q^v,R^v]$, $u \in V$, for all states $v \in V$.

Remark 4 *Let $x \in \mathbb{R}^V$ be a potential vector, and let $\mathcal{Q}$ and $\mathcal{R}$ be subsets of strategies as above. Then,* (1) *and* (15) *imply*

$$\Gamma(x)[\mathcal{Q},\mathcal{R}] \;\equiv\; \Gamma[\mathcal{Q},\mathcal{R}](x).$$

1.5 Uniformly Best Response

Given a stationary strategy, e.g., of the maximizer $\alpha \in \mathcal{K}(\Gamma)$ in a stochastic game Γ, we can compute for each initial state $v \in V$ a best response $\beta(\alpha, v) \in \mathcal{L}(\Gamma)$ of the minimizer, i.e., for which we have

$$g^v(\alpha, \beta(\alpha, v)) \;=\; \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\alpha, \beta).$$

We call $\beta(\alpha) \in \mathcal{L}(\Gamma)$ a *uniformly best response* of the minimizer, if the equalities

$$g^v(\alpha, \beta(\alpha)) = \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\alpha, \beta)$$

hold simultaneously for all states $v \in V$. Of course, we can interchange the roles of maximizer and minimizer, and define analogously the best response of the maximizer.

The following claims are well known. For completeness, we include the proofs in Appendix B.

Lemma 1 *In a stochastic game Γ there exists a uniformly best response for every stationary strategy. In fact this best response can be chosen as a pure strategy at every state.*

Let us denote by $\beta(\alpha)$ and $\alpha(\beta)$ some arbitrarily chosen uniformly best responses to $\alpha \in \mathcal{K}(\Gamma)$ of the minimizer and to $\beta \in \mathcal{L}(\Gamma)$ of the maximizer, respectively.

Lemma 2 *Given a stochastic game Γ and two stationary strategies of the maximizer $\alpha, \alpha' \in \mathcal{K}(\Gamma)$ there exists a third strategy $\alpha'' \in \mathcal{K}(\Gamma)$ such that*

$$g^v(\alpha'', \beta(\alpha'')) \;\geq\; \max\{g^v(\alpha, \beta(\alpha)),\, g^v(\alpha', \beta(\alpha'))\} \quad \textit{for all } v \in V.$$

Analogously, for any two strategies $\beta, \beta' \in \mathcal{L}(\Gamma)$ of the minimizer there exists a third strategy $\beta'' \in \mathcal{L}(\Gamma)$ such that

$$g^v(\alpha(\beta''), \beta'') \;\leq\; \min\{g^v(\alpha(\beta), \beta),\, g^v(\alpha(\beta'), \beta')\} \quad \textit{for all } v \in V.$$

Corollary 1 *If in a stochastic game Γ we have max-min strategies $\alpha(v) \in \mathcal{K}(\Gamma)$ for all initial states $v \in V$, then there exists a* uniform max-min *strategy $\bar{\alpha} \in \mathcal{K}(\Gamma)$, i.e., for which*

$$\min_{\beta \in \mathcal{L}(\Gamma)} g^v(\bar{\alpha}, \beta) \;=\; \max_{\alpha \in \mathcal{K}(\Gamma)} \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\alpha, \beta) \tag{18}$$

holds for all states $v \in V$ simultaneously. Analogously, if in Γ we have min-max strategies $\beta(v) \in \mathcal{L}(\Gamma)$ for all initial vertices $v \in V$, then there exists a uniform min-max *strategy $\bar{\beta} \in \mathcal{L}(\Gamma)$, i.e., for which*

$$\max_{\alpha \in \mathcal{K}(\Gamma)} g^v(\alpha, \bar{\beta}) \;=\; \min_{\beta \in \mathcal{L}(\Gamma)} \max_{\alpha \in \mathcal{K}(\Gamma)} g^v(\alpha, \beta) \tag{19}$$

holds simultaneously for all states $v \in V$.

Proof It is enough to prove the statement for the maximizer, since the claim is symmetric. By applying Lemma 2 repeatedly, we can derive the existence of a strategy $\bar{\alpha} \in \mathcal{K}(\Gamma)$ such that

$$g^v(\bar{\alpha}, \beta(\bar{\alpha})) \;\geq\; \max\{g^v(\alpha(u), \beta(\alpha(u))) \mid u \in V\} \;=\; g^v(\alpha(v), \beta(\alpha(v)))$$

holds for all $v \in V$. Since $\alpha(v)$ is a max-min strategy for initial state v, we must have equalities in the above relations for all states $v \in V$. In other words, $\bar{\alpha}$ provides the best value for the maximizer in all states simultaneously, and hence it is a uniform max-min strategy. □

The following claim states that if a pair of stationary strategies have the property that each player is playing optimally against the opponent's stationary strategies, then it remains optimal, even if each the opponent is given the full power of playing history-dependent strategies.

Corollary 2 *Suppose that for some state v and for a pair of stationary strategies $(\bar{\alpha}, \bar{\beta}) \in \mathcal{K}(\Gamma) \times \mathcal{L}(\Gamma)$, it holds that*

$$\max_{\alpha \in \mathcal{K}(\Gamma))} g^v(\Gamma, \alpha, \bar{\beta}) = \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\Gamma, \bar{\alpha}, \beta) = g^v(\Gamma, \bar{\alpha}, \bar{\beta}).$$

Then $(\bar{\alpha}, \bar{\beta})$ are optimal strategies and $g^v(\Gamma, \bar{\alpha}, \bar{\beta})$ is the value of the game starting from state v.

Proof This follows from the fact that if we fix one player's strategy, say $\bar{\alpha}$, then we obtain a Markov decision process, in which there is an optimal stationary strategy. In other words, $\min_\beta g^v(\Gamma, \bar{\alpha}, \beta) = \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\Gamma, \bar{\alpha}, \beta)$, where the first minimum is over history-dependent strategies (See Appendix B.) □

1.6 Uniformly Optimal Strategies

A stochastic game is called *subgame perfect*, if it has stationary strategies $\bar{\alpha} \in \mathcal{K}(\Gamma)$ and $\bar{\beta} \in \mathcal{L}(\Gamma)$ for which

$$g^v(\bar{\alpha}, \bar{\beta}) \;=\; \max_{\alpha \in \mathcal{K}(\Gamma)} g^v(\alpha, \bar{\beta}) \;=\; \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\bar{\alpha}, \beta) \tag{20}$$

holds simultaneously for all states $v \in V$. Such stationary strategies are called *uniformly optimal*.

The following claim is well known.

Corollary 3 *If in a stochastic game Γ we have both max-min and min-max strategies and they provide the same g^v value for all initial states $v \in V$, then Γ is subgame perfect.*

Proof By Corollary 1 we have stationary strategies $\bar{\alpha}$ and $\bar{\beta}$ satisfying (18) and (19). If we additionally assume that the the max-min and min-max values are the same g^v for every state $v \in V$ then we must have $g^v = g^v(\bar{\alpha}, \bar{\beta})$ for all states $v \in V$, implying that these strategies are uniformly optimal. □

The main focus of this paper is to prove the existence of a canonical form (which will be defined in Section 2) for some classes of stochastic games. We know that the existence of the canonical form for a stochastic game Γ implies by a theorem of Vrieze [Vri80, Theorem 8.1.8] that Γ is subgame perfect (for an independent proof see Theorem 1 below). Therefore, we may restrict our attention, without any loss of generality to subgame perfect games, that is, to games satisfying the condition

(A1) Γ has values $g \in \mathbb{R}^V$ and uniformly optimal stationary strategies $\bar{\alpha} \in \mathcal{K}(\Gamma)$ and $\bar{\beta} \in \mathcal{L}(\Gamma)$ for the maximizer and minimizer, respectively.

1.7 Special Cases

The following classes of stochastic games are known to be solvable in uniformly optimal stationary strategies:

- Stochastic games with perfect information (PI-games): in this class, introduced by Gillette [Gil57], the set of states is partitioned in two sets $V = V_W \cup V_B$; V_W is controlled by WHITE, and V_B is controlled by BLACK. It is assumed that $|L^v| = 1$ for all $v \in V_W$ and $|K^v| = 1$ for all $v \in V_B$. The fact that a *saddle point* exists in *pure positional strategies*, in this class of games, was proved by Gillette [Gil57] and Liggett and Lippman [LL69]. If one of the sets V_B or V_W is empty, we obtain a *Markov decision process*; see, for example, [MO70], and if both are empty $V_B = V_W = \emptyset$, we get a *weighted Markov chain*.

- BWR-games: In [BEGM09], we considered another special class of games, which was first suggested in [GKK88], and recently considered under the name of *Stochastic Mean payoff games* in [CH08], and which is (polynomially)equivalent with the class of PI-games. Each such game, which we call a *BWR-game*, is played by two players, WHITE and BLACK, on an arc-weighted directed graph $G = (V = V_B \cup V_W \cup V_R, E)$, with given local rewards $r \in \mathbb{R}^E$ and 3 types of vertices: Black V_B, controlled by BLACK; White V_W, controlled by WHITE; and Random V_R, controlled by nature. When the play is at a white (black) vertex, WHITE (resp., BLACK) selects a outgoing arc and BLACK pays WHITE the reward on that arc. When the play is at a random vertex v, a vertex u is picked with specified probability p^{vu} and again BLACK pays WHITE the value r^{vu} on the arc (v, u). The play continues forever, and WHITE aims to maximize (BLACK aims to minimize) the limiting average payoff, defined as in (7). The special case when there are no random nodes, is known as *cyclic games* or *mean payoff games* (or BW-games), which were initially considered for complete bipartite digraphs in [Mou76b, Mou76a], for all (not necessarily complete) bipartite digraphs in [EM79], and for arbitrary digraphs in [GKK88]. A further special case of this was considered extensively in the literature under the name of *parity games* [BV01a, BV01b, CJH04, Hal07, Jur98, JPZ06], and later generalized also to include random nodes in [CH08]. The game is reduced to the *minimum mean cycle problem* in case $V_W = V_R = \emptyset$, see for example [Kar78]. In the special case of a BWR-game when all rewards are zero except at a single node t called the terminal, at which there is a self-loop with reward 1, we obtain the so-called *simple stochastic games* (SS-game), introduced by Condon [Con92]. In these games, the objective of White is to maximize the probability of reaching the terminal while Black wants to minimize this probability.

- Switching controller games (SC-games): in this class, suggested first by Maschler (see Filar [Fil81]), the set of states is partitioned in two sets V_W, V_B: it is assumed that $p^{vu}_{k\ell} = \psi^{vu}_k$ for all $v \in V_W$ and $p^{vu}_{k\ell} = \gamma^{vu}_\ell$ for all $v \in V_B$, that is, the transition probabilities at V_W and V_B are controlled by WHITE, and BLACK, respectively. The existence of a saddle point in uniformly optimal stationary strategies was shown by Filar [Fil81] and Bewley and Kohlberg [BK78]. Clearly this class includes the class of PI-games.

- Additive rewards, additive transition probabilities (ARAT-games): this is the case when, for all $v \in V$, $r^v_{k\ell} = q^v_k + s^v_\ell$, and $p^{k\ell}_{vu} = \lambda^v \psi^{vu}_k + (1 - \lambda^v)\gamma^{vu}_\ell$, for some constants $0 \le \lambda^v \le 1$, that is, both the rewards and transition probabilities are *separable* into two parts, one controlled by WHITE, and the other controlled by BLACK. The existence of a saddle point in uniformly optimal stationary strategies was shown by Raghavan, Tijs and Vrieze [RTV85]. This class also includes the class of PI-games.

- Additive transition probabilities (AT-games): in this case, only the transition probabilities are separable, that is, for all $v \in V$, $p^{k\ell}_{vu} = \lambda^v \psi^{vu}_k + (1 - \lambda^v)\gamma^{vu}_\ell$, for some constants $0 \le \lambda^v \le 1$. This class generalizes both SC-games and ARAT-games. The fact that saddle point exists in uniformly optimal stationary strategies was only shown quite recently by Flesch, Thuijsman and Vrieze [FTV07].

A class of stochastic games is said to possess the *ordered field property* if for any game in the class, with *rational* rewards and transition probabilities, there exists a *rational* pair of optimal strategies. Among the above classes, only SC-games (and hence PI- and BWR-games), and ARAT-games are known to have this property. On the other hand, for AT-games, there exists an example showing that this property does not hold [RTV85]. The interest in the ordered filed property stems from the fact that the accuracy ε needed to solve the game is bounded *exponentially* in the input size, and hence any algorithm that runs in time polynomial in $\log \frac{1}{\varepsilon}$ (and the other input parameters) and approximates the game value within ε will be *weakly polynomial*.

1.8 Value Vectors in Stochastic Games

In the sequel we shall assume that the stochastic game Γ has values in stationary strategies. Then, we can associate to every state $v \in V$ its value g^v. In the sequel our main focus will be on determining

these values, or for a given vector $g \in \mathbb{R}^V$ to verify if that is indeed the vector of values of the given game. To emphasize this role of the vector g, we shall call it sometimes a *value vector*.

To a stochastic game Γ and a value vector $g \in \mathbb{R}^V$, we associate the *value* matrices $G^v(g) \in \mathbb{R}^{K^v \times L^v}$ for all states $v \in V$, defined as follows:

$$(G^v(g))_{k\ell} \;=\; \sum_{u \in V} p^{vu}_{k\ell} g^u. \tag{21}$$

If the value vector g is indeed the vector of values of Γ, then we shall also refer to $G^v(g)$ as $G^v(\Gamma)$, or simply G^v when it is not ambiguous.

To an arbitrary value vector $g \in \mathbb{R}^V$ let us associate $\bar{K}^v = \bar{K}^v(g) = \bar{K}^v(g, \Gamma) = \Omega(G^v(g)) \subseteq \Delta(K^v)$ and $\bar{L}^v = \bar{L}^v(g) = \bar{L}^v(g, \Gamma) = \Lambda(G^v(g)) \subseteq \Delta(L^v)$, the sets of extremal optimal mixed strategies, for all states $v \in V$ as defined in Subsection 1.1. It is easy to see that a necessary condition for g to be the vector of values, in stationary strategies $\bar{\alpha} \in \mathcal{K}(\Gamma)$ and $\bar{\beta} \in \mathcal{L}(\Gamma)$, is that

$$g^v \;=\; \text{Val}\,(G^v(g)), \quad \text{and } \alpha^v \in \bar{K}^v \text{ and } \beta^v \in \bar{L}^v \text{ for all states } v \in V. \tag{22}$$

It was shown by Federgruen [Fed80] (see also Vrieze [Vri80, Lemma 8.1.3]) that every ergodic stochastic game satisfying condition (A1) has a potential transformation such that the local value becomes equal to the global value at each state. We will need a slightly reformulated version, thus for the sake of completeness, we restate this claim and provide a short proof in the appendix.

Lemma 3 *Assume that a stochastic game Γ satisfies condition (A1) and let $g = g(\Gamma) \in \mathbb{R}^V$ be its value vector. Then, there exists a potential vector $x \in \mathbb{R}^V$ satisfying*

$$g^v \;=\; \text{Val}\,(A^v(x)) \qquad \textit{for all states} \quad v \in V. \tag{23}$$

Proof See Appendix C. □

Let us note that equalities (23) does not imply the existence of a canonical form, in general. Though such a potential transformation is helpful in the sense that the local matrix game values are equal with the global values achievable from the given state, it may still be very difficult to verify if those values are indeed the global game values, and finding the optimal stationary strategy remains also difficult. The following example demonstrates that despite the above equalities, the locally (unique) optimal strategies may still be far from being globally optimal.

Example 1.3 *In Figure 3 we show a small BW game which satisfies the value equalities in* (23) *with the zero potential, and in which the locally optimal (unique) strategies are not globally optimal.*

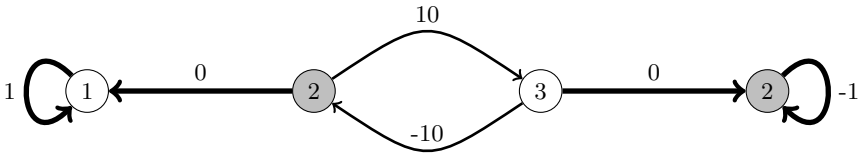


Figure 3: In this BW game we have 4 states, white nodes representing states controlled by the maximizer and black (gray) states are controlled by the minimizer. All transitions are zero or one. Local rewards are indicated along the arcs. In this example with $x = 0 \in \mathbb{R}^4$ we have the equalities (23) hold with values $g = (1, 0, 0, -1)$ for the states. Thick arcs in the picture indicate the locally optimal strategies, which are however not globally optimal, since they provide e.g., for states 2 and 3 the incorrect values of 1 and -1, respectively. The optimal strategies are in fact the thin arcs $(2, 3)$ and $(3, 2)$ from these states, guaranteeing the correct 0 value for them.

Remark 5 *For discounted stochastic games, a saddle point always exists in uniformly optimal strategies, and the vector of values g_δ satsifies the so-called Shapley equations [Sha53]:*

$$g^v_\delta \;=\; \text{Val}\,((1-\delta)A^v + \delta G^v(g_\delta)), \qquad \textit{for all } v \in V. \tag{24}$$

2 Canonical Form

Let $\Gamma = (p^{vu}_{k\ell}, r^{vu}_{k\ell} |\ k \in K^v, \ell \in L^v, v, u \in V)$ be a stochastic game. We will consider the following properties:

(B1) There is a potential transformation $x \in \mathbb{R}^V$ such that every locally optimal pair of strategies in $\Gamma(x)$ is also globally optimal, and the local and global values are equal at each state, that is, $\forall \bar{\alpha}, \bar{\beta} \in \mathcal{K}(\Gamma) \times \mathcal{L}(\Gamma)$:

$$\left(\max_{\alpha^v \in \Delta(K^v)} \alpha^v A^v(x) \bar{\beta}^v = \min_{\beta^v \in \Delta(L^v)} \bar{\alpha}^v A^v(x) \beta^v = \bar{\alpha}^v A^v(x) \bar{\beta}^v \quad \forall v \in V \right) \quad \Rightarrow$$
$$\left(\max_{\alpha \in \mathcal{K}(\Gamma)} g^v(\Gamma, \alpha, \bar{\beta}) = \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\Gamma, \bar{\alpha}, \beta) = g^v(\Gamma, \bar{\alpha}, \bar{\beta}) = \bar{\alpha}^v A^v(x) \bar{\beta}^v \quad \forall v \in V \right).$$

(B2) There exist $g \in \mathbb{R}^V$ and two potential vectors $x, y \in \mathbb{R}^V$ such that for all $v \in V$, there is an optimal pair of strategies for G^v that guarantees White and Black, respectively, a value of $g^v = \mathrm{Val}(G^v(g))$ in the transformed games $A^v(x)$ and $A^v(y)$:

$$\forall v \in V: \ g^v = \operatorname*{Val}_{K^v \times L^v}(G^v(g)) = \operatorname*{Val}_{\bar{K}^v(g) \times L^v}(A^v(x)) = \operatorname*{Val}_{K^v \times \bar{L}^v(g)}(A^v(y)). \tag{25}$$

(B3) There exist $g \in \mathbb{R}^V$ and a potential vector $x \in \mathbb{R}^V$ such that (B2) is satisfied with $x = y$, that is, for all $v \in V$, there is an optimal pair of strategies for G^v, which is also optimal for the transformed matrix game $A^v(x)$, and g^v is the value of both matrix games:

$$\forall v \in V: \ g^v = \operatorname*{Val}_{K^v \times L^v}(G^v(g)) = \operatorname*{Val}_{\bar{K}^v(g) \times L^v}(A^v(x)) = \operatorname*{Val}_{K^v \times \bar{L}^v(g)}(A^v(x)).$$

(B4) There exist $g \in \mathbb{R}^V$ and a potential vector $x \in \mathbb{R}^V$ such that for all $v \in V$, *every* optimal pair of strategies for $A^v(x)$ is also optimal for $G^v(g)$, and g^v is the value of both matrix games:

$$\text{(B4-i)} \quad \forall v \in V: \ g^v = \operatorname*{Val}_{K^v \times L^v}(G^v(g)) = \operatorname*{Val}_{K^v \times L^v}(A^v(x)),$$

$$\text{(B4-ii)} \quad \forall v \in V: \ \Omega(A^v(x)) \times \Lambda(A^v(x)) \subseteq \Omega(G^v(g)) \times \Lambda(G^v(g)).$$

(B5) There exist $g \in \mathbb{R}^V$, a potential vector $x \in \mathbb{R}^V$, and a pair of strategies $(\bar{\alpha}, \bar{\beta}) \in \Omega(G^v(g)) \times \Lambda(G^v(g))$, such that

$$\forall v \in V: \ g^v = \operatorname*{Val}_{K^v \times L^v}(G^v(g)) = \operatorname*{Val}_{\{\bar{\alpha}^v\} \times \hat{L}^v}(A^v(x)) = \operatorname*{Val}_{\hat{K}^v \times \{\bar{\beta}^v\}}(A^v(x)),$$

where $\hat{K}^v = \{k \in K^v: \ (G^v(g)\bar{\beta}^v)_k = g^v\}$ and $\hat{L}^v = \{\ell \in L^v: \ (\bar{\alpha}^v G^v(g))_\ell = g^v\}$.

(B6) There exists a potential vector $x \in \mathbb{R}^V$ such that

$$\forall v \in V: \ g^v(\Gamma) = \operatorname*{Val}_{K^v \times L^v}(A^v(x)).$$

Theorem 1 *The following implications hold for any stochastic game* Γ*:*

[I1]: $(A1) \Leftrightarrow (B2)$ *[I2]:* $(B3) \Rightarrow (A1)$ *[I3]:* $(B4) \Rightarrow (B3)$
[I4]: $(B1) \Leftrightarrow (B4)$ *[I5]:* $(B5) \Rightarrow (B3)$ *[I6]:* $(A1) \Rightarrow (B6)$

Proof of [I1]. This is Theorem 8.1.8 in [Vri80]. We give an elementary proof of the implication $(B2) \Rightarrow (A1)$ in the appendix.

Proof of [I2]. It follows immediately from [I1] by setting $x = y$ in $(B2)$.

Proof of [I3]. This follows immediately since (B4-ii) implies

$$\underset{K^v \times L^v}{\text{Val}}(A^v(x)) = \underset{\bar{K}^v \times L^v}{\text{Val}}(A^v(x)) = \underset{K^v \times \bar{L}^v}{\text{Val}}(A^v(x)).$$

Proof of [I4]. Suppose that (B1) holds. Then there exist $g, x \in \mathbb{R}^V$ such that for all $\bar{\alpha} \in \mathcal{K}(\Gamma)$, $\bar{\beta} \in \mathcal{L}(\Gamma)$, and all states $v \in V$, if $\bar{\alpha}^v A^v(x)\beta^v \geq g^v$ for all $\beta^v \in \Delta(L^v)$ and $\alpha^v A^v(x)\bar{\beta}^v \geq g^v$ for all $\alpha^v \in \Delta(K^v)$, then $g^v(\Gamma) = g^v(\Gamma, \bar{\alpha}, \bar{\beta}) = g^v$. In view of (22), the latter statement implies that $g^v = \text{Val}\,(G^v(g))$ and $(\bar{\alpha}, \bar{\beta})$ is an optimal pair of strategies in the matrix game $G^v(g)$. Thus, we get (B4).

Suppose on the other hand that (B4) holds. Let us choose for every state $v \in V$, an arbitrary pair of locally optimal strategies $(\bar{\alpha}^v, \bar{\beta}^v)$ in the matrix game $A^v(x)$. By the implications (B4)⇒(B3)⇒(B2)⇒(A1) (and the proof of (B2)⇒(A1)), it follows that $(\bar{\alpha}, \bar{\beta})$ is a globally optimal strategy guaranteeing the vector of values g in Γ .

Proof of [I5]. Note that the $g, x, \bar{\alpha}, \bar{\beta}$ satisfy

$$(\bar{\alpha}^v G^v(g))_\ell \geq g^v \text{ for all } \ell \in L^v, \text{ and } (G^v(g)\bar{\beta}^v)_k \leq g^v \text{ for all } k \in K^v, \tag{26}$$

$$(\bar{\alpha}^v A^v(x))_\ell \geq g^v \text{ for all } \ell \in \hat{L}^v, \text{ and } (A^v(x)\bar{\beta}^v)_k \leq g^v \text{ for all } k \in \hat{K}^v. \tag{27}$$

In order to satisfy (B3), we modify the vector of potentials x into $\hat{x} = x - C \cdot g$, where $C \geq 0$ is chosen sufficiently large. Clearly, it is enough to show that $(\bar{\alpha}_v A^v(\hat{x}))_\ell \geq g^v$ for all $\ell \in L^v$ and $(A^v(\hat{x})\bar{\beta}^v)_k \leq g^v$ for all $k \in K^v$. We show the first set of inequalities; the second set can be shown similarly.

Consider any $\ell \in L^v$; then $(\bar{\alpha}^v A^v(\hat{x}))_\ell = (\bar{\alpha}^v A^v(x))_\ell - C(g^v - (\bar{\alpha}^v G^v(g))_\ell)$. Thus, if $\ell \in \hat{L}^v$, then $(\bar{\alpha}^v A^v(\hat{x}))_\ell = (\bar{\alpha}^v A^v(x))_\ell$ and hence (27) already implies the claim. On the other hand, if $\ell \notin \hat{L}^v$, then (26) implies that $g^v - (\bar{\alpha}^v G^v(g))_\ell < 0$, and hence the statement follows if we choose $C \geq 0$ sufficiently large.

Proof of [I6]. This is just Lemma 3 above. □

The central definition of this paper is the following.

Definition 1 *We say that a stochastic game* Γ *admits a canonical form if satisfies property (B1) (or equivalently(B4)).*

It is known that for cyclic games (the case of BWR-games when $V_R = \emptyset$), there exists such a transformation such that, in the transformed game, the *locally* optimal strategies are *globally* optimal, and hence, the value and optimal strategies become obvious [GKK88]. For the special case of Markov decision processes (the case of PI-games when V_B or V_W is empty), the potentials mentioned in the theorem correspond to the *dual* variables in the standard linear programming formulation; see e.g. [MO70] and also Appendix B.

As illustrated in Example 1.3, property (B6) is not sufficient to imply (B1). In the ergodic case, i.e., when all the values are equal, it becomes sufficient. This suggests the following defintion.

Definition 2 *We say that a discounted/undiscounted stochastic game* Γ *admits an* ergodic canonical form *if there exist a constant* $c \in \mathbb{R}$ *and a potential vector* $x \in \mathbb{R}^V$ *such that*

$$\forall v \in V : \; c = \underset{K^v \times L^v}{\text{Val}}(A^v(x)). \tag{28}$$

If game admits an ergodic canonical form, then it easy to see that property (B1) holds.

Proposition 1 *If a discounted/undiscounted stochastic* Γ *game admits an ergodic canonical form (28) then (i) every locally optimal strategy is globally optimal and (ii) If* $\delta = 1$*, the game is ergodic:* c *is its value for every initial position* $v_0 \in V$*.*

Proof Let us apply the potential transformation x that brings Γ to ergodic canonical form (28). Now consider the game $\Gamma_\delta(x)$. If WHITE (BLACK) applies a locally optimal strategy then after every own move (s)he will get (pay) an expected value of c, while for each move of the opponent the expected local reward will be at least (at most) c. Thus, if both players choose their locally optimal strategies then the expected local reward b_i equals c for every step i. Hence, the δ-discounted value is $(1-\delta)\sum_{i=0}^{\infty} c\delta^i = c$ for each $\delta \in [0,1)$ and its limit, as $\delta \to 1^-$, which is the undiscounted value equals c, too. The statements then follow from (16) and (17). □

In Section 5 we will give an algorithmic proof for the existence of canonical form for the discounted case.

Theorem 2 *Given a stochastic game Γ and discount factor $\delta < 1$, there is a discounted potential transformation (14) reducing Γ to an ergodic canonical form.*

The existence of canonical form may give rise to new algorithms for solving stochastic games. For instance, in the absence of random positions in a BWR-game (that is, for a BW-game), a pseudo-polynomial algorithm reducing the game to canonical form was given in [GKK88], see also [Pis99]. For BWR-games with constant number of random nodes, a pseudo-polynomial algorithm was given in [BEGM10b] to reduce any ergodic BWR-game to canonical form, and this algorithm has been extended in [BEGM11] to general ergodic stochastic games (whenever they admit a canonical form). We note these algorithms do not go through discounting, and thus yield direct algorithmic proofs of the existence of canonical form in the ergodic case. Obtaining a similar direct result and a corresponding algorithm in the general case remains a challenging open problem.

Remark 6 *Canonical forms can also be motivated by the so-called* certifying algorithms *(see e.g. [KMMS03]). For instance, if the game satisfies (B1) then the potential x can be used as a certificate for optimality: to do this, we transform the game and then compute, at each state v, a locally optimal pair of strategies (α^v, β^v). This gives a vector of values g, with g^v being the value of the local game at state v. Once, we have g and a pair of strategies (α, β), we can verify optimality by solving two Markov decision processes. If a game satisfies (B2) (or (B3)), then to certify optimality one needs, in addition to the potential vectors x and y, the global vector of values g, and a pair of locally optimal strategies (α, β).*

Remark 7 *It is worth noting that the decision problem of checking if a game Γ satisfies (B2) is in NP, since given $x, y, g \in \mathbb{R}^V$, we can verify (25) by solving a pair of linear programs. A similar remark holds for checking (B4-i), given $x, g \in \mathbb{R}^V$. Checking (B4-ii) requires checking if the polytopes $\Omega(A^v(x))$ and $\Lambda(A^v(x))$ (given by their facet descriptions) are subsets of $\Omega(G^v(g))$ and $\Lambda(G^v(g))$, respectively. This can be done by solving a number of $O(|K^v| + |L^v|)$ linear programs.*

3 A Sufficient Condition for the Existence of Canonical Form in Stochastic Games

In this section we provide a sufficient condition for the existence of the canonical form in subgame perfect stochastic games.

Let us assume now that Γ satisfies condition (A1), and consider the value vector $g = g(\Gamma)$, the corresponding value matrices $G^v = G^v(\Gamma)$, and strategy sets $\bar{K}^v = \bar{K}^v(\Gamma) = \Omega(G^v(\Gamma))$ and $\bar{L}^v = \bar{L}^v(\Gamma) = \Lambda(G^v(\Gamma))$ for all states $v \in V$.

We shall show in this section that Γ can be brought to its canonical form by a suitable potential transformation if, in addition to (A1), one of the following condition holds:

(A2) There exists an $\epsilon > 0$, such that for all states $v \in V$ and for all strategies $\bar{\alpha}^v \in \bar{K}^v$, $\bar{\beta}^v \in \bar{L}^v$ we have

$$\begin{aligned} (\bar{\alpha}^v G^v(g))_\ell &\geq g^v + \epsilon \text{ for all } \ell \in L^v \setminus \bar{L}^v, \text{ and} \\ \left(G^v(g)\bar{\beta}^v\right)_k &\leq g^v - \epsilon \text{ for all } k \in K^v \setminus \bar{K}^v. \end{aligned}$$

Let us remark that given a value vector $g \in \mathbb{R}^V$ condition (A2) can be tested efficiently by using linear programming (and by using a rational approximation of the potentially irrational g^v values).

We are ready now to state the main result of this section.

Theorem 3 *If a stochastic game Γ satisfies conditions (A1) and (A2), then it admits a canonical form (in short, (A1)∧(A2)⇒(B1)).*

In the rest of this section, we give a proof of Theorem 3, an use it in the next section to show that AT-games admit a canonical form. In Section 5, we show that every discounted stochastic game can be brought into canonical form by a potential transformation, and use this fact in Section 6 to give an independent proof (which does not use the result in [FTV07]) that AT-games satisfy condition (A1)) that undiscounted BWR-games also admit a canonical form. We shall then use this result to give an independent proof that PI-games, SC-games, and ARAT-games (which satisfy condition (A2)) admit a canonical form.

Proof of Theorem 3. Let us first consider the strategy sets $\mathcal{K}^* = \{\bar{K}^v(\Gamma) \mid v \in V\}$ and $\mathcal{L}^* = \{\bar{L}^v(\Gamma) \mid v \in V\}$, and define the restricted game $\Gamma^* = \Gamma[\mathcal{K}^*, \mathcal{L}^*]$ as in Section 1.4.

Let us then note by (22) that Γ^* has the same vector of values g as Γ and has the same set of optimal stationary strategies. Consequently, any uniformly optimal strategy of Γ is also uniformly optimal for Γ^*. Thus, since Γ satisfies condition (A1), so does Γ^*. Therefore, we can apply Lemma 3 for the stochastic game Γ^* and obtain a potential vector $x \in \mathbb{R}^V$ satisfying

$$g^v \;=\; \text{Val}\ (A^v[\mathcal{K}^*, \mathcal{L}^*](x)) \;=\; \text{Val}\ (A^v(x)[\mathcal{K}^*, \mathcal{L}^*]) \quad \text{for all states}\ \ v \in V, \tag{29}$$

where the second equation follows by Remark 4.

Let us next apply a special potential transformation, modifying the potential vector by

$$\hat{x} \;=\; x \;-\; Cg$$

for a suitably large constant $C \geq 0$.

We claim that for all $\bar{\alpha}^v \in \bar{K}^v(g)$ and $\bar{\beta}^v \in \bar{L}^v(g)$ and suitably large C we have

$$\begin{aligned} (\bar{\alpha}^v A^v(\hat{x}))_\ell \;&>\; g^v \text{ for all } \ell \in L^v \setminus \bar{L}^v(g) \text{ and} \\ \left(A^v(\hat{x})\bar{\beta}^v\right)_k \;&<\; g^v \ \text{ for all } k \in K^v \setminus \bar{K}^v(g). \end{aligned} \tag{30}$$

for all states $v \in V$.

To see these, let us note that for $\alpha^v \in \bar{K}^v(g) \cup K^v$ and $\beta^v \in \bar{L}^v(g) \cup L^v$ we have the equality

$$(\alpha A^v(\hat{x})\beta) \;=\; (\alpha A^v(x)\beta) - C\,(g^v - (\alpha G^v(g)\beta)) \tag{31}$$

by the above definition of $\hat{x}$. Thus, condition (A2) implies that the last term above can be made positive for $(\alpha, \beta) = (\bar{\alpha}, \ell)$ and negative for $(\alpha, \beta) = (k, \bar{\beta})$. Thus, since we have only finitely many such pairs, with a suitably large constant C, we can insure that inequalities (30) are satisfied.

Let us remark that (30) (together with (29)) means that the maximizer cannot locally gain by using a pure strategy $k \in K^v \setminus \bar{K}^v(g)$ with some positive weight $\alpha_k > 0$ in a mixed strategy α, and similarly the minimizer cannot locally gain form using strategies $\ell \in L^v \setminus \bar{L}^v(g)$. In other words, for all $\bar{\alpha}^v \in \bar{K}^v(g)$, $\bar{\beta}^v \in \bar{L}^v(g)$, $\alpha^v \in \Delta(K^v) \setminus \bar{K}^v(g)$ and $\beta^v \in \Delta(L^v) \setminus \bar{L}^v(g)$ we have

$$\left(\alpha^v A^v(\hat{x})\bar{\beta}^v\right) < g^v \quad \text{ and } \quad (\bar{\alpha}^v A^v(\hat{x})\beta^v) > g^v. \tag{32}$$

This implies that both conditions of (B4) hold, and hence by [I4], the potential transformation with $\hat{x}$ provides a canonical form for Γ. □

Corollary 4 *If a stochastic game satisfies conditions (B3) and (A2), then it has a canonical form (that is, (B3)∧(A2)⇒(B1)).*

Proof The proof of Theorem 3 actually shows that from the conditions of (B3) with the same arguments we can derive the validity of (32), and hence proving the existence of a canonical form. (It also immediately follows from Theorem 3 and the implication [I2].) □

Example 3.1 *Vrieze [Vri80, Chapter 8] showed an example, see Figure 4, for a stochastic game which has values and uniformly optimal stationary strategies, and which has no canonical form. We can see that condition (A2) is violated. In this game we have* $V = \{1, 2, 3\}$*, states 2 and 3 are absorbing with* $|K^2| = |K^3| = |L^2| = |L^3| = 1$*, while in state 1 we have* $|K^1| = |L^1| = 3$*. The reward matrix of state one is shown in Figure 4 together with the transition probabilities which are all zero or one.*

This game has values, $g = (0, -1, 1)$*, and unique uniformly optimal stationary strategies, namely* $\alpha^1 = (\frac{1}{2}, \frac{1}{2}, 0)$ *and* $\beta^1 = (\frac{1}{2}, \frac{1}{2}, 0)$*, and the trivial strategies in states 2 and 3. We have*

$$G^1 \;=\; \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$$

For a potential vector $x \in \mathbb{R}^V$ *we can assume w.l.o.g. that* $x_1 = 0$*, and thus we have*

$$A^1(x) \;=\; \begin{pmatrix} 1 & -1 & -1 - x_3 \\ -1 & 1 & -1 - x_2 \\ 1 - x_3 & 1 - x_2 & 0 \end{pmatrix}.$$

Here $\bar{K}^1 = \{(\frac{1}{2}, \frac{1}{2}, 0), (1, 0, 0)\}$*,* $\bar{L}^1 = \{(\frac{1}{2}, \frac{1}{2}, 0), (0, 1, 0)\}$*, and only the first vectors are optimal in the matrix game with payoffs* $A^1(x)$ *(for any potential transformation), and thus* α^1 *and* β^1 *given above are the unique optimal strategies. For the canonical form for some potential vector* $x \in \mathbb{R}^V$ *(*$x_1 = 0$*) we would need the inequalities that* $\alpha^1 A^1(x) \geq 0$ *and* $A^1(x)\beta^1 \leq 0$*, implying that* $-1 - \frac{x_2+x_3}{2} \geq 0$ *and* $1 - \frac{x_2+x_3}{2} \leq 0$*, leading to a contradiction. Consequently, this example does not have a canonical form.*

4 Stochastic Games with Additive Transitions

Recall that is this case, each player controls a part of the transition probabilities. More precisely, let Γ be a stochastic game, and assume that for each possible transition $(v, u) \in V$ there are probability distributions $\psi^{vu} \in [0, 1]^{K^v}$ and $\gamma^{vu} \in [0, 1]^{L^v}$ such that

$$p^{vu}_{k\ell} \;=\; \lambda^v \psi^{vu}_k + (1 - \lambda^v)\gamma^{vu}_\ell \tag{33}$$

holds for all strategies $k \in K^v$ and $\ell \in L^v$ with some constants $0 \leq \lambda^v \leq 1$, for all states $v \in V$.

Recently Flesch, Thuijsman and Vrieze [FTV07] showed that AT games satisfy condition (A1). We are going to show here that they also satisfy condition (A2), and hence admit a canonical form.

Lemma 4 *AT games satisfy condition (A2).*

Proof Let us consider an AT game Γ with transition probabilities as in (33), and denote by $g = g(\Gamma)$ its value vector. Let us fix a state $v \in V$ and define vectors $d \in \mathbb{R}^{K^v}$ and $f \in \mathbb{R}^{L^v}$ by

$$d_k \;=\; \sum_{u \in V} \psi^{vu}_k g^u \qquad \text{and} \qquad f_\ell \;=\; \sum_{u \in V} \gamma^{vu}_\ell g^u$$

for all $k \in K^v$ and $\ell \in L^v$. Let us then observe that for any strategies $\alpha^v \in \Delta(K^v)$ and $\beta^v \in \Delta(L^v)$ we have the equation

$$\alpha^v G^v \beta^v \;=\; \lambda^v \alpha^v d \;+\; (1 - \lambda^v) f \beta^v.$$

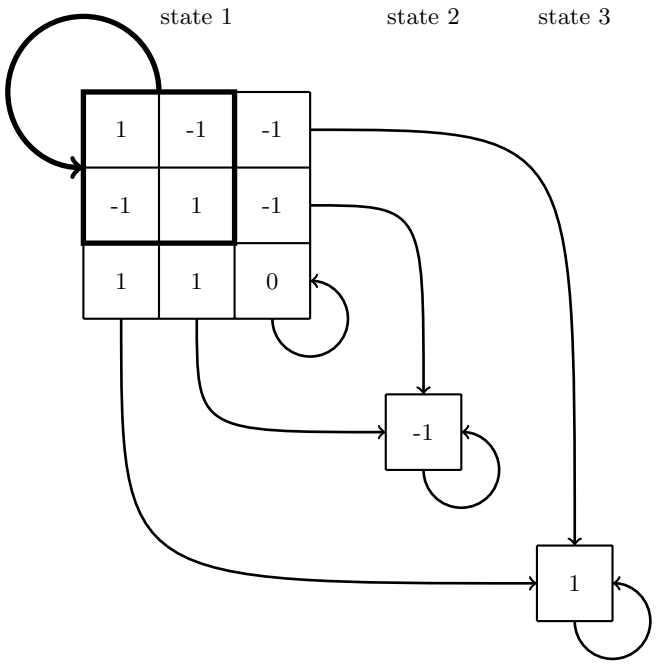


Figure 4: In this game Γ we have $|V| = 3$ states, $|K^1| = |L^1| = 3$, $|K^2| = |L^2| = 1$, and $|K^3| = |L^3| = 1$. The numbers in the state matrices are the local rewards. All transition probabilities are zero or one, and arcs in the picture indicate the probability 1 transitions. The thick arc in the picture indicates that for pairs of strategies from the top left 2×2 area in state 1 the game remains in state 1 with probability 1. This game has values and uniformly optimal stationary strategies, but it does not have a canonical form.

Let $d_{max} = \max_{k \in K^v} d_k$ and $f_{min} = \min_{\ell \in L^v} f_\ell$, set $I^v = \{k \in K^v \mid d_{max} = d_k\}$ and $J^v = \{\ell \in L^v \mid f_{min} = f_\ell\}$ Then we have $\bar{K}^v = \Delta(I^v)$ and $\bar{L}^v = \Delta(J^v)$, for all states $v \in V$. Furthermore, if we choose $\epsilon > 0$ such that it satisfies the inequalities

$$\epsilon \leq d_{max} - d_k \quad \text{and} \quad \epsilon \leq f_\ell - f_{min}$$

for all indices $k \in K^v \setminus I^v$ and $\ell \in L^v \setminus J^v$ and for all states $v \in V$, then (A2) follows. □

Corollary 5 *AT games admit a canonical form.*

Proof Immediate by Theorems 3 and 4. □

Let us also note that computing the values and the canonical form of an AT-game looks even more difficult than for the other known classes when canonical form exists. The main reason is that AT-games may have irrational values, irrational potentials and irrational coefficients in the uniformly optimal strategies.

Example 4.1 *Rhagavan, Tijs and Vrieze [RTV85] showed an example, see Figure 5, for an AT game in which the optimal values and strategies are irrational. This example is ergodic, with states* $V = \{1, 2\}$ *and values* $g^1 = g^2 = -(6 - \sqrt{30})^2$. *The vector* $x = (0, 22 - 4\sqrt{30})$ *is a potential transformation providing the canonical form for this example. We have* $K^1 = K^2 = L^1 = L^2 = \{1, 2\}$, *and the strategies* $\alpha^1 = \beta^1 = (\frac{-4+\sqrt{30}}{2}, \frac{6-\sqrt{30}}{2})$, *and* $\alpha^2 = \beta^2 = (\frac{-9+2\sqrt{30}}{3}, \frac{12-2\sqrt{30}}{3})$ *are the uniformly optimal stationary strategies.*

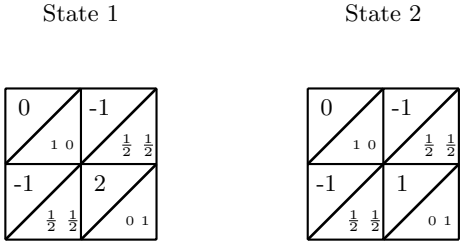


Figure 5: This example has two states with $|K^1| = |L^1| = |K^2| = |L^2| = 2$. In each cell in the figure we have the reward in the top left corner, while the transition probabilities to states 1 and 2 (in this order) are in bottom right area. This is an ergodic AT-game with irrational values and optimal strategies.

5 Canonical Form for Discounted Games: An Algorithmic Proof

Let us note that the existence of canonical form for the discounted case follows immediately from the fact that the vector of values g_β satisfies the Shapley equations (5). Indeed, let us use the potentials $x = \frac{-g_\delta}{1-\delta}$ in (40). Then it follows from (3), (15) and (24) that

$$\begin{aligned} \text{Val}\,[A^v(x)] &= \text{Val}\,[A^v - \frac{g^v_\delta}{1-\delta}J + \frac{\delta}{1-\delta}G^v(g_\delta)] \\ &= \frac{1}{1-\delta}\ \text{Val}\,[(1-\delta)A^v + \delta G^v(g_\delta)] - \frac{g^v_\delta}{1-\delta} = 0. \end{aligned}$$

Thus, after the transformation, all the local values are equal to $c = 0$, and a locally optimal strategy at each position achieves this local value. Proposition 1 implies that the transformed game is in canonical form.

In the rest of this section, we give a pseudo-polynomial-time algorithm, that brings any δ-discounted game which satisfies the ordered field property into canonical form by repeatedly applying potential transformations. More precisely, the running time of the algorithm is polynomial in the total number of bits needed to represent the rewards and the transition probabilities, $\frac{1}{1-\delta}$, and $\log \frac{1}{\epsilon}$, where ϵ is the necessary accuracy at which we can guarantee that function $m_x(v) := \text{Val}\,[A^v(x)]$ is constant for all v.

Given a δ-discounted stochastic game $\Gamma = (p^{vu}_{k\ell},\ r^{vu}_{k\ell}\ \mid\ k \in K^v,\ \ell \in L^v,\ u, v \in V)$, let $[r] = [r^- : r^+]$ denote the range of the local rewards, that is, $r^+ = \max\{r^{vu}_{k\ell} \mid v, u \in V,\ k \in K^v,\ \ell \in L^v\}$ and $r^- = \min\{r^{vu}_{k\ell} \mid v, u \in V,\ k \in K^v,\ \ell \in L^v\}$, Similarly, let $[m_x] = [m^-_x : m^+_x]$ denote the range of the function m. We will find a potential x such that function $m_x : V \to \mathbb{R}$ is constant, that is, $m^-_x = m^+_x$.

The following simple procedure reduces $|[m_x]| \stackrel{\text{def}}{=} m^+_x - m^-_x$ to within an arbitrary accuracy ε.

Algorithm 1 Pumping algorithm

Input: a δ-discounted game Γ, an accuracy ϵ, and two parameters $a, b \in [0, 1]$.
Output: a potential $x : V \to \mathbb{R}$ s.t. $|m_x(v) - m_x(u)| \le \varepsilon$ for all $u, v \in V$

```
initialize x = 0
while |[m_x]| ≥ ε do
  for all v s.t. x^v ≥ m_x^- + a|[m_x]| do
    x^v := x^v − b|[m_x]|
  end for
end while
return x
```

Lemma 5 *When run with $a = b = \frac{1}{2}$, the above procedure terminates in $N = \frac{\log|[r]| - \log\varepsilon}{1-\log(1+\beta)}$ iterations.*

Proof We show that the range of m decreases in each iteration by a factor

$$c = \frac{1+\delta}{2} = 1 - \frac{1-\delta}{2}. \tag{34}$$

In fact, a, b will be chosen such that this is the maximum possible decrease in one iteration. Without loss of generality, we can assume that $[m_0] = [0, 1]$ is the unit interval. Indeed, if $[m_0]$ is just one point then the game is already in an ergodic canonical form and the problem is solved; otherwise, there is a unique (bijective) linear map of $[m_0]$ onto $[0, 1]$, that can be applied to the local rewards without changing the set of locally optimal strategies.

Given two parameters $a, b \in [0, 1]$, let us define potential $x = x_{a,b}$ as follows: $x^v = -b$ for a vertex $v \in V$ whenever $m_0(v) \geq a$ and $x^v = 0$ otherwise. We will show that the optimal choice $a = b = \frac{1}{2}$ results in $[m_x] = [0, c]$, where c is given by (34).

Indeed, it is easy to verify using (3) that

$$(1-\delta)b \leq m_0(v) - m_x(v) \leq b \text{ if } m_0(v) \geq a, \text{ while}$$

$$0 \leq m_x(v) - m_0(v) \leq \delta b \text{ if } m_0(v) < a.$$

Clearly, $a \geq b$ should hold, since otherwise $m_x(v)$ could become negative for a vertex v such that $m_0(v) = a$. On the other hand, to get the largest decrease in range, we have to minimize c subject to $m_x(v) \notin [c, 1]$ for all $v \in V$. Hence, $c \geq a + \delta b$ and $c \geq 1 - (1-\delta)b$. To optimize, we set $a + \delta b = 1 - (1-\delta)b$ which results in $a + b = 1$.

Finally, we have to minimize $c = a + \delta b$ subject to $0 \leq b \leq a$, $a + b = 1$, and $0 \leq \delta \leq 1$. Obviously, the optimal c is given by (34) when $a = b = \frac{1}{2}$.

Thus, in one iteration the range $[m_0]$ is reduced at least by the factor (34). Using $[m_0] \subseteq [r]$, we must have, after N iterations,

$$|[m_x]| \leq |[r]| \left(\frac{1+\beta}{2}\right)^N \leq \varepsilon, \tag{35}$$

by our choice of N. □

Clearly, the above procedure converges to a canonical form as $\epsilon \to 0$. Furthermore, for games with the ordered field property, such as BWR-games, an ARAT-games, and SC-games, we can use (11) to estimate the necessary accuracy at which we can guarantee that function m_x is constant. For instance given a BWR-game, let us assume that (i) $\beta = 1 - B'/B \in [0, 1)$ is a rational number; (ii) all local rewards, are integral in the range $[-R, R]$; (iii) probabilities $p^{vu}_{k\ell}$ are rational numbers with the least common denominator D. Then it is enough to take $\epsilon = (1/(DBB'))^{O(n^4)} \cdot (1/R)^{O(nh)}$ for the BWR-case, where $h = |E|$ is the number of edges of the graph G and $n = |V|$ is the number of states (see, e.g., [BEGM09]).

Let us remark, however, that the constant $1 - \log(1+\delta)$ in the running time tends to 0, as $\delta \to 1^-$. More precisely, if $y = 1 - \delta \to 0^+$ then $1 - \log(1+\delta) = 1 - \log(2-y) \sim y/(2\ln 2)$, and thus we obtain for the number of iterations $N \sim 2\ln 2 \frac{(\log R - \log\varepsilon)}{(1-\delta)}$. There are examples [Con92, BEGM10a, Mil11] which shows that $\delta > 1 - 2^{-n}$ might be needed for a "sufficiently good" approximation.

On the other hand, Andersson and Miltersen [AM09] showed that, for PI-games, it is enough to take $\delta = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$, so that the an optimal pair of strategies in the discounted game remains optimal the undiscounted one. Thus, for the undiscounted BWR-games the limit transition $\delta \to 1^-$ provides a finite but exponential in the worst case algorithm. Note, however, that this is *not* yet enough to prove the existence of canonical form for PI-games, since the canonical form obtained through discounting will contain a factor $\delta < 1$ in it. In the next section, we overcome this problem by taking δ to the limit.

6 BWR-games

Recall that a BWR-game is defined by the quadruple $\Gamma = (G, \{p^{vu}, r^{vu}\}_{v,u\in V})$, where $G = (V = V_W \cup V_B \cup V_R, E)$ is a digraph on n vertices that may have loops and multiple arcs, but no terminal vertices[2], i.e., vertices of out-degree 0; $\{p^{vu}\}$ is the set of probability distributions for all $v \in V_R$ specifying the probability p^{vu} of a move form v to u; and $(r^{vu})_{(v,u)\in E} \in \mathbb{R}^E$ is a local reward vector. As usual, we assume that $\sum_{u \mid (v,u)\in E} p^{vu} = 1 \; \forall v \in V_R$. For convenience we will assume that $p^{vu} > 0$ whenever $(v, u) \in E$ and $v \in V_R$, and set $p^{vu} = 0$ for $(v, u) \notin E$.

In this case, it will be enough to consider pure stationary strategies. In particular, we define a strategy $\alpha \in \mathcal{K}(\Gamma)$ (respectively, $\beta \in \mathcal{L}(\Gamma)$) as a mapping that assigns a move $(v, u) \in E$ to each position $v \in V_W$ (respectively, $v \in V_B$). A pair of strategies $s = (\alpha, \beta)$ is called a *situation*.

In Section 6.2 we will prove our main result for the undiscounted case.

Theorem 4 *Any BWR-game can be brought by a potential transformation to canonical form.*

Overview of the technique. Our proof of Theorem 4 proceeds in the spirit of [Gil57, LL69] (see also [MO70], Chapter 4): First, we consider the discounted BWR-game and consider the function $m : V \to \mathbb{R}$, where $m(v) :=$ Val (A^v) is the maximum (minimum, or average) local reward for $v \in V_W$ (resp., $v \in V_B$, or $v \in V_R$). Starting form 0 potentials, and changing the potentials by a very simple iterative procedure, we can show that the function m can be made constant, in a number of iterations proportional to $(1 - \delta)^{-1}$. (However, such approach requires exponential time in general, since one must choose $\delta > 1 - \varepsilon/2^n$ to approximate the value of an undiscounted BWR-game within accuracy ε; see, e.g., [Con92, BEGM09].) We then reduce the undiscounted BWR-games to canonical form by just taking the limit $\delta \to 1^-$. However, in doing so we face one difficulty: some of the potentials tend to ∞ as $\delta \to 1^-$. We overcome this by modifying the potentials somehow, and then using a convergence result of Blackwell [Bla62] to finish the proof.

6.1 Canonical Form for BWR-games

For BWR-games, the canonical form admits a simpler interpretation. Specifically, let us use the following notation throughout this section: Given a vector $f \in \mathbb{R}^{V\times V}$ and subset of edges $E' \subseteq E$, we write $\bar{M}_{E'}[f^{vu}]$ to *symbolically* mean

$$\bar{M}_{E'}[f^{vu}] \equiv \begin{cases} \max_{u|(v,u)\in E'} f^{vu}, & \text{for } v \in V_W, \\ \min_{u|(v,u)\in E'} f^{vu}, & \text{for } v \in V_B, \\ \sum_{u|(v,u)\in E'} p^{vu} f^{vu}, & \text{for } v \in V_R. \end{cases}$$

When $E' = E$, we will write $\bar{M}_{E'}[f^{vu}]$ as $\bar{M}[f^{vu}]$.

Note that for BWR-games, condition (A2) is trivially satisfied. Thus by Corollary 4, it is enough for proving the existence of canonical to show that (B3) holds. With the the above notation, condition (B3) can now be re-written as follows in the BWR-case:

(B3′) There exist $g \in \mathbb{R}^V$ and a potential vector $x \in \mathbb{R}^V$ such that (i) for all $v \in V$, $g^v = \bar{M}[g^u] = \bar{M}[r^{vu}(x)]$ and, moreover, (ii) for every $v \in V_W \cup V_B$ there is a move $(v, u) \in E$ such that $g^v = g^u = r^{vu}(x)$, or in other words, move (v, u) is locally optimal and it respects the value of vector g.

6.2 Proof of Theorem 4

In deriving Theorem 4 from Theorem 2, we face one difficulty: some of the potentials tend to ∞ as $\beta \to 1^-$. We overcome this by modifying the potentials somehow, and then using a convergence result of Blackwell [Bla62]. By implication [I5] of Theorem 1, it is enough to show that (B5) is satisfied, that is,

[2]This assumption is without loss of generality since one can add a loop to each terminal vertex

(B5′) there exist $g \in \mathbb{R}^V$ and a potential vector $x \in \mathbb{R}^V$ such that (i) for all $v \in V$, $g^v = \bar{M}[g^u]$, and $g^v = \bar{M}_{\mathrm{Ext}(g)}[r^{vu}(x)]$, and (iii) for every $v \in V_W \cup V_B$, there is a move $(v,u) \in E$ such that $g^v = g^u = r^{vu}(x)$,

where

$$\mathrm{Ext}(g) = \{(v,u) \in E \mid g^u = g^v,\ v \in V_W \cup V_B\} \cup \{(v,u) \in E \mid v \in V_R\},$$

is the subset of extremal edges with respect to g.

From Theorem 2, we know that, for any $0 \leq \delta < 1$, there exist $c_\delta \in \mathbb{R}$ and $x = x_\delta \in \mathbb{R}^V$ such that

$$c_\delta = \bar{M}[r^{vu} + x^v_\delta - \delta x^u_\delta] \quad \text{for all } v \in V. \tag{36}$$

Furthermore, from (17), we know that the value of the game when started at vertex $v \in V$ is

$$g^v_\delta = c_\delta - (1-\delta)x^v_\delta = \bar{M}[r^{vu} + \delta(x^v_\delta - x^u_\delta)]. \tag{37}$$

Note that the values g^v_δ satisfy the Shapley equations (5):

$$g^v_\delta = \bar{M}[(1-\delta)r^{vu} + \delta g^u_\beta], \tag{38}$$

for all $v \in V$. (Indeed, using (37), we have

$$\begin{aligned}
\bar{M}[(1-\delta)r^{vu} + \delta g^u_\delta] &= \bar{M}[(1-\delta)r^{vu} + \delta(c_\delta - (1-\delta)x^u_\delta)] \\
&= (1-\delta)\bar{M}[r^{vu} - \delta x^u_\delta] + \delta c_\delta \\
&= (1-\delta)\bar{M}[r^{vu} + \delta(x^v_\delta - x^u_\delta)] + \delta(c_\delta - (1-\delta)x^v_\delta) \\
&= (1-\delta)g^v_\delta + \delta g^v_\delta = g^v_\delta.)
\end{aligned}$$

By Theorem 2, for each $\delta \in [0,1)$, there exists an optimal situation s_δ in the δ-discounted BWR-game, and potential x_δ satisfying (36) and (37). Let us consider all such situations as $\delta \to 1^-$. Among this infinite sequence of situations, one situation s appears infinitely many times, since the total number of possible strategies is finite. Let us fix such a situation s and consider the corresponding infinite subsequence $\{\delta_i\}_{i=0}^\infty$ for which s is optimal in the corresponding game. Then $\lim_{i\to\infty} \delta_i = 1$ and (36), (37) and (38) hold for every $\delta \in \{\delta_i\}_{i=0}^\infty$. For $v \in V$, let $g^v = \lim_{i\to\infty} g^{\delta_i}(v)$ and note that this limit exists (by (12)). Furthermore, since (38) is satisfied for all δ_i, it is also satisfied in the limit as $i \to \infty$, i.e.,

$$g^v = \bar{M}[g^u]. \tag{39}$$

Note that, in the *non-ergodic* case, as $\delta \to 1^-$, (37) implies that $|x^v_\delta| \to \infty$, for some vertices $v \in V$; otherwise all the values g^v are equal to $\lim_{\delta\to 1} c_\delta$, independent of the starting position. We will modify the potentials, in this case, to guarantee that they become finite, without affecting the value of the game.

Consider any $\delta \in \{\delta_i\}_{i=0}^\infty$. From (37), we can express the potential at $v \in V$ as follows

$$x^v_\delta = \frac{c_\delta - g^v_\delta}{1-\delta}. \tag{40}$$

Define, for $v \in V$, the new potential:

$$y^v_\delta = x^v_\delta - \frac{c_\delta - g^v}{1-\delta} = \frac{g^v - g^v_\delta}{1-\delta}. \tag{41}$$

In particular, substituting $y^v_\delta - y^u_\delta = x^v_\delta - x^u_\delta + \frac{g^v - g^u}{1-\delta}$, we have by (37) and (39),

$$g^v_\delta = \bar{M}_{\mathrm{Ext}(g)}[r^{vu} + \delta(y^v_\delta - y^u_\delta)]. \tag{42}$$

Let $P(\alpha,\beta)$ be the transition matrix obtained by extending P by setting the entries corresponding to $s = (\alpha,\beta)$ to 1, $Q(\alpha,\beta)$ and $a(\alpha,\beta)$ be the corresponding limiting transition matrix and expected local reward vector, respectively. Recall that $g_\delta = (1-\delta)\sum_{i=0}^\infty \delta^i P(\alpha,\beta)^i a(\alpha,\beta)$, $g = \lim_{\delta\to 1} g_\delta = Q(\alpha,\beta)a(\alpha,\beta)$; see Equations (16) and (17).

Rewriting $Q(\alpha,\beta)a(\alpha,\beta) = (1-\delta)\sum_{i=0}^{\infty}\delta^i Q(\alpha,\beta)a(\alpha,\beta)$, for any $\delta \in (0,1)$, we obtain

$$\begin{aligned} y &= \lim_{i\to\infty} y^{\delta_i} = \lim_{\delta\to 1} y^{\delta} = \lim_{\delta\to 1}(1-\delta)^{-1}(g - g_\delta) \\ &= \lim_{\delta\to 1}\sum_{i=0}^{\infty}\delta^i(Q(\alpha,\beta) - P(\alpha,\beta)^i)a(\alpha,\beta) = -[(I-(P(\alpha,\beta)-Qa(\alpha,\beta)))^{-1} - Q(\alpha,\beta)]a(\alpha,\beta), \end{aligned}$$

where the last equality follows by (63) in Appendix A. So y exists in the limit and it satisfies (42) with $\delta = 1$. In other words, $y, g \in \mathbb{R}^V$ satisfy the conditions stated in (B5′).

7 Canonical forms for SC- and ARAT-games

Let Γ be an additive stochastic games (that is, $P^{vu} = \lambda^v \psi^{vu} e^T + (1-\lambda^v)e\gamma^{vu}$, for some $\lambda^v \in [0,1]$, $\psi^{vu} \in [0,1]^{K^v}$, $\gamma^{vu} \in [0,1]^{L^v}$, where e is the vector of all ones). It will be convenient to consider the following generalization of SC and ARAT-games [KaR10, Sin89]:

(A3) For each $v \in V$, if $\lambda^v \in (0,1)$ then $A^v = q^v e^T + e s^v$, for some $q^v \in \mathbb{R}^{K^v}$ and $s^v \in \mathbb{R}^{L^v}$.

We note that any additive game satisfying (A3) enjoys the following further property:

(A4) For every $v \in V$, if $\lambda^v > 0$ (resp., if $\lambda^v < 1$), then there exists a finite set $\Delta_W^v \subseteq \Delta(K^v)$ (resp., $\Delta_B^v \subseteq \Delta(L^v)$) such that, for all $x \in \mathbb{R}^V$, $\Omega(A^v(x)) \subseteq \Delta_W^v$ (resp., $\Lambda(A^v(x)) \subseteq \Delta_B^v$.

Indeed, let Γ be a game satisfying (A3). Then, for $\lambda^v = 1$, the claim follows from the following lemma, which can be derived from results of Shapley and Snow [SS50], and Parthasarathy and Raghavan [PR81].

Lemma 6 ([PR81]) *Let $A^v > 0$ for all v. If $\alpha^v \in \Omega(A^v(x))$ (resp., $\beta^v \in \Lambda(A^v(x))$) for some $x \in \mathbb{R}^V$, then there exists a non-singular submatrix $\bar{A}^v$ of A^v such that $\alpha^v = \frac{e^T(\bar{A}^v)^{-1}}{e^T(\bar{A}^v)^{-1}e}$ (resp., $\alpha^v = \frac{(\bar{A}^v)^{-1}e}{e^T(\bar{A}^v)^{-1}e}$).*

Clearly, we may assume without loss of generality (by adding a sufficiently large finite constant $C \geq 0$ to every entry $r^{vu}_{k\ell}$) that $A^v > 0$, for all v. Thus Lemma 6, and its symmetric version for $\lambda^v = 0$, imply that if we take

$$\begin{aligned} \Delta_W^v &= \left\{ \frac{e^T(\bar{A}^v)^{-1}}{e^T(\bar{A}^v)^{-1}e} \;:\; \bar{A}^v \text{is a non-singular submatrix of } A \right\} \cap \Delta^v(K^v) \;\text{ if } \lambda^v > 0, \text{ and} \\ \Delta_B^v &= \left\{ \frac{(\bar{A}^v)^{-1}e}{e^T(\bar{A}^v)^{-1}e} \;:\; \bar{A}^v \text{is a non-singular submatrix of } A \right\} \cap \Delta^v(L^v), \;\text{ if } \lambda^v < 1, \end{aligned}$$

we would satisfy (A4) for any v such that $\lambda^v \in \{0,1\}$.

Let us now consider a state v such that $\lambda^v \in (0,1)$. Then, for any $x \in \mathbb{R}^V$, the extremum locally optimal strategies for $A^v(x)$ are pure [RTV85], and hence it is enough to take $\Delta_W^v = K^v$ and $\Delta_B^v = L^v$. (To see this, consider the LP formulation for solving the local matrix game $A^v(x)$: $\max\{g^v :\; \alpha^v A^v(x) \geq g^v e^T,\; \alpha^v e = 1,\; \alpha^v \geq 0\}$. It is easy to see that this equivalent to finding

$$\max_{\alpha^v:\; \alpha^v e = 1, \alpha^v \geq 0} \alpha^v (q^v - \lambda^v \sum_{u\in V} x^u \psi^{vu}) + \min_{\ell \in L^v}\left(s^v_\ell - (1-\lambda^v)\sum_{u\in V} x^u \gamma^{vu}_\ell \right) + x^v,$$

which is attainted at a pure strategy $\alpha^v \in K^v$. A symmetric argument shows that the optimal is also attained at a pure strategy $\beta^v \in L^v$.)

Theorem 5 *If an additive stochastic game Γ satisfies condition (A3), then it is equivalent to a BWR-game on $\sum_{v\in V}(3 + |\Delta_W^v| + |\Delta_B^v|)$ states, where Δ_W^v and Δ_B^v are the finite sets strategies guaranteed by (A4).*

Proof Let $\Gamma = (p^{k\ell}_{vu}, r^{k\ell}_{vu} |\ k \in K_v,\ \ell \in L_v, v, u \in V)$ be a game satisfying (A3). We assume without loss of generality that, for all $v \in V$ $\Delta^v_W \supseteq K^v$ and $\Delta^v_B \supseteq L^v$ (otherwise we can extend the sets Δ^v_W and Δ^v_B with all pure strategies without obviously violating condition (A4)). We construct a BWR-game $\tilde{\Gamma} = (G = (\tilde{V} = \tilde{V}_W \cup \tilde{V}_B \cup \tilde{V}_R, E), \{\tilde{p}_{vu}, \tilde{r}_{vu}\}_{v,u\in\tilde{V}})$ as follows. For every state $v \in V$, we have a random node $v \in \tilde{V}_R$. If $\lambda^v > 0$, then we have a white node $v_W \in \tilde{V}_W$, and a set of $|\Delta^v_W|$ random nodes $\{(\alpha^v) :\ \alpha^v \in \Delta^v_W\} \subseteq \tilde{V}_R$. Similarly if $\lambda^v < 1$, we have a black node $v_B \in \tilde{V}_B$, and a set of $|\Delta^v_B|$ random nodes $\{(\beta^v) :\ \beta^v \in \Delta^v_B\} \subseteq \tilde{V}_R$. The arcs are defined as follows. We have arcs (v, v_W) and (v, v_B) with local rewards 0 and transition probabilities λ^v and $1-\lambda^v$, respectively. For $\alpha^v \in \Delta^v_W$ and $u \in V$, we have arcs $((\alpha^v), u)$ of local reward 0, $(v_W, (\alpha^v))$ of local reward $m(\alpha^v) := \min_{\beta^v\in\Delta(L^v)} \alpha^v A^v \beta^v$ if $\lambda^v = 1$, and of local reward $m(\alpha^v) := \frac{\alpha^v q^v}{\lambda^v}$, if $\lambda^v \in (0,1)$. At the random node (α^v), the probability on arc $((\alpha^v), u)$ is set to $\alpha^v \psi^{vu}$. Similarly, for $\beta^v \in \Delta^v_B$ and $u \in V$, we have arcs $((\beta^v), u)$ of local reward 0, $(v_B, (\beta^v))$ of local reward $m(\beta^v) := \max_{\alpha^v\in\Delta(K^v)} \alpha^v A^v \beta^v$ if $\lambda^v = 0$, and of local reward $m(\beta^v) := \frac{s^v \beta^v}{1-\lambda^v}$ if $\lambda^v \in (0,1)$. At the random node (β^v), the probability on arc $((\beta^v), u)$ is set to $\gamma^{vu}\beta^v$.

By the equivalence [I4] of Theorem 1, It is enough to show that Γ satisfies (B4). By the same equivalence and Theorem 4, applied to $\tilde{\Gamma}$, there exist vectors $\tilde{g}, \tilde{x} \in \mathbb{R}^{\tilde{V}}$ such that for every $v \in V$ the following holds:

- $$\begin{aligned} \tilde{g}^v &= \lambda^v \tilde{g}^{v_W} + (1-\lambda^v)\tilde{g}^{v_B} && (43)\\ \tilde{g}^v &= \tilde{x}^v - \lambda^v \tilde{x}^{v_W} - (1-\lambda^v)\tilde{x}^{v_B}; && (44) \end{aligned}$$

- there exists $\bar{\alpha}^v \in \Delta^v_W$, such that

$$\begin{aligned} \tilde{g}^{v_W} &= \tilde{g}^{(\bar{\alpha}^v)} = \max_{\alpha^v\in\Delta^v_W} \tilde{g}^{(\alpha^v)}, && (45)\\ \tilde{g}^{v_W} &= m(\bar{\alpha}^v) + \tilde{x}^{v_W} - \tilde{x}^{(\bar{\alpha}^v)} = \max_{\alpha^v\in\Delta^v_W} (m(\alpha^v) + \tilde{x}^{v_W} - \tilde{x}^{(\alpha^v)}), && (46)\\ \tilde{g}^{v_W} &> m(\alpha^v) + \tilde{x}^{vv_W} - \tilde{x}^{(\alpha^v)} \quad \text{for all } \alpha^v \in \Delta^v_W \text{ such that } \tilde{g}^{(\alpha^v)} < \tilde{g}^{(\bar{\alpha}^v)}; && (47) \end{aligned}$$

- for every $\alpha^v \in \Delta^v_W$

$$\begin{aligned} \tilde{g}^{(\alpha^v)} &= \sum_{u\in V} \alpha^v \psi^{vu} \tilde{g}^u && (48)\\ &= \tilde{x}^{(\alpha^v)} - \sum_{u\in V} \alpha^v \psi^{vu} \tilde{x}^u; && (49) \end{aligned}$$

- there exists $\bar{\beta}^v \in \Delta^v_B$, such that

$$\begin{aligned} \tilde{g}^{v_B} &= \tilde{g}^{(\bar{\beta}^v)} = \min_{\beta^v\in\Delta^v_B} \tilde{g}^{(\beta^v)}, && (50)\\ \tilde{g}^{v_B} &= m(\bar{\beta}^v) + \tilde{x}^{v_B} - \tilde{x}^{(\bar{\beta}^v)} = \min_{\beta^v\in\Delta^v_B} (m(\beta^v) + \tilde{x}^{v_B} - \tilde{x}^{(\beta^v)}), && (51)\\ \tilde{g}^{v_B} &< m(\beta^v) + \tilde{x}^{v_B} - \tilde{x}^{(\beta^v)} \quad \text{for all } \beta^v \in \Delta^v_B \text{ such that } \tilde{g}^{(\beta^v)} > \tilde{g}^{(\bar{\beta}^v)}; && (52) \end{aligned}$$

- for every $\beta^v \in \Delta^v_B$

$$\begin{aligned} \tilde{g}^{(\beta^v)} &= \sum_{u\in V} \gamma^{vu} \beta^v \tilde{g}^u && (53)\\ &= \tilde{x}^{(\beta^v)} - \sum_{u\in V} \gamma^{vu} \beta^v \tilde{x}^u. && (54) \end{aligned}$$

We claim that the vectors x, g defined by $g^v = 3\tilde{g}^v$ and $x^v = \tilde{x}^v$, for all $v \in V$, satisfy (B4) in Γ. Indeed, for any $v \in V$, (43), (45), (48), (50) and (53) imply that

$$g^v = \lambda^v \max_{\alpha^v \in \Delta_W^v} \sum_{u \in V} \alpha^v \psi^{vu} g^u + (1-\lambda^v) \min_{\beta^v \in \Delta_B^v} \sum_{u \in V} \gamma^{vu} \beta^v g^u = \underset{K^v \times L^v}{\mathrm{Val}}(G^v(g)), \tag{55}$$

where the last equality follows from the assumption that $\Delta_W^v \supseteq K^v$ and $\Delta_B^v \supseteq L^v$.

On the other hand, (46), (49), (51) and (54) imply that

$$2\tilde{g}^{v_W} = \max_{\alpha^v \in \Delta_W^v} \left(m(\alpha^v) + \tilde{x}^{v_W} - \sum_{u \in V} \alpha^v \psi^{vu} x^u \right), \tag{56}$$

$$2\tilde{g}^{v_B} = \min_{\beta^v \in \Delta_B^v} \left(m(\beta^v) + \tilde{x}^{v_B} - \sum_{u \in V} \gamma^{vu} \beta^v x^u \right), \tag{57}$$

with the further implications, following by (47) and (52), that

$$\text{if } \alpha^v \text{ is a maximizer in (56) then } \alpha^v \text{ is a maximizer in the first term in (55),} \tag{58}$$

$$\text{if } \beta^v \text{ is a minimizer in (57) then } \beta^v \text{ is a minimizer in the second term in (55).} \tag{59}$$

Multiplying (58) and (59) respectively by λ^v and $1-\lambda^v$, summing and using (43) and (44), we obtain

$$g^v = \lambda^v \max_{\alpha^v \in \Delta_W^v} \left(m(\alpha^v) + x^v - \sum_{u \in V} \alpha^v \psi^{vu} x^u \right) + (1-\lambda^v) \min_{\beta^v \in \Delta_B^v} \left(m(\beta^v) + x^v - \sum_{u \in V} \gamma^{vu} \beta^v x^u \right) \tag{60}$$

If $\lambda^v = 1$, then (60) gives

$$g^v = \max_{\alpha^v \in \Delta_W^v} \min_{\beta^v \in \Delta(L^v)} (\alpha^v A^v \beta^v + x^v - \sum_{u \in V} \alpha^v \psi^{vu} x^u) = \underset{K^v \times L^v}{\mathrm{Val}} A^v(x), \tag{61}$$

where the last equation follows from (A4). Furthermore, (58) implies that, if (α^v, β^v) are locally optimal for $A^v(x)$ then they are also optimal for $G^v(g)$. A similar conclusion can be made when $\lambda^v = 0$.

Suppose now that $\lambda^v \in (0,1)$. Then (60) gives

$$\begin{aligned} g^v &= \max_{\alpha^v \in \Delta_W^v} \left(\alpha^v q^v - \lambda^v \sum_{u \in V} \alpha^v \psi^{vu} x^u \right) + \min_{\beta^v \in \Delta_B^v} \left(s^v \beta^v - \sum_{u \in V} \gamma^{vu} \beta^v x^u \right) + x^v \\ &= \max_{\alpha^v \in \Delta_W^v} \min_{\beta^v \in \Delta_B^v} (\alpha^v A^v \beta^v + x^v - \sum_{u \in V} \alpha^v \psi^{vu} x^u) = \underset{K^v \times L^v}{\mathrm{Val}} A^v(x), \end{aligned} \tag{62}$$

where the last equation follows from (A4). Furthermore, (58) and (59) imply that, if (α^v, β^v) are locally optimal for $A^v(x)$ then they are also optimal for $G^v(g)$. This together with (61) (and the similar equation for $\lambda^v = 0$) and (50) imply (B4). □

Remark 8 *From an algorithmic point of view, we note that the above reduction (in Theorem 5), in general, can yield only the global value vector g and the potential vector x that brings the game to canonical form. This is because the constructed BWR-game, in its canonical form, might have exponentially many locally optimal strategies, and an arbitrary one of them might only be a* max-min *or a* min-max *strategy. However, having the vectors g and x we can find the optimal strategies by solving the local matrix games.*

Corollary 6 *Every PI-game, SC-game, or ARAT game admits a canonical form.*

Corollary 7 *BWR-games, ARAT-games and PI-games are polynomial-time equivalent.*

Proof Obviously, a BWR-game $\Gamma = (G = (V = V_W \cup V_B \cup V_R, E), \{p^{vu}, r^{vu}\}_{v,u \in V})$ can be written as a PI-game $\tilde{\Gamma} = (\tilde{V} = \tilde{V}_W \cup \tilde{V}_B, \{\tilde{p}^{vu}, \tilde{r}^{vu}\}_{v,u \in \tilde{V}})$., where:

- $\tilde{V}_W = V_W \cup V_R$, $\tilde{V}_B = V_B$;
- for all $v \in V_W$, $K^v = \{u \in V : \ (v,u) \in E\}$, $\tilde{r}^{vu}_{u1} = r^{vu}$, $\tilde{p}^{vu}_{u1} = 1$, and $\tilde{r}^{vu}_{u'1} = \tilde{p}^{vu}_{u'1} = 0$, for all $u' \neq u$;
- for all $v \in V_B$, $L^v = \{u \in V : \ (v,u) \in E\}$, $\tilde{r}^{vu}_{1u} = r^{vu}$, $\tilde{p}^{vu}_{1u} = 1$, and $\tilde{r}^{vu}_{1u'} = \tilde{p}^{vu}_{1u'} = 0$, for all $u' \neq u$;
- for all $v \in V_R$, $|L^v| = |K^v| = 1$, and $\tilde{r}^{vu}_{11} = r^{vu}$, and $\tilde{p}^{vu}_{11} = p^{vu}$.

Obviously also every PI-game is an ARAT-game.

Conversely, Theorem 5 and the fact that an ARAT-game satisfies (A4) with $|\Delta^v_W| = |K^v|$ and $|\Delta^v_B| = |L^v|$ implies that any ARAT-game can be reduced to a polynomial-size BWR-game. □

7.1 BWR-, PI-, and ARAT-games are solvable in subexponential time

Zwick and Paterson [ZP96] observed that undiscounted BW-games are polynomial-time reducible to the discounted ones. In fact, it is enough to choose any $\delta > 1 - 1/(4n^3R)$, when rewards are integral with maximum absolute value R; see [ZP96], Theorem 5.2. Furthermore, they showed that the discounted BW-games are polynomial-time reducible to SS-Games; see [ZP96] Theorem 6.1. Andersson and Miltersen [AM09] has recently modified this reduction in [ZP96] to show that any discounted PI-game can be represented as an SS-game, where the probabilities are bilinear in δ, the original transition probabilities, and original rewards. Furthermore, he also showed that any undiscounted PI-game is reduced to a discounted one with $\delta = 1 - ((n!)^2 2^{2n+3} M^{2n^2})^{-1}$, when the rewards and transition probabilities are assumed to be rational with integral numerators and denominators of maximum absolute value M.

Halman [Hal07] showed that any SS-game with $m = |V_B| + |V_W|$ deterministic nodes can be solved in randomized *strongly* subexponential-time $2^{O(\sqrt{m \log m})} \operatorname{poly}(|V_R|)$. We observe further that the reduction in [AM09] can only increase the number of random nodes. Thus we obtain the following result.

Corollary 8 *Any BWR-game, PI-game, or ARAT-game on n states is solvable in strongly $2^{O(\sqrt{n \log n})}$ $\operatorname{poly}(n)$ expected time.*

References

[AM09] D. Andersson and P. B. Miltersen. The complexity of solving stochastic games on graphs. In *Proc. 20th ISAAC*, volume 5878 of *LNCS*, pages 112–121, 2009.

[BEGM09] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. Every stochastic game with perfect information admits a canonical form. RRR-09-2009, RUTCOR, Rutgers University, 2009.

[BEGM10a] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A lower bound for discounting algorithms solving two-person zero-sum limit average payoff stochastic games. RRR-22-2010, RUTCOR, Rutgers University, 2010.

[BEGM10b] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A pumping algorithm for ergodic stochastic mean payoff games with perfect information. In *Proc. 14th IPCO*, pages 341–354, 2010.

[BEGM11] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A potential reduction algorithm for two-person zero-sum limiting average payoff stochastic games, manuscript, 2011.

[BK78] T. Bewley and E. Kohlberg. On stochastic games with stationary optimal strategies. *Mathematics of Operations Research*, 3(2):104–125, 1978.

[Bla62] D. Blackwell. Discrete dynamic programming. *Ann. Math. Statist.*, 33:719–726, 1962.

[BV01a] E. Beffara and S. Vorobyov. Adapting Gurvich-Karzanov-Khachiyan's algorithm for parity games: Implementation and experimentation. Technical Report 2001-020, Department of Information Technology, Uppsala University, available at: https://www.it.uu.se/research/reports/#2001, 2001.

[BV01b] E. Beffara and S. Vorobyov. Is randomized Gurvich-Karzanov-Khachiyan's algorithm for parity games polynomial? Technical Report 2001-025, Department of Information Technology, Uppsala University, available at: https://www.it.uu.se/research/reports/#2001, 2001.

[CH08] K. Chatterjee and T. A. Henzinger. Reduction of stochastic parity to stochastic mean-payoff games. *Inf. Process. Lett.*, 106(1):1–7, 2008.

[CJH04] K. Chatterjee, M. Jurdziński, and T. A. Henzinger. Quantitative stochastic parity games. In *Proc. 15th SODA*, pages 121–130, 2004.

[Con92] A. Condon. The complexity of stochastic games. *Information and Computation*, 96:203–224, 1992.

[EM79] A. Eherenfeucht and J. Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8:109–113, 1979.

[Fed80] A. Federgruen. Successive approximation methods in undiscounted stochastic games. *Operations Research*, 1:794–810, 1980.

[Fil81] J. A. Filar. Ordered field property for stochastic games when the player who controls transitions changes from state to state. *J. of Optimization Theory and Applications*, 34(4):503–515, 1981.

[FTV07] J. Flesch, F. Thuijsman, and O.J. Vrieze. Stochastic games with additive transitions. *European Journal of Operational Research*, 179(2):483 – 497, 2007.

[Gal58] T. Gallai. Maximum-minimum Sätze über Graphen. *Acta Mathematica Academiae Scientiarum Hungaricae*, 9:395–434, 1958.

[Gil57] D. Gillette. Stochastic games with zero stop probabilities. In M. Dresher, A. W. Tucker, and P. Wolfe, editors, *Contribution to the Theory of Games III*, volume 39 of *Annals of Mathematics Studies*, pages 179–187. Princeton University Press, 1957.

[GKK88] V. Gurvich, A. Karzanov, and L. Khachiyan. Cyclic games and an algorithm to find minimax cycle means in directed graphs. *USSR Computational Mathematics and Mathematical Physics*, 28:85–91, 1988.

[Hal07] N. Halman. Simple stochastic games, parity games, mean payoff games and discounted payoff games are all LP-type problems. *Algorithmica*, 49(1):37–50, 2007.

[HK66] A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Science, Series A*, 12(5):359–370, 1966.

[HL31] G. H. Hardy and J. E. Littlewood. Notes on the theory of series (xvi): two tauberian theorems. *J. of London Mathematical Society*, 6:281–286, 1931.

[How60] R. A. Howard. *Dynamic programming and Markov processes*. Technology press and Willey, New York, 1960.

[JPZ06] M. Jurdzinski, M. Paterson, and U. Zwick. A deterministic subexponential algorithm for solving parity games. In *Proc. 17th SODA*, pages 117–123, 2006.

[Jur98] M. Jurdziński. Deciding the winner in parity games is in UP ∩ co-UP. *Inf. Process. Lett.*, 68(3):119–124, 1998.

[Kar78] R. M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Math.*, 23:309–311, 1978.

[KaR10] N. Krishnamurthy and T. Parthasarathy andG. Ravindran. Orderfield property of mixtures of stochastic games. *Mathematical Statistics and Probability*, 72(1):246–275, 2010.

[KMMS03] D. Kratsch, R. M. McConnell, K. Mehlhorn, and J. P. Spinrad. Certifying algorithms for recognizing interval graphs and permutation graphs. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 158–167, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[KS63] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. Springer, 1963.

[LL69] T. M. Liggett and S. A. Lippman. Stochastic games with perfect information and time-average payoff. *SIAM Review*, 4:604–607, 1969.

[Mil11] P. B. Miltersen. Discounted stochastic games poorly approximate undiscounted ones, manuscript. Technical report, 2011.

[MN81] J.F. Mertens and A. Neyman. Stochastic games. *Int. J. Game Theory*, 10:5366, 1981.

[MO70] H. Mine and S. Osaki. *Markovian decision process*. American Elsevier Publishing Co., New York, 1970.

[Mou76a] H. Moulin. Extension of two person zero sum games. *Journal of Mathematical Analysis and Application*, 5(2):490–507, 1976.

[Mou76b] H. Moulin. Prolongement des jeux à deux joueurs de somme nulle. *Bull. Soc. Math. France, Memoire*, 45, 1976.

[Pis99] N. N. Pisaruk. Mean cost cyclical games. *Mathematics of Operations Research*, 24(4):817–828, 1999.

[PR81] T. Parthasarathy and T. E. S. Raghavan. An orderfield property for stochastic games when one player controls transition probabilities. *Journal of Optimization Theory and Applications*, 33:375–392, 1981. 10.1007/BF00935250.

[RTV85] T. E. S. Raghavan, S. H. Tijs, and O. J. Vrieze. On stochastic games with additive reward and transition structure. *Journal of Optimization Theory and Applications*, 47:451–464, 1985. 10.1007/BF00942191.

[Sha53] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Science, USA*, 39:1095–1100, 1953.

[Sin89] S. Sinha. *A contribution to the theory of stochastic games*. PhD thesis, Indian Statistical Institute, , New Delhi, India, 1989.

[SS50] L. S. Shapley and R. N. Snow. Basic solutions of discrete games. *Annals of Mathematical Studies*, 24:27–35, 1950.

[vN28] J. v. Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

[Vri80] O. J. Vrieze. *Stochastic games with finite state and action spaces*. PhD thesis, Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands, 1980.

[ZP96] U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theoret. Comput. Sci.*, 158(1-2):343–359, 1996.

A Related Results from Theory of Markov Chains

Given a $n \times n$ transition matrix P, the Cesáro partial sums $\frac{1}{k+1}\sum_{i=1}^{k} P^i$ converge, as $k \to \infty$, to the limit Markov matrix Q such that:

(i) $PQ = QP = QQ = Q$; (ii) $rank(I-P) + rankQ = n$;

(iii) For each n-vector c system $Px = x,\ Qx = c$ has a unique solution.

(iv) matrix $I - (P - Q)$ is non-singular and

$$H(\delta) = \sum_{i=0}^{\infty} \delta^i (P^i - Q) \to H = (I - (P-Q))^{-1} - Q \ \text{ as } \ \delta \to 1^-; \tag{63}$$

(v)

$$H(\delta)Q = QH(\delta) = HQ = QH = 0 \ \text{ and } \ (I-P)H = H(I-P) = I - Q.$$

Claim (iv) (which is used in Section 6.2 was proved in 1962 by Blackwell, [Bla62], while for the remaining four claims, he cited the text-book in finite Markov chains by Kemeny and Snell [KS63] (that was published, in fact, one year later, in 1963).

B Stochastic Games and Linear Programming

It is well known that the one-player variant of a stochastic game is a Markovian decision process, and can be solved via linear programming (see e.g., [MO70]). In this section for completeness we recall some of the related results and models using our terminology and notation. We provide here a completely elementary proof based only on the theory of linear programming.

Let us consider a game Γ, and an arbitrary strategy of the maximizer $\alpha = (\alpha^v \mid v \in V) \in \mathcal{K}(\Gamma)$. We shall associate to the pair (Γ, α) a linear programming problem $LP(\alpha)$, having the values g^v and potentials x^v for $v \in V$ as its variables:

$$\sum_{v \in V} g^v \quad \to \quad \max$$

$$\sum_{u \in V} (\alpha^v P^{vu})_\ell \, g^u \quad \geq \quad g^v \quad \forall v \in V \text{ and } \ell \in L^v \tag{64}$$

$$(\alpha^v A^v)_\ell \ + \ x^v \ - \sum_{u \in V} (\alpha^v P^{vu})_\ell \, x^u \quad \geq \quad g^v \quad \forall v \in V \text{ and } \ell \in L^v. \tag{65}$$

Let us note that conditions (64) and (65) imply that for any $\beta = (\beta^v \mid v \in V) \in \mathcal{L}(\Gamma)$ we have

$$\alpha^v G^v(g) \beta^v \geq g^v \quad \text{ and } \quad \alpha^v A^v(x) \beta^v \geq g^v \tag{66}$$

for all states $v \in V$, where $G^v(g)$ is the matrix defined in (21). We shall show that for some optimal solution $x, g \in \mathbb{R}^V$ of $LP(\alpha)$ there exists a strategy $\overline{\beta}$ for which the inequalities in (66) are equations for all states $v \in V$ simultaneously. We prove that such a $\beta(\alpha) = \overline{\beta}$ is a uniform best response to α. Furthermore, we can also show that $\overline{\beta}$ can be chosen to be a pure strategy at each of the states.

To this end let us first recall some remarkable properties of this LP. Let us refer to the inequalities of the above LP as $(64)(v, \ell)$ and $(65)(v, \ell)$ for all $v \in V$ and $\ell \in L^v$.

Let us start by showing that $LP(\alpha)$ is feasible and bounded, that is, it has a finite optimum.

Lemma 7 *Problem $LP(\alpha)$ has a finite optimum.*

Proof For feasibility let us observe that (64) is satisfied with $g^v = D$, $v \in V$ for any constant D. Thus, choosing $D \leq \min_{v \in V, \ell \in L^v} (\alpha^v A^v)_\ell$ and $x^v = 0$, $v \in V$, all constraints (65) are satisfied, too.

To see the boundedness, let us recall from the theory of linear programming that $LP(\alpha)$ is unbounded if and only if the corresponding homogenized system of inequalities has a feasible solution with positive objective function value, that is, if there are vectors (y, f) such that

$$\sum_{v \in V} f^v \quad = \quad 1 \tag{67}$$

$$\sum_{u \in V} (\alpha^v P^{vu})_\ell \, f^u \quad \geq \quad f^v \quad \forall v \in V \text{ and } \ell \in L^v \tag{68}$$

$$y^v \; - \sum_{u \in V} (\alpha^v P^{vu})_\ell \, y^u \quad \geq \quad f^v \quad \forall v \in V \text{ and } \ell \in L^v. \tag{69}$$

Let us then define $f^* = \max_{v \in V} f^v$, set $M = \{v \in V \mid f^v = f^*\}$, and let $w \in M$ be a state in M for which $y^w \leq y^v$ for all $v \in M$.

Then, inequalities (68)(v, ℓ) require that f^v for $v \in M$ is not larger than a convex combination of the f^u values, $u \in V$, implying that $(\alpha^v P^{vu})_\ell = 0$ for all $v \in M$, $u \notin M$, and $\ell \in L^v$. Since (67) implies $f^* > 0$, inequalities (69)(w, ℓ) for $\ell \in L^w$ imply that y^w is strictly larger (by $f^* > 0$) than a convex combination of the values y^v, $v \in M$, which is impossible by the choice of w. This contradiction proves that the system (67)-(69) is infeasible, proving that $LP(\alpha)$ is bounded. □

Let us next show that a simple family of linear transformations of potentials does not change the feasibility in $LP(\alpha)$. To simplify notations, let us introduce $\lambda(g, v, \ell)$ to denote the left hand side value in inequality (64)(v, ℓ), and let $\mu(x, v, \ell)$ denote the left hand side value in (65)(v, ℓ).

Lemma 8 *If (x, g) is feasible in $LP(\alpha)$, $C, D \in \mathbb{R}$ are constants, $C \geq 0$, and*

$$y^v = x^v + D - Cg^v$$

for all states $v \in V$, then (y, g) is also feasible in $LP(\alpha)$.

Proof The left hand sides $\lambda(g, v, \ell)$ do not change by the above operation, and for (65)(v, ℓ) we have

$$\mu(y, v, \ell) \;=\; \mu(x, v, \ell) + C\,(\lambda(g, v, \ell) - g^v) \geq g^v \tag{70}$$

for all states $v \in V$ and $\ell \in L^v$ by the feasibility of (x, g) and by the fact that $\sum_{u \in V} (\alpha^v P^{vu})_\ell \;=\; 1$. □

For a feasible solution (x, g) let us define $I^v = I^v(g) \subseteq L^v$ to be the set of pure strategies $\ell \in L^v$ for which (64)(v, ℓ) is tight, that is for which $\lambda(g, v, \ell) = g^v$.

Lemma 9 *For any optimal solution (x, g) of $LP(\alpha)$ we have $I^v(g) \neq \emptyset$ for all states $v \in V$.*

Proof Assume indirectly that

$$\lambda(g, v, \ell) > g^v \tag{71}$$

for all $\ell \in L^v$ for some state $v \in V$. Let us then consider the potential vector y as defined in Lemma 8 with $D = 0$ and $C > 0$. Then (y, g) is again feasible by Lemma 8, and has the same objective function value as (x, g), thus it is again optimal. Furthermore, by (70) and (71) we have

$$\mu(y, v, \ell) \;=\; \mu(x, v, \ell) + C\,(\lambda(g, v, \ell) - g^v) > g^v$$

for all $\ell \in L^v$. Thus, increasing g^v by a small positive quantity will not change the feasibility of (y, g), since on the right hand side g^v is involved only in strict inequalities. This contradicts the optimality (y, g), proving our claim. □

For a feasible solution (x, g) let us define $J^v = J^v(x, g) \subseteq L^v$ to be the set of pure strategies $\ell \in L^v$ for which (65)(v, ℓ) is tight, that is, for which $\mu(x, v, \ell) = g^v$.

Lemma 10 *For any optimal solution (x, g) of $LP(\alpha)$, there exists a potential $y \in \mathbb{R}^V$ such that (y, g) is also optimal in $LP(\alpha)$ and that*

$$\emptyset \;\neq\; J^v(y,g) \subseteq I^v(g) \tag{72}$$

holds for all states $v \in V$.

Proof By Lemma 9 we can assume that $I^v(g) \neq \emptyset$ for all states $v \in V$. It is enough to show that there exists a potential $y \in \mathbb{R}^V$ such that (y, g) is optimal in $LP(\alpha)$ and $J^v(y, g) \cap I^v(g) \neq \emptyset$ for all $v \in V$, since we can then apply Lemma 8 and modify the potential y to satistfy (72).

Let us now fix the value vector g, and denote by

$$Y(g) = \{y \in \mathbb{R}^V \mid (\alpha^v A^v)_\ell \;+\; y^v \;-\sum_{u\in V} (\alpha^v P^{vu})_\ell \, y^u \geq \; g^v \quad \forall v \in V \text{ and } \ell \in I^v(g)\}.$$

Clearly, $Y(g)$ is a closed and convex non-empty set in $\mathbb{R}^V$ (since $x \in Y(g)$). Let us further denote by $\tilde{Y}(g)$ the set of potentials $y \in Y(g)$ such that (y, g) is feasible for $LP(\alpha)$ and satisfies $J^v(y, g) \subseteq I^v(g)$ (possibly, $J^v(y, g) = \emptyset$), for all states $v \in V$. By Lemma 8, every $y \in Y(g)$ can be transformed into a vector of potentials in $\tilde{Y}(g)$, by choosing a sufficiently large but finite C; for each $y \in Y(g)$, we will fix an arbitrary such vector in $\tilde{Y}(g)$ and denote it by $\tilde{y}(g)$. Let us define

$$\epsilon^v(y) \;=\; \min_{\ell \in I^v(g)} \mu(y, v, \ell) - g^v \;\geq\; 0$$

for all states $v \in V$ and potentials $y \in Y(g)$, and call a state $v \in V$ *tight* with respect to (y, g) if $\epsilon^v(y) = 0$. Finally, let us denote by $T(y) \subseteq V$ the subset of tight states with respect to a given $y \in Y(g)$.

Let us first note that for all $y \in Y(g)$ and for all subsets $S \subseteq V \setminus T(y)$ we must have a $v \in S$, a $u \notin S$ and an $\ell \in I^v(g)$ such that $(\alpha P^{vu})_\ell > 0$ holds, since otherwise we would have $(\tilde{y}(\tilde{g}), \tilde{g})$ feasible in $LP(\alpha)$ for some small enough $\epsilon > 0$, where

$$\tilde{g}^u = \begin{cases} g^u + \epsilon & \text{if } u \in S \\ g^u & \text{otherwise,} \end{cases}$$

contradicting the optimality of (x, g). Let us call this property (*). This property implies, in particular, that we must have $T(y) \neq \emptyset$ for all $y \in Y(g)$ (since no positive probability arc leaves the set $S = V$).

Let us next note that $T(\lambda y + (1-\lambda)y') \subseteq T(y) \cap T(y')$ holds for all $y, y' \in Y(g)$ and $0 < \lambda < 1$. Consequently, we must have $U = \{v \in V \mid v \in T(y) \;\forall\; y \in Y(g)\} \neq \emptyset$, since otherwise, if for all $v \in V$ there is a potential vector $y^v \in Y(g)$ such that $v \notin T(y^v)$, then by the previous observation we would have $T(\frac{1}{|V|}\sum_{v\in V} y^v) = \emptyset$, contradicting the above consequence of property (*). This also implies, in particular, that if $U \neq V$, then the vector $\bar{y} = \frac{1}{|V|}\sum_{v\notin U} y^v$ satsifies $T(\bar{y}) = U$.

Let us also note that for any $y \in Y(g)$ and any state $u \notin T(y)$, decreasing y^u by a small positive quantity keeps $y \in Y(g)$, since y^u is involved with a positive coefficient only in non-tight inequalities. Let us call this property (**).

Let us now introduce variables z^v for $v \in S = V \setminus U$, define

$$y(z)^u = \begin{cases} \bar{y}^u - z^u & \text{if } u \in S, \\ \bar{y}^u & \text{otherwise,} \end{cases}$$

and consider the linear programming problem LPZ:

$$\max\{\sum_{u\in S} z^u \mid y(z) \in Y(g),\; z \geq 0\}.$$

We claim that this LP is feasible, bounded, and hence has a finite optimum $\tilde{z}$. Then we must have $T(y(\tilde{z})) = V$, since otherwise by property (**) we could slightly increase the value of $\tilde{z}^u$ for a $u \in V \setminus T(y(\tilde{z}))$, contradicting the optimality of $\tilde{z}$ (note that all the states in U remain tight, by defintion of U).

To complete the proof, we only need to show that LPZ is feasible and bounded. Since $z = 0$ yields $y(z) = \bar{y} \in Y(g)$, feasibility of LPZ follows. For the boundedness let us recall again from the theory of linear programming that LPZ is not bounded only if the homogenized system of inequalities has a solution with a strictly positive objective function value. Assume indirectly that z is such a solution and let $Z > 0$ be the largest component of z (which then must be positive, since the sum of the components is positive by our indirect assumption). Let us further denote by $M = \{v \in S \mid z^v = Z\}$ the set of states with this maximal z^v value. Let us also extend in our mind the vector z with zero coefficients for states $u \in U$. Then the homogeneous inequality for $v \in M$ and $\ell \in I^v(g)$ states that a convex combination of the z^u, $u \in V$ values is at least as large as Z, which of course is possible only if $(\alpha P^{vu})_\ell = 0$ for all $v \in M$, $u \notin M$ and $\ell \in I^v(g)$. In this case we could add the same sufficiently small $\epsilon > 0$ to all g^v, $v \in M$ components, and get another feasible solution $(\tilde{y}(\tilde{g}), \tilde{g})$, contradicting the optimality of (x, g) in $LP(\alpha)$.

Finally, note that the pair $(\tilde{y}, g)$ satisfy the statement of the lemma, where $y = y(\tilde{z})$. □

Corollary 9 *Let $\bar{x}, \bar{g}$ be an optimal solution in $LP(\alpha)$. Then there exists a (pure) strategy $\overline{\beta} \in \mathcal{L}(\Gamma)$ for which we have equalities in* (66)*, for all states $v \in V$ simultaneously.*

Proof By Lemma 10 we can assume that $\emptyset \neq J^v(\bar{x}, \bar{g}) \subseteq I^v(\bar{g})$ for all states $v \in V$. Let us then choose $\overline{\beta}^v \in \Delta(J(\bar{x}, \bar{g}))$ (possibly a pure strategy) for all states $v \in V$. □

Lemma 11 *Let $\alpha_1, \alpha_2 \in \mathcal{K}(\Gamma)$ be two strategies of the maximizer. Let us further assume that (x_i, g_i) is a feasible solution in $LP(\alpha_i)$ for $i = 1, 2$. Let us then define α_3 and g_3 as the maxima of the two, that is, for each state $v \in V$ we set $\alpha_3^v = \alpha_1^v$ and $g_3^v = g_1^v$ if $g_1^v \geq g_2^v$, and set $\alpha_3^v = \alpha_2^v$ and $g_3^v = g_2^v$ if $g_1^v < g_2^v$. Then there exists a potential vector $x_3 \in \mathbb{R}^V$ such that (x_3, g_3) is feasible in $LP(\alpha_3)$.*

Proof

Let us next define $S_1 = \{v \in V \mid g_1^v \geq g_2^v\}$, and set $S_2 = V \setminus S_1$. Then, for any $v \in S_i$, $i = 1, 2$ we have

$$\sum_{u \in V} (\alpha_3^v P^{vu})_\ell \, g_3^u = \sum_{u \in V} (\alpha_i^v P^{vu})_\ell \max(g_i^u, g_{3-i}^u) \geq \sum_{u \in V} (\alpha_i^v P^{vu})_\ell \, g_i^u \geq g_i^v$$

by the feasibility of (x_i, g_i) in $LP(\alpha_i)$ showing the feasibility of g_3 in (64) of $LP(\alpha_3)$.

Let us next define $y^v = x_i^v$ for all $v \in S_i$, $i = 1, 2$, and set $x_3^v = y^v - C g_3^v$ for an appropriately large positive constant C, and consider a state $v \in S_i$:

$$\begin{aligned} \mu(x_3, v, \ell) &= \mu(x_i, v, \ell) + C\,(\lambda(g_i, v, \ell) - g_i^v) \\ &\qquad + \sum_{u \in S_{3-i}} (\alpha_i^v P^{vu})_\ell \left(C(g_{3-i}^u - g_i^u) + x_i^u - x_{3-i}^u\right) \\ &\geq \mu(x_i, v, \ell) \geq g_i^v = g_3^v, \end{aligned}$$

since we have $\lambda(g_i, v, \ell) \geq g_i^v$ by the feasibility of (x_i, g_i) in $LP(\alpha_i)$, and $g_{3-i}^u \geq g_i^u$ for all $u \in S_{3-i}$ by the definition of S_{3-i}. It follows that (x_3, g_3) is feasible in $LP(\alpha_3)$. □

Corollary 10 *Let $\bar{x}, \bar{g}$ be an optimal solution and let (x, g) be an arbitrary feasible solution in $LP(\alpha)$. Then we have $\bar{g}^v \geq g^v$ for all states $v \in V$.*

Proof Otherwise, we could construct a new feasible solution $(\hat{x}, \hat{g})$, $\hat{g}^v = \max(\bar{g}^v, g^v)$, $v \in V$ by Lemma 11 (applied with $\alpha_1 = \alpha_2 = \alpha$), which has a strictly larger objective functions value, a contradiction with the optimality of g. □

Proof of Lemma 1. We can view the problem of computing the best response of the minimizer as a special stochastic game $\widetilde{\Gamma}$ in which the action sets are $\widetilde{K}^v = \{\alpha\}$ and $\widetilde{L}^v = L^v(\Gamma)$ for all states $v \in V$, the rewards are $\widetilde{r}_{\alpha\ell}^{vu} = \sum_{k \in K^v(\Gamma)} \alpha_k^v r_{k\ell}^{vu}(\Gamma)$ and the transition probabilities are $\widetilde{p}_{\alpha\ell}^{vu} =$

$\sum_{k\in K^v(\Gamma)} \alpha_k^v p_{k\ell}^{vu}(\Gamma)$ for all states $v,u \in V$ and actions $\ell \in \widetilde{L}^v$. Let us denote by $\widetilde{A}^v$, $v \in V$ the reward matrices of $\widetilde{\Gamma}$.

Let us now consider an optimal solution $(\widetilde{x},\widetilde{g})$ to $LP(\alpha)$. By Lemma 10 we can assume that $\emptyset \neq J^v(\widetilde{x},\widetilde{g}) \subseteq I^v(\widetilde{g})$ holds for all $v \in V$. Thus, the existence of $\overline{\beta}$ by Corollary 9 and the inequalities (64) and (65) for $(\widetilde{x},\widetilde{g})$ proves that for $\widetilde{\Gamma}$ we have

$$\text{Val } (G^v(\widetilde{g})) \;=\; \widetilde{g}^v \quad \text{ and } \quad \text{Val } \left(\widetilde{A}^v(\widetilde{x})\right) \;=\; \widetilde{g}^v \quad \text{ for all states } \; v \in V.$$

Furthermore, we have $\bar{L}^v = \Lambda(G^v(\widetilde{g})) = \Delta(I^v(\widetilde{g}))$ and $\bar{K}^v = \Omega(G^v(\widetilde{g})) = \widetilde{K}^v$ since we have a $1 \times |\widetilde{L}^v|$ matrix game in state $v \in V$. Thus, by $\emptyset \neq J^v(\widetilde{x},\widetilde{g}) \subseteq I^v(\widetilde{g})$, for all states $v \in V$ we get

$$\underset{\widetilde{K}^v \times \bar{L}^v}{\text{Val}} \left(\widetilde{A}^v(\widetilde{x})\right) \;=\; \underset{\bar{K}^v \times \widetilde{L}^v}{\text{Val}} \left(\widetilde{A}^v(\widetilde{x})\right) \;=\; \text{Val } \left(\widetilde{A}^v(\widetilde{x})\right) \;=\; \widetilde{g}^v$$

for all states $v \in V$. Hence, $(\widetilde{x},\widetilde{g})$ transforms $\widetilde{\Gamma}$ into a weak canonical form (B2). by te implication (B2) $\Rightarrow$ (A1) in Theorem 1 (whose elementary proof is given in Appendix C), $\widetilde{\Gamma}$ has values $\widetilde{g}$ that can be realized by a uniformly optimal stationary strategy. In fact, by Corollary 9, the strategy $\overline{\beta}$ is such a uniform optimum in $\widetilde{\Gamma}$ (and it can be chosen to be a pure strategy at each of the states). Hence $\beta(\alpha) = \overline{\beta}$ is a (pure) uniform best response strategy. □

Proof of Lemma 2. Let $\alpha_1 = \alpha$ and $\alpha_2 = \alpha'$ and denote by (x_i, g_i) an optimal solution of $LP(\alpha_i)$ for $i = 1, 2$. Then by Lemma 11 we can define α_3 and g_3 (such that $g_3^v = \max(g_1^v, g_2^v)$) and derive the existence of x_3 such that (x_3, g_3) is feasible in $LP(\alpha_3)$. Then, $\alpha'' = \alpha_3$ proves the claim. □

C Proof of Lemma 3

Let us first fix the strategy $\bar{\beta}$ of the minimizer, and compute the uniformly best response by the maximizer by solving a controlled Markov chain problem by linear programming (see Appendix B). This LP provides us with a potential vector $y \in \mathbb{R}^V$ such that

$$g^v \geq \text{ Val } (A^v(y)) \tag{73}$$

holds for all states $v \in V$, according to Corollary 9. Let us next fix $\bar{\alpha}$ and compute similarly the best response of the minimizer, providing analogously a potential vector $z \in \mathbb{R}^V$ satisfying

$$g^v \leq \text{ Val } (A^v(z)) \tag{74}$$

for all states $v \in V$. Since adding a constant to a potential vector does not change the potential transformation and the value matrices, we can assume w.l.o.g. that

$$y \leq z. \tag{75}$$

Let us define a matrix valued mapping $B^v(d)$ for all states $v \in V$ and vectors $d \in \mathbb{R}^V$ by

$$B^v(d) \;=\; A^v(d) - d^v J_{|K^v|\times|L^v|}.$$

Then we have by (75) that $B^v(z) \leq B^v(y)$ (componentwise), and since the value function of matrix games is monotone we can conclude by (73) and (74), and by the fact that changing the payoff matrix by a constant changes the value of the game by the same constant that

$$g^v - z^v \leq \text{ Val } (B^v(z)) \leq \text{ Val } (B^v(y)) \leq g^v - y^v, \tag{76}$$

for all states $v \in V$. Note that if $g - z \leq g - d \leq g - y$, then we have $B^v(z) \leq B^v(d) \leq B^v(y)$ for all $v \in V$, and hence by the above cited properties of the value function and by (76)

$$\text{Val } (B^v(z)) \leq \text{ Val } (B^v(d)) \leq \text{ Val } (B^v(y)) \text{ for all states } v \in V. \tag{77}$$

Since the mapping $F : g - d \mapsto \text{ Val }(B(d))$ (where $\text{ Val }(B(d)) = (\text{ Val }(B^v(d)) : v \in V)$) is Lipschitz-continuous and since by property (77) and (76) it maps the compact box $[g-z, g-y]$ into a subset of itself, we can conclude by Brouwer's theorem that F has a fixed point, that is there exists a potential vector $x \in [y, z]$ (i.e. $g - x \in [g - z, g - y]$) for which $g - x = F(g - x) = \text{ Val }(B(x))$. This implies that $g = \text{ Val }(A(x))$, completing our proof. □

D Proof of the Implication $(B2) \Rightarrow (A1)$

Let $g, x, y \in \mathbb{R}^V$ be the vectors satisfying condition (B2). Then there exist strategies $\bar{\alpha} \in \mathcal{K}(\Gamma)$ and $\bar{\beta} \in \mathcal{L}(\Gamma)$ such that, for all states $v \in V$, the following hold: (1) $\bar{\alpha}^v G^v(g)\beta \geq g^v$ and $\bar{\alpha}^v A^v(x)\beta \geq g^v$ for all $\beta \in \Delta(L^v)$, and (2) $\alpha^v G^v(g)\bar{\beta}^v \leq g^v$ and $\alpha^v A^v(y)\bar{\beta}^v \leq g^v$ for all $\alpha \in \Delta(K^v)$.

Fix a starting position $v_0 = w$. It is enough to show that player 1 can guarantee at least g^w while player 2 can guarantee at most g^w. We only show the former statement since the latter can be shown similarly. At time i, we will let player 1 play his/her locally optimal (stationary) strategy $\bar{\alpha}^v$ whenever (s)he is at position $v_i = v$, while player 2 chooses an arbitrary, not necessarily stationary, strategy $\beta^{\mathcal{H}}$, where $\mathcal{H} \in \mathcal{H}_i(v)$ is the history of the play leading to $v_i = v$ and $\mathcal{H}_i(v)$ is the set of all such histories. Let us note that $\sum_{\mathcal{H}\in\mathcal{H}_i(v)} \Pr[\mathcal{H}|\ v_i = v] = 1$ and denote by $\beta^{v,i} \in \Delta(L^v)$ the Markovian strategy given by $\sum_{\mathcal{H}\in\mathcal{H}_i(v)} \beta^{\mathcal{H}} \Pr[\mathcal{H}|\ v_i = v]$.

Consider a play $w = v_0, v_1, v_2, \ldots$ (where each v_i is a random variable). By (7) and the fact that potential transformations do not change the Cesáro sum (Section 1.4), it is enough to show that $\mathbb{E}[b_i(x)] \geq g^w$ for all i. Note that

$$\begin{aligned}
\mathbb{E}[b_i(x)] &= \sum_v \sum_{\mathcal{H}\in\mathcal{H}_i(v)} \mathbb{E}[b_i(x)|\ v_i = v,\ \mathcal{H}] \cdot \Pr[\mathcal{H}|\ v_i = v] \cdot \Pr[v_i = v] \\
&= \sum_v \sum_{\mathcal{H}\in\mathcal{H}_i(v)} \bar{\alpha}^v A^v(x)\beta^{\mathcal{H}} \Pr[\mathcal{H}|\ v_i = v] \cdot \Pr[v_i = v] \\
&= \sum_v \bar{\alpha}^v A^v(x)\beta^{v,i} \cdot \Pr[v_i = v] \\
&\geq \sum_v g^v \cdot \Pr[v_i = v]. \qquad (78)
\end{aligned}$$

We prove by induction on $i = 0, 1, 2, \ldots$, that $\sum_v g^v \cdot \Pr[v_i = v] \geq g^w$, which will imply the lemma by (78). Indeed, the statement is trivially true for $i = 0$. For any i, we have

$$\begin{aligned}
\sum_v g^v \cdot \Pr[v_i = v] &= \sum_v g^v \cdot \sum_u \sum_{\mathcal{H}\in\mathcal{H}_{i-1}(u)} \Pr[v_i = v|v_{i-1} = u,\ \mathcal{H}] \cdot \Pr[\mathcal{H}|\ v_{i-1} = u] \cdot \Pr[v_{i-1} = u] \\
&= \sum_v g^v \cdot \sum_u \sum_{\mathcal{H}\in\mathcal{H}_{i-1}(u)} \bar{\alpha}^u P^{uv} \beta^{\mathcal{H}} \cdot \Pr[\mathcal{H}|\ v_{i-1} = u] \cdot \Pr[v_{i-1} = u] \\
&= \sum_v g^v \cdot \sum_u \bar{\alpha}^u P^{uv} \beta^{u,i-1} \cdot \Pr[v_{i-1} = u] \\
&= \sum_u \bar{\alpha}^u \cdot (\sum_v P^{uv} g^v) \beta^{u,i-1} \cdot \Pr[v_{i-1} = u] \\
&= \sum_u \bar{\alpha}^u \cdot G^u(g)\beta^{u,i-1} \cdot \Pr[v_{i-1} = u] \\
&\geq \sum_u g^u \cdot \Pr[v_{i-1} = u]
\end{aligned}$$

and the latter is at least g^w by the induction hypothesis. □