A Simple P-Matrix Linear Complementarity Problem for Discounted Games*

Marcin Jurdziński and Rahul Savani**

Department of Computer Science, University of Warwick, UK
{mju,rahul}@dcs.warwick.ac.uk

Abstract. The values of a two-player zero-sum binary discounted game are characterized by a P-matrix linear complementarity problem (LCP). Simple formulas are given to describe the data of the LCP in terms of the game graph, discount factor, and rewards. Hence it is shown that the unique sink orientation (USO) associated with this LCP coincides with the strategy valuation USO associated with the discounted game. As an application of this fact, it is shown that Murty's least-index method for P-matrix LCPs corresponds to both known and new variants of strategy improvement algorithms for discounted games.

Keywords: Discounted game, linear complementarity problem, P-matrix, strategy improvement algorithm, unique sink orientation, zero-sum game.

1 Introduction

Discounted (stochastic) games were introduced by Shapley [15]. The monograph of Filar and Vrieze [6] discusses discounted (stochastic) games in detail. For clarity, we only consider non-stochastic discounted games in this paper. One motivation for studying these games is that there is a polynomial time reduction to discounted games from parity games (via mean-payoff games) [14,21], which are equivalent to model checking for the modal mu-calculus. A polynomial-time algorithm for parity games is a long-standing open question.

Our contribution is a transparent reduction from binary discounted games to the P-matrix linear complementarity problem (LCP). The simple formulas for the LCP data allow us to show that the unique sink orientation of the cube associated with the P-matrix LCP [16] is the same as the strategy valuation USO for the game. As an application of this fact, it is shown that Murty's least-index method for P-matrix LCPs corresponds to both known and new variants of strategy improvement algorithms for binary discounted games. For games with outdegree greater than two, one gets *generalized* LCPs. Discounted games can be reduced in polynomial time to simple stochastic games [21]. Recently (non-binary) simple stochastic games have been reduced to P-matrix (generalized)

* This research was supported in part by EPSRC projects EP/D067170/1 and EP/E022030/1.

** Corresponding author.

A. Beckmann, C. Dimitracopoulos, and B. Löwe (Eds.): CiE 2008, LNCS 5028, pp. 283–293, 2008.

LCPs [17,8]. The monograph of Cottle et. al. [4] is the authoritative source on the linear complementarity problem.

2 Discounted Games

A (perfect-information binary) discounted game Γ=$(S, \lambda, \rho, r_\lambda, r_\rho, \beta, S_{\text{Min}}, S_{\text{Max}})$ consists of: a set of states $S = \{ 1, 2, \ldots, n \}$; *left* and *right successor* functions $\lambda, \rho : S \to S$, respectively; *reward* functions $r_\lambda, r_\rho : S \to \mathbb{R}$ for left and right edges respectively, with $r_\lambda(s) = r_\rho(s)$ if $\lambda(s) = \rho(s)$; a *discount factor* $\beta \in [0, 1)$; and a partition $(S_{\text{Min}}, S_{\text{Max}})$ of the set of states. A sequence $\langle s_0, s_1, s_2, \ldots \rangle \in S^\omega$ is a *play* if for all $i \in \mathbb{N}$, we have that $\lambda(s_i) = s_{i+1}$ or $\rho(s_i) = s_{i+1}$. We define the *(β-)discounted payoff* $\mathcal{D}(\pi, \beta)$ of a play $\pi = \langle s_0, s_1, s_2, \ldots \rangle$ by $\mathcal{D}(\pi, \beta) = \sum_{i=0}^{\infty} \beta^i r(s_i, s_{i+1})$, with $r(s_i, s_{i+1})$ denoting $r_\lambda(s_i)$ or $r_\rho(s_i)$ as appropriate.

A function $\mu : S_{\text{Min}} \to S$ is a positional strategy for player Min if for every $s \in S_{\text{Min}}$, we have that $\mu(s) = \lambda(s)$ or $\mu(s) = \rho(s)$. Strategies $\chi : S_{\text{Max}} \to S$ for player Max are defined analogously. We write Π_{Min} and Π_{Max} for the sets of positional strategies for player Min and Max, respectively. For strategies $\mu \in \Pi_{\text{Min}}$ and $\chi \in \Pi_{\text{Max}}$, and a state $s \in S$, we write $\text{Play}(s, \mu, \chi)$ for the play $\langle s_0, s_1, s_2, \ldots \rangle$, such that $s_0 = s$, and for all $i \in \mathbb{N}$, we have that $s_i \in S_{\text{Min}}$ implies $\mu(s_i) = s_{i+1}$, and $s_i \in S_{\text{Max}}$ implies $\chi(s_i) = s_{i+1}$. A function $\sigma : S \to S$ is a (combined) positional strategy. For a combined positional strategy $\sigma : S \to S$, we write $\text{Play}(s, \sigma)$ for the play $\text{Play}(S, \sigma{\restriction}S_{\text{Min}}, \sigma{\restriction}S_{\text{Max}})$.

For every $s \in S$, we define the *lower value* $\text{Val}_*(s, \beta)$ and the *upper value* $\text{Val}^*(s, \beta)$ by

$$\text{Val}_*(s, \beta) = \max_{\chi \in \Pi_{\text{Max}}} \min_{\mu \in \Pi_{\text{Min}}} \mathcal{D}(\text{Play}(s, \mu, \chi), \beta),$$
$$\text{Val}^*(s, \beta) = \min_{\mu \in \Pi_{\text{Min}}} \max_{\chi \in \Pi_{\text{Max}}} \mathcal{D}(\text{Play}(s, \mu, \chi), \beta).$$

The inequality $\text{Val}_*(s, \beta) \leq \text{Val}^*(s, \beta)$ always holds. We say that the *value* exists in a state $s \in S$, if we have $\text{Val}_*(s, \beta) = \text{Val}^*(s, \beta)$; we then write $\text{Val}(s, \beta)$ for $\text{Val}_*(s, \beta) = \text{Val}^*(s, \beta)$. We say that the discounted game is *positionally determined* if for all $s \in S$, the value exists in s.

We identify functions $\overline{v} : S \to \mathbb{R}$ and n-vectors $\overline{v} \in \mathbb{R}^n$. For $s \in S$, depending on which interpretation is more natural in context, we write either $\overline{v}(s)$ or $\overline{v}_s$. We do the same for n-vectors of variables, for which Latin letters v, w, and z are typically used. We say that $\overline{v} : S \to \mathbb{R}$ is a solution of the optimality equations $\text{Opt}(\Gamma)$ if for all $s \in S$, we have

$$\overline{v}(s) = \begin{cases} \min\{\, r_\lambda(s) + \beta \cdot \overline{v}(\lambda(s)),\ r_\rho(s) + \beta \cdot \overline{v}(\rho(s)) \,\} & \text{if } s \in S_{\text{Min}}, \\ \max\{\, r_\lambda(s) + \beta \cdot \overline{v}(\lambda(s)),\ r_\rho(s) + \beta \cdot \overline{v}(\rho(s)) \,\} & \text{if } s \in S_{\text{Max}}. \end{cases} \tag{1}$$

Theorem 1 ([15]). *Every discounted game is positionally determined. Moreover, the optimality equations Opt(Γ) have a unique solution $\overline{v} : S \to \mathbb{R}$, and for every $s \in S$, we have Val$(s, \beta) = \overline{v}$.*

It follows from the existence of a solution to the optimality equations that there exist optimal *pure* positional strategies [21]. Hence, without loss of generality, we consider only pure positional strategies.

3 A P-Matrix LCP for Discounted Games

3.1 An LCP for Discounted Games

Consider the following set of constraints over variables $v(s), w(s), z(s)$, for all $s \in S$:

$$v(s) + w(s) = r_\lambda(s) + \beta v(\lambda(s)), \text{ if } s \in S_{\text{Min}}, \tag{2}$$
$$v(s) - w(s) = r_\lambda(s) + \beta v(\lambda(s)), \text{ if } s \in S_{\text{Max}}, \tag{3}$$
$$v(s) + z(s) = r_\rho(s) + \beta v(\rho(s)), \text{ if } s \in S_{\text{Min}}, \tag{4}$$
$$v(s) - z(s) = r_\rho(s) + \beta v(\rho(s)), \text{ if } s \in S_{\text{Max}}, \tag{5}$$
$$w(s), z(s) \geq 0 \tag{6}$$
$$w(s) \cdot z(s) = 0. \tag{7}$$

Non-negative variables $w(s)$ and $z(s)$ should be thought of as slack variables which turn inequalities such as $v(s) \leq r_\lambda(s) + \beta v(\lambda(s))$ if $s \in S_{\text{Min}}$, or $v(s) \geq r_\rho(s) + \beta v(\rho(s))$ if $s \in S_{\text{Max}}$, into equations. Note that variables w are slacks for left successors, and variables z are slacks for right successors. The natural inequalities for left and right successors, turned into equations (2)–(5) using non-negative slack variables (6), together with the *complementarity* condition (7), for all $s \in S$, yield the following characterization.

Proposition 1. *There is a unique solution $\overline{v}, \overline{w}, \overline{z} : S \to \mathbb{R}$ of constraints (2)–(7), and $\overline{v}$ is the unique solution of Opt(Γ).*

A linear complementarity problem [4] LCP(M, q) is the following set of constraints:

$$w = Mz + q, \tag{8}$$
$$w, z \geq 0, \tag{9}$$
$$w_s \cdot z_s = 0, \text{ for every } s \in S, \tag{10}$$

where M is an $n \times n$ real matrix, $q \in \mathbb{R}^n$, and w and z are n-vectors of real variables. In order to turn constraints (2)–(7) into a linear complementarity problem LCP(M, q), we rewrite equations (2)–(5) in matrix notation and eliminate variables $v(s)$, for all $s \in S$.

For a predicate p, we define $[p] = 1$ if p holds, and $[p] = 0$ if p does not hold. For $\sigma : S \to S$, define the $n \times n$ matrix T_σ by $(T_\sigma)_{st} = [\sigma(s) = t]$, for all $s, t \in S$. For every $n \times n$ matrix A, we define the matrix $\widehat{A}$ by setting $(\widehat{A})_{st} = (-1)^{[s \in S_{\text{Min}}]} A_{st}$, for every $s, t \in S$. Observe that $\widehat{A}$ is obtained from A by multiplying all entries in every row s, such that $s \in S_{\text{Min}}$, by -1.

Equations (2)–(3) and (4)–(5) can be written as

$$\widehat{I}v = w + \widehat{I}r_\lambda + \beta\widehat{T_\lambda}v,$$
$$\widehat{I}v = z + \widehat{I}r_\rho + \beta\widehat{T_\rho}v,$$

respectively, where v, w, and z are n-vectors of real variables, and $r_\lambda, r_\rho \in \mathbb{R}^n$ are the vectors of rewards. By eliminating v we get

$$w + \widehat{I}r_\lambda = (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}(z + \widehat{I}r_\rho),$$

and hence we obtain an LCP(M, q) equivalent to constraints (2)–(7), where

$$M = (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}, \tag{11}$$
$$q = M\widehat{I}r_\rho - \widehat{I}r_\lambda. \tag{12}$$

Proposition 2. *There is a unique solution $\overline{w}, \overline{z} \in \mathbb{R}^n$ of the LCP(M, q), and $(\widehat{I} - \beta\widehat{T_\lambda})^{-1}(\overline{w} + \widehat{I}r_\lambda) = (\widehat{I} - \beta\widehat{T_\rho})^{-1}(\overline{z} + \widehat{I}r_\rho)$ is the unique solution of Opt(Γ).*

Invertibility of $(\widehat{I} - \beta\widehat{T_\lambda})$ and $(\widehat{I} - \beta\widehat{T_\rho})$ is guaranteed by Theorem 4.

3.2 The P-Matrix Property

For an $n \times n$ matrix A and $\alpha \subseteq S$, such that $\alpha \neq \emptyset$, the *principal submatrix* $A_{\alpha\alpha}$ of A is the matrix obtained from A by removing all rows and columns in $S \setminus \alpha$. A principal minor of A is the determinant of a principal submatrix of A. An $n \times n$ matrix is a P-matrix [4] if all of its principal minors are positive. The importance of P-matrices for LCPs is captured by the following theorem.

Theorem 2 (Theorem 3.3.7, [4]). *A matrix $M \in \mathbb{R}^{n \times n}$ is a P-matrix if and only if the LCP(M, q) has a unique solution for every $q \in \mathbb{R}^n$.*

There are many algortithms for LCPs that work for P-matrices, but not in general. As stated by the following theorem, the matrices that arise from discounted games are P-matrices.

Theorem 3. *The matrix $M = (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}$ is a P-matrix.*

Proof. By Proposition 2, every LCP (M, q) arising from a discounted game has a unique solution. Given M, every $q \in \mathbb{R}^n$ can arise from a game (to see this, set $r_\lambda = 0$ in (12) and note that M is invertible), hence M is a P-matrix by Theorem 2. □

We give an alternative proof of Theorem 3 that does not rely on the fixed point theorem underlying Theorem 1 and Proposition 2. For this we recall the following two theorems from linear algebra. An $n \times n$ matrix A is *strictly row-diagonally dominant* if for every i, $1 \leq i \leq n$, we have $|A_{ii}| > \sum_{j \neq i} |A_{ij}|$.

Theorem 4 (Levy-Desplanques [9]). *Every strictly row-diagonally dominant square matrix is invertible.*

A *convex combination* of $n \times n$ matrices B and C is a matrix $QB + (I - Q)C$, where Q is a diagonal matrix with diagonal entries $q_1, q_2, \ldots, q_n \in [0, 1]$.

Theorem 5 (Johnson-Tsatsomeros [10]). *Let $A = BC^{-1}$, where B and C are square real matrices. Then A is a P-matrix iff every convex combination of B and C is invertible.*

Proof (alternative proof of Theorem 3). For every $\beta \in [0, 1)$, both $(\widehat{I} - \beta\widehat{T_\lambda})$ and $(\widehat{I} - \beta\widehat{T_\rho})$ are strictly row-diagonally dominant, and so is every convex combination of them. By Theorem 4, every such convex combination is invertible, and hence by Theorem 5, the matrix $M = (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}$ is a P-matrix. □

It is well-known that one-player discounted games, where $S = S_{\text{Min}}$ or $S = S_{\text{Max}}$, can be solved in polynomial time via a simple linear program [5]. We briefly note that in this case the matrix M is hidden-K, giving another proof that the LCP (M, q) is solvable via a linear program [13]. A matrix X is a Z-matrix if all off-diagonal entries are non-positive. A P-matrix M is hidden-K if and only if there exist Z-matrices X and Y such that $MX = Y$ and $Xe > 0$, where e is the all-one vector (see pg.212 of [4]). Without loss of generality, suppose $S = S_{\text{Max}}$, so $\widehat{I} = I$, $\widehat{T_\lambda} = T_\lambda$, and $\widehat{T_\rho} = T_\rho$. Then, by (11), we have $M(I - \beta T_\rho) = (I - \beta T_\lambda)$, which gives the hidden-K property.

3.3 Understanding q and M

For every $\sigma : S \to S$ with $\sigma(s) \in \{\lambda(s), \rho(s)\}$, let $\overline{v}^\sigma \in \mathbb{R}^n$ be the vector of discounted payoffs of σ-plays $\big\langle \mathcal{D}(\text{Play}(s, \sigma), \beta) \big\rangle_{s \in S}$. We define $r_\sigma \in \mathbb{R}^n$ as follows. For $s \in S$,

$$r_\sigma(s) = \begin{cases} r_\lambda(s) & \text{if } \sigma(s) = \lambda(s), \\ r_\rho(s) & \text{if } \sigma(s) = \rho(s). \end{cases}$$

Proposition 3. *For $\sigma : S \to S$, we have $\overline{v}^\sigma = (\widehat{I} - \beta\widehat{T_\sigma})^{-1}\widehat{I}r_\sigma$.*

Proof. The discounted payoff of the play $\text{Play}(s, \sigma)$ is the unique solution of the system of equations $v = r_\sigma + \beta T_\sigma v$, which is equivalent to $\widehat{I}v = \widehat{I}r_\sigma + \beta\widehat{T_\sigma}v$, and hence $\overline{v}^\sigma = (\widehat{I} - \beta\widehat{T_\sigma})^{-1}\widehat{I}r_\sigma$. □

Proposition 4. *If $q \in \mathbb{R}^n$ is as defined in (12), then $q = \widehat{I}\big(\overline{v}^\rho - (r_\lambda + \beta T_\lambda \overline{v}^\rho)\big)$.*

Proof. By Proposition 3, we have

$$q = M\widehat{I}r_\rho - \widehat{I}r_\lambda = (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}\widehat{I}r_\rho - \widehat{I}r_\lambda = (\widehat{I} - \beta\widehat{T_\lambda})\overline{v}^\rho - \widehat{I}r_\lambda\ . \qquad \square$$

For $\sigma : S \to S$, define the $n \times n$ matrix D_σ in the following way. For $s \in S$, let $\text{Play}(s, \sigma) = \langle s_0, s_1, \ldots, s_{k-1}, \langle t_0, t_1, \ldots, t_{\ell-1} \rangle^\omega \rangle$. Then for $t \in S$, we define

$$(D_\sigma)_{st} = \begin{cases} \beta^i & \text{if } t = s_i \text{ for some } i,\ 0 \le i < k, \\ \frac{\beta^{k+i}}{1-\beta^\ell} & \text{if } t = t_i \text{ for some } i,\ 0 \le i < \ell, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 5. *For $\sigma : S \to S$, we have $\overline{v}^\sigma = D_\sigma r_\sigma$.*

Proof. Let $s \in S$ and $\mathrm{Play}(s,\sigma) = \langle s_0, s_1, \ldots, s_{k-1}, \langle t_0, t_1, \ldots, t_{\ell-1}\rangle^\omega\rangle$. Then we have

$$\begin{aligned}
\mathcal{D}(\mathrm{Play}(s,\sigma),\beta) &= \sum_{i=0}^{k-1} \beta^i r_\sigma(s_i) + \beta^k \sum_{j=0}^{\infty}\sum_{i=0}^{\ell-1} \beta^{j\ell+i} r_\sigma(t_i) \\
&= \sum_{k-1} \beta^i r_\sigma(s_i) + \sum_{i=0}^{\ell-1} \big(\beta^{k+i} \sum_{j=0}^{\infty} \beta^{j\ell}\big) r_\sigma(t_i) \\
&= \sum_{k-1} \beta^i r_\sigma(s_i) + \sum_{i=0}^{\ell-1} \frac{\beta^{k+i}}{1-\beta^\ell} \cdot r_\sigma(t_i) \\
&= (D_\sigma r_\sigma)_s. \qquad \square
\end{aligned}$$

By Proposition 5, the discounted payoff of the play $\mathrm{Play}(s,\sigma)$ is equal to $(D_\sigma r_\sigma)_s = \sum_{t\in S} (D_\sigma)_{st} \cdot r_\sigma(t)$. Therefore, we can think of $(D_\sigma)_{st}$ as the coefficient of the contribution of the reward $r_\sigma(t)$ on the edge that leaves state $t \in S$, towards the total discounted payoff of the play, which is starting from state s, and that is following strategy σ onwards.

Lemma 1. *Let M be the $n \times n$ matrix as defined in (11). Then for every $s, t \in S$, we have $M_{st} = (-1)^{[s\in S_{Min}]+[t\in S_{Min}]}\big((D_\rho)_{st} - \beta(D_\rho)_{\lambda(s)t}\big)$.*

Proof. The following follows from Propositions 3 and 5, and from $\widehat{I}^{-1} = \widehat{I}$:

$$\begin{aligned}
M &= (\widehat{I} - \beta\widehat{T_\lambda})(\widehat{I} - \beta\widehat{T_\rho})^{-1}\widehat{I}\widehat{I}^{-1} \\
&= (\widehat{I} - \beta\widehat{T_\lambda}) D_\rho \widehat{I}.
\end{aligned}$$

Therefore, for all $s, t \in S$, we have

$$\begin{aligned}
M_{st} &= (-1)^{[s\in S_{\mathrm{Min}}]} \cdot (-1)^{[t\in S_{\mathrm{Min}}]} (D_\rho)_{st} - (-1)^{[s\in S_{\mathrm{Min}}]} \beta \cdot (-1)^{[t\in S_{\mathrm{Min}}]} (D_\rho)_{\lambda(s)t} \\
&= (-1)^{[s\in S_{\mathrm{Min}}]+[t\in S_{\mathrm{Min}}]}\big((D_\rho)_{st} - \beta(D_\rho)_{\lambda(s)t}\big). \qquad \square
\end{aligned}$$

4 Algorithms

4.1 Unique Sink Orientations of Cubes

A unique sink orientation (USO) of an n-dimensional hypercube is an orientation of its edges such that every face has a unique sink. The USO problem is to find the unique sink of the n-cube, using calls to an oracle that gives the orientation of edges adjacent to a vertex. For more details about USOs see [19].

For an LCP (M, q), the vector q is *nondegenerate* if it is not a linear combination of any $n-1$ columns of $(I, -M)$. Every P-matrix LCP (M, q) of dimension n with nondegenerate q corresponds to a USO $\psi(M, q)$ of the n-cube [16].

A *principal pivot transform* (PPT) of the LCP (M, q) is a related LCP with the role of w_i and z_i exchanged for all $i \in \alpha$ for some $\alpha \subseteq \{1, \ldots, n\}$. We denote by M_i the i-th column of M and by $e_i = I_i$ the i-th unit vector. For each $\alpha \subseteq \{1, \ldots, n\}$, define the $n \times n$ matrix B^α as,

$$(B^\alpha)_i = \begin{cases} -M_i, & \text{if } i \in \alpha \ , \\ e_i, & \text{if } i \notin \alpha \ . \end{cases}$$

The α-PPT of (M, q), written (M^α, q^α), is found as follows. Start with the matrix $A = [I, -M, q]$, which comes from the equation $Iw - Mz = q$, see (8). Obtain A' from A by exchanging I_i with $-M_i$ for all $i \in \alpha$. Then $(B^\alpha)^{-1}A' = [I, -M^\alpha, q^\alpha]$.

The vertices of $\psi(M, q)$ correspond to the subsets $\alpha \subseteq \{1, \ldots, n\}$. At vertex α, the n adjacent edges are oriented according to the sign of $q^\alpha = \big((B^\alpha)^{-1}q\big)$. For exactly one α, we have $q^\alpha \geq 0$, so that $z = 0$ is a trivial solution of the LCP (M^α, q^α); this is the sink of $\psi(M, q)$.

For a binary discounted game Γ, each subset $\alpha \subseteq \{1, \ldots, n\}$ corresponds to a choice of right-successor function, ρ^α, with

$$\rho^\alpha(s) = \begin{cases} \lambda(s) & \text{if } s \in \alpha, \\ \rho(s) & \text{if } s \notin \alpha, \end{cases}$$

for all $s \in S$. The sink of $\psi(M, q)$ is an α such that ρ^α is an optimal (combined) strategy.

4.2 Strategy Improvement and the Strategy Valuation USO

In this section we outline strategy improvement algorithms for solving discounted games. Such algorithms also exist for other classes of zero-sum games, such as parity games, mean-payoff games, and simple stochastic games [1,20]. For the *all-switching* variant of strategy improvement, no super-linear examples are known for any of these classes of games.

Underlying strategy improvement algorithms are corresponding USOs. For binary games, as considered here, these are USOs of cubes, for games with out-degree larger than two, USOs of grids; see [7].

Definition 1. *For a pair of strategies, a state $s \in S$ is* switchable *if the optimality equation s, given by* (1)*, does not hold. That is, for a right successor function ρ, used to denote a strategy pair, and a left successor function λ, used to denote the alternative choices to ρ, a state $s \in S_{Max}$ is switchable under strategy pair ρ if*

$$r_\lambda(s) + \beta \cdot \overline{v}^\rho(\lambda(s)) > r_\rho(s) + \beta \cdot \overline{v}^\rho(\rho(s)), \tag{13}$$

and state $s \in S_{Min}$ is switchable under ρ if

$$r_\lambda(s) + \beta \cdot \overline{v}^\rho(\lambda(s)) < r_\rho(s) + \beta \cdot \overline{v}^\rho(\rho(s)). \tag{14}$$

The rewards of the game are nondegenerate if there is no strategy pair σ such that for some state $s \in S$ we have $r_\lambda(s) + \beta \cdot \overline{v}^\sigma(\lambda(s)) = r_\rho(s) + \beta \cdot \overline{v}^\sigma(\rho(s))$. For the purpose of defining the strategy valuation USO $\tau(\Gamma)$, we only consider nondegenerate rewards. We associate a vertex of $\tau(\Gamma)$ with the strategy pair σ.

Definition 2. *For a game Γ, the strategy valuation USO $\tau(\Gamma)$ is defined as follows. At vertex σ an edge is outgoing if and only if the corresponding state is switchable.*

Proposition 6. *For a game Γ and the corresponding (M, q) defined by* (11)-(12)*, we have $\tau(\Gamma) = \psi(M, q)$.*

Proof. For $s \in S$, we have $q_s = (-1)^{[s \in S_{\mathrm{Min}}]}\big((r_\rho + \beta\overline{v}^\rho(\rho(s)) - (r_\lambda + \beta\overline{v}^\rho(\lambda(s))\big)$, by Proposition 6. Thus, if $s \in S_{\mathrm{Max}}$, then $q_s < 0$ if and only if (13) is satisfied, and if $s \in S_{\mathrm{Min}}$, then $q_s < 0$ if and only if (14) is satisfied. □

For a fixed strategy χ of Max, a *best response* of Min, $BR(\chi)$, is a strategy that for all $s \in S_{\mathrm{Min}}$ does not satisfy (14). For a state $s \in S$, there are two opposite facets, i.e., $(n-1)$-dimensional faces, of $\tau(\Gamma)$ such that in one all strategies are consistent with $\lambda(s)$, and in the other all are consistent with $\rho(s)$. Thus, the strategy χ of Max defines a subcube of $\tau(\Gamma)$ as the intersection of the facets that are consistent with χ. $BR(\chi)$ is the sink in this subcube.

Algorithm 1. [Strategy Improvement for Max]

Input: Discounted game Γ

repeat:

1. $\rho' \leftarrow \big(BR(\rho\restriction S_{\mathrm{Max}}), \rho\restriction S_{\mathrm{Max}}\big)$.
2. Obtain ρ'' from ρ' by switching at a nonempty subset of switchable $s \in S_{\mathrm{Max}}$ under ρ'.
3. $\rho \leftarrow \rho''$

until $\rho'' = \rho'$.

The proof of correctness of strategy improvement for simple stochastic games in Section 3.3 of [2] can be easily adapted to discounted games using the fact that, for every strategy σ, the matrix $D_\sigma = (I - \beta T_\sigma)^{-1}$ is nonnegative and has positive diagonal.

Algorithm 1 has the following interpretation in terms of $\tau(\Gamma)$. In Step 1., find the best response of Min as the sink in the subcube of $\tau(\Gamma)$ consistent with χ. In Step 2., from this sink, jump to the antipodal vertex in the subcube spanned by the chosen set of outgoing edges (switchable states). The algorithm can be seen as repeating Step 2. in the *strategy-improvement* USO $\tau_{\mathrm{Max}}(\Gamma)$, which is an *inherited* USO where the vertices correspond to the strategies of Max only. To obtain $\tau_{\mathrm{Max}}(\Gamma)$ from $\tau(\Gamma)$, we "drop" the dimensions corresponding to Min: at

vertex γ in $\tau_{\mathrm{Max}}(\Gamma)$, the orientation is consistent with that at $(BR(\gamma), \gamma)$ in $\tau(\Gamma)$, which is the sink in the subcube of $\tau(\Gamma)$ consistent with γ. For more details on inherited USOs, see Section 3 of [19]. It is a long-standing open question whether the all-switching variant of strategy improvement is polynomial.

With degnerate rewards, there is at least one edge that does not have a well-defined orientation. Strategy improvement still works, by considering any such edge as incoming to the current vertex.

4.3 Murty's Least-Index Method

In this section we outline Murty's least-index method for P-matrix LCPs. We show that, applied to the LCP (M, q) derived from a discounted game Γ according to (11) and (12), the least-index method can be considered as a strategy improvement algorithm.

Algorithm 2. [Murty's least-index method]

Input: LCP (M, q). Initialization: Set $\alpha := \emptyset$, $\bar{q} := q$.

while $\bar{q} \not\geq 0$ **do:**

$s \leftarrow \min_{\{1,\ldots,n\}}\{i \mid \bar{q}_i < 0\}$;
$\alpha \leftarrow \alpha \oplus \{s\}$;
$\bar{q} \leftarrow (B^{\alpha})^{-1} q$.

For a proof of the correctness of this Murty's least-index method, see [16]. Given Lemma 6, we see that, applied to the LCP derived from Γ, in each iteration Algorithm 2 makes a single switch in a switchable state with the lowest index.

Proposition 7. *Suppose Murty's least-index method is applied to the LCP arising from a discounted game Γ. If $S_{Max} = \{1, \ldots, k\}$ ($S_{Min} = \{1, \ldots, k\}$) for some $k \in \{1, \ldots, n\}$, then the algorithm corresponds to a single-switch variant of strategy improvement for Min (Max).*

Proof. Suppose $S_{\mathrm{Max}} = \{1, \ldots, k\}$. Then before any states of Min are switched, we have $q_1, \ldots, q_k \geq 0$, i.e., Max is playing a best response. Then, if possible, a single switch for Min with lowest index is made. □

Murty's least-index method gives a new algorithm for binary discounted games, and hence also for binary mean-payoff and parity games. For a given game, the method depends on an initial strategy pair, and an ordering of the states. As described by Proposition 7, for certain orderings of the states the method corresponds to a single-switch variant of strategy improvement in which the subroutine of computing best responses is also done via single-switch strategy improvement; for general orderings however it is a different algorithm.

5 Further Research

There are several algorithms for P-matrix LCPs that should be investigated in the context of discounted and simple stochastic games. For example, there is the Cottle-Dantzig prinicpal pivoting method [3] and Lemke's algorithm [12], which are pivoting methods. There are also interior point methods known for P-matrix LCPs [11].

The reduction from mean-payoff games to discounted games requires "large" discount factors. Can we design efficient algorithms for smaller discount factors? For small enough discount factor, the matrix M is close to the identity matrix and hence hidden-K, so the LCP can be solved as a linear program.

Whether all-switching strategy improvement is a polynomial-time algorithm is a long-standing open question. An exponential lower bound has been given for USOs in [18], but so far games that give rise to these example have not been constructed. What about upper bounds for strategy improvement for one-player discounted games? Are the inherited (strategy improvement) USOs, which we know to be acyclic, linearly inducible? Do they at least satisfy the Holt-Klee condition, which is known to hold for P-matrix LCPs, but is not necessarily preserved by inheritance [7]?

Acknowledgements. We thank Hugo Gimbert for stimulating us to formulate and study an LCP for solving discounted games.

References

1. Condon, A.: The complexity of stochastic games. Information and Computation 96, 203–224 (1992)
2. Condon, A.: On algorithms for simple stochastic games. In: Advances in Computational Complexity Theory, pp. 51–73. American Mathematical Society (1993)
3. Cottle, R.W., Dantzig, G.B.: Complementary pivot theory of mathematical programming. Linear Algebra and Its Applications 1, 103–125 (1968)
4. Cottle, R.W., Pang, J.-S., Stone, R.E.: The Linear Complementarity Problem. Academic Press (1992)
5. Derman, C.: Finite State Markov Decision Processes. Academic Press (1972)
6. Filar, J., Vrieze, K.: Competitive Markov Decision Processes. Springer, Heidelberg (1997)
7. Gärtner, B., Morris, W.D., Rüst, L.: Unique sink orientations of grids. In: Jünger, M., Kaibel, V. (eds.) IPCO 2005. LNCS, vol. 3509, pp. 210–224. Springer, Heidelberg (2005)
8. Gärtner, B., Rüst, L.: Simple stochastic games and P-matrix generalized linear complementarity problems. In: Liśkiewicz, M., Reischuk, R. (eds.) FCT 2005. LNCS, vol. 3623, pp. 209–220. Springer, Heidelberg (2005)
9. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press (1985)
10. Johnson, C.R., Tsatsomeros, M.J.: Convex sets of nonsingular and P-matrices. Linear and Multilinear Algebra 38, 233–239 (1995)
11. Kojima, M., Noma, T., Megiddo, N., Yoshise, A. (eds.): A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems. LNCS, vol. 538. Springer, Heidelberg (1991)

12. Lemke, C.E.: Bimatrix equilibrium points and mathematical programming. Management Science 11, 681–689 (1965)
13. Mangasarian, O.L.: Linear complementarity problems solvable by a single linear program. Mathematical Programming 10, 263–270 (1976)
14. Puri, A.: Theory of Hybrid Systems and Discrete Event Systems. PhD thesis, University of California, Berkeley (1995)
15. Shapley, L.S.: Stochastic games. Proc. Nat. Acad. Sci. U.S.A. 39, 1095–1100 (1953)
16. Stickney, A., Watson, L.: Digraph models of Bard-type algorithms for the linear complementarity problem. Mathematics of Operations Research 3, 322–333 (1978)
17. Svensson, O., Vorobyov, S.: Linear complementarity and P-matrices for stochastic games. In: Virbitskaite, I., Voronkov, A. (eds.) PSI 2006. LNCS, vol. 4378, pp. 408–421. Springer, Heidelberg (2007)
18. Szabó, T., Schurr, I.: Jumping doesn't help in abstract cubes. In: Integer Programming and Combinatorial Optimization (IPCO). LNCS, vol. 11, pp. 225–235. Springer, Heidelberg (2005)
19. Szabó, T., Welzl, E.: Unique sink orientations of cubes. In: IEEE Symposium on Foundations of Computer Science (FOCS), pp. 547–555 (2001)
20. Vöge, J., Jurdziński, M.: A discrete strategy improvement algorithm for solving parity games (Extended abstract). In: Emerson, E.A., Sistla, A.P. (eds.) CAV 2000. LNCS, vol. 1855, pp. 202–215. Springer, Heidelberg (2000)
21. Zwick, U., Paterson, M.: The complexity of mean payoff games on graphs. Theoretical Computer Science 158, 343–359 (1996)