A formula for the value of a stochastic game

Luc Attia[a,1] and Miquel Oliu-Barton[b,1,2]

[a]Centre de Mathématiques Appliquées, École Polytechnique, 91128 Palaiseau, France; and [b]Centre de Recherches en Mathématiques de la Décision, Université Paris-Dauphine, Paris Sciences et Lettres, 75016 Paris, France

Edited by Robert J. Aumann, The Hebrew University, Jerusalem, Israel, and approved November 12, 2019 (received for review May 21, 2019)

In 1953, Lloyd Shapley defined the model of stochastic games, which were the first general dynamic model of a game to be defined, and proved that competitive stochastic games have a discounted value. In 1982, Jean-François Mertens and Abraham Neyman proved that competitive stochastic games admit a robust solution concept, the value, which is equal to the limit of the discounted values as the discount rate goes to 0. Both contributions were published in PNAS. In the present paper, we provide a tractable formula for the value of competitive stochastic games.

stochastic games | repeated games | dynamic programming

1. Introduction

A. Motivation. Stochastic games are the first general model of dynamic games. Introduced by Shapley (1) in 1953, stochastic games extend the model of strategic-form games, which is due to von Neumann (2), to dynamic situations in which the environment (henceforth, the state) changes in response to the players' choices. They also extend the model of Markov decision problems to competitive situations with more than one decision maker.

Stochastic games proceed in stages. At each stage, the players choose actions which are available to them at the current state. Their choices have 2 effects: They generate a stage reward for each player, and they determine the probability for the state at the next stage. Consequently, the players are typically confronted with a trade-off between getting high rewards in the present and trying to reach states that will ensure high future rewards. Stochastic games and their applications have been studied in several scientific disciplines, including economics, operations research, evolutionary biology, and computer science. In addition, mathematical tools that were used and developed in the study of stochastic games are used by mathematicians and computer scientists in other fields. We refer the readers to Solan and Vieille (3) for a summary of the historical context and the impact of Shapley's seminal contribution.

The present paper deals with finite competitive stochastic games, that is, 2-player stochastic games with finitely many states and actions, and where the stage rewards of the players add up to zero. Shapley (1) proved that these games have a discounted value, which represents what playing the game is worth to the players when future rewards are discounted at a constant positive rate. Bewley and Kohlberg (4) proved that the discounted values admit a limit as the discount rate goes to 0. Building on this result, Mertens and Neyman (5, 6) proved that finite competitive stochastic games admit a robust solution concept, the value, which represents what playing the game is worth to the players when they are sufficiently patient.

Finding a tractable formula for the value of finite competitive stochastic games was a major open problem for nearly 40 y, which is settled in the present contribution. While opening an additional path for faster computations, our approach may also bring additional quantitative and qualitative insights into the model of stochastic games.

B. Outline of this Paper. This paper is organized as follows: *Section 2* states our results on finite competitive stochastic games, namely a formula for the λ-discounted values (proved in *Section 3*), and a formula for the value (proved in *Section 4*). *Section 5* describes the algorithmic implications and tractability of these 2 formulas. *Section 6* concludes with remarks and extensions.

2. Context and Main Results

To state our results precisely, we recall some definitions and well-known results about 2-player zero-sum games (*Section 2A*) and about finite competitive stochastic games (*Section 2B*). In *Section 2C* we give a brief overview of the relevant literature on finite competitive stochastic games. Our results are described in *Section 2D*.

Notation. Throughout this paper, $\mathbb{N}$ denotes the set of positive integers. For any finite set E we denote the set of probabilities over E by $\Delta(E) = \{f : E \to [0,1] \mid \sum_{e \in E} f(e) = 1\}$ and its cardinality by $|E|$.

A. Preliminaries on Zero-Sum Games. The aim of this section is to recall some well-known definitions and facts about 2-player zero-sum games, henceforth zero-sum games.

Definition. A zero-sum game is described by a triplet (S, T, ρ), where S and T are the sets of possible strategies for player 1 and player 2, respectively, and $\rho : S \times T \to \mathbb{R}$ is a payoff function. It is played as follows: Independently and simultaneously, the first player chooses $s \in S$ and the second player chooses $t \in T$. Player 1 receives $\rho(s,t)$ and player 2 receives $-\rho(s,t)$. The zero-sum game (S, T, ρ) has a value whenever

$$\sup_{s \in S} \inf_{t \in T} \rho(s,t) = \inf_{t \in T} \sup_{s \in S} \rho(s,t).$$

In this case, we denote this common quantity by val ρ.

Optimal strategies. Let (S, T, ρ) be a zero-sum game which has a value. An optimal strategy for player 1 is an element $s^* \in S$ so

Significance

Stochastic games were introduced by the Nobel Memorial Prize winner Lloyd Shapley in 1953 to model dynamic interactions in which the environment changes in response to the players' behavior. The theory of stochastic games and its applications have been studied in several scientific disciplines, including economics, operations research, evolutionary biology, and computer science. In addition, mathematical tools that were used and developed in the study of stochastic games are used by mathematicians and computer scientists in other fields. This paper contributes to the theory of stochastic games by providing a tractable formula for the value of finite competitive stochastic games. This result settles a major open problem which remained unsolved for nearly 40 y.

Author contributions: L.A. and M.O.-B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]L.A. and M.O.-B. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: miquel.oliu.barton@normalesup.org.

that $\rho(s^*, t) \geq \mathsf{val}\,\rho$ for all $t \in T$. Similarly, $t^* \in T$ is an optimal strategy for player 2 if $\rho(s, t^*) \leq \mathsf{val}\,\rho$ for all $s \in S$.

The value operator. The following properties are well known:

i) Minmax theorem: Let (S, T, ρ) be a zero-sum game. Suppose that S and T are 2 compact subsets of some topological vector space, ρ is a continuous function, the map $s \mapsto \rho(s, t)$ is concave for all $t \in T$, and the map $t \mapsto \rho(s, t)$ is convex for all $s \in S$. Then (S, T, ρ) has a value and both players have optimal strategies.

ii) Monotonicity: Suppose that (S, T, ρ) and (S, T, ν) have a value, and $\rho(s, t) \leq \nu(s, t)$ holds for all $(s, t) \in S \times T$. Then $\mathsf{val}\,\rho \leq \mathsf{val}\,\nu$.

Matrix games. In the sequel, we identify every real matrix $M = (m_{a,b})$ of size $p \times q$ with the zero-sum game (S_M, T_M, ρ_M), where $S_M = \Delta(\{1, \ldots, p\})$, $T_M = \Delta(\{1, \ldots, q\})$ and where

$$\rho_M(s, t) = \sum_{a=1}^{p} \sum_{b=1}^{q} s(a)\, m_{a,b}\, t(b) \qquad \forall (s, t) \in S_M \times T_M.$$

The value of the matrix M, denoted by $\mathsf{val}\,M$, is the value of (S_M, T_M, ρ_M) which exists by the minmax theorem. The following properties are well known:

iii) Continuity: Suppose that $M(t)$ is a matrix with entries that depend continuously on some parameter $t \in \mathbb{R}$. Then the map $t \mapsto \mathsf{val}\,M(t)$ is continuous.

iv) A formula for the value: For any matrix M, there exists a square submatrix $\hat{M}$ of M so that $\mathsf{val}\,M = \frac{\det \hat{M}}{\varphi(\hat{M})}$, where $\varphi(\hat{M})$ denotes the sum of all of the cofactors of $\hat{M}$, with the convention that $\varphi(\hat{M}) = 1$ if $\hat{M}$ is of size 1×1.

Comments. Property *i*) is taken from Sion (7), a generalization of von Neumann's (2) minmax theorem, while property *iv*) was established by Shapley and Snow (8). The other 2 properties are straightforward.

B. Stochastic Games. We present now the standard model of finite competitive stochastic games, henceforth stochastic games for simplicity. We refer the reader to Sorin's book (ref. 9, chap. 5) and to Renault's notes (10) for a more detailed presentation of stochastic games.

Definition. A stochastic game is described by a tuple (K, I, J, g, q, k), where $K = \{1, \ldots, n\}$ is a finite set of states, for some $n \in \mathbb{N}$; I and J are the finite action sets of player 1 and player 2, respectively; $g : K \times I \times J \to \mathbb{R}$ is a reward function to player 1; $q : K \times I \times J \to \Delta(K)$ is a transition function; and $1 \leq k \leq n$ is an initial state.

The game proceeds in stages as follows: At each stage $m \geq 1$, both players are informed of the current state $k_m \in K$, where $k_1 = k$. Then, independently and simultaneously, player 1 chooses an action $i_m \in I$ and player 2 chooses an action $j_m \in J$. The pair (i_m, j_m) is then observed by both players, from which they can infer the stage reward $g(k_m, i_m, j_m)$. A new state k_{m+1} is then chosen according to the probability distribution $q(k_m, i_m, j_m)$, and the game proceeds to stage $m + 1$.

Discounted stochastic games. For any discount rate $\lambda \in (0, 1]$, we denote by $(K, I, J, g, q, k, \lambda)$ the stochastic game (K, I, J, g, q, k) where player 1 maximizes, in expectation, the normalized λ-discounted sum of rewards

$$\sum\nolimits_{m \geq 1} \lambda(1 - \lambda)^{m-1} g(k_m, i_m, j_m),$$

while player 2 minimizes this amount.

In the following, the discount rate λ and the initial state k are considered as parameters, while (K, I, J, g, q) is fixed.

Strategies. A behavioral strategy, henceforth a strategy, is a decision rule from the set of possible observations of a player to the set of probabilities over the set of the player's actions. Formally, a strategy for player 1 is a sequence of mappings $\sigma = (\sigma_m)_{m \geq 1}$, where $\sigma_m : (K \times I \times J)^{m-1} \times K \to \Delta(I)$. Similarly, a strategy for player 2 is a sequence of mappings $\tau = (\tau_m)_{m \geq 1}$, where $\tau_m : (K \times I \times J)^{m-1} \times K \to \Delta(J)$. The sets of strategies are denoted, respectively, by Σ and $\mathcal{T}$.

The expected payoff. By the Kolmogorov extension theorem, together with an initial state k and the transition function q, any pair of strategies $(\sigma, \tau) \in \Sigma \times \mathcal{T}$ induces a unique probability $\mathsf{P}^k_{\sigma,\tau}$ over the sets of plays $(K \times I \times J)^{\mathbb{N}}$ on the sigma algebra generated by the cylinders. Hence, to any pair of strategies $(\sigma, \tau) \in \Sigma \times \mathcal{T}$ corresponds a unique payoff $\gamma^k_\lambda(\sigma, \tau)$ in the discounted game $(K, I, J, g, q, k, \lambda)$,

$$\gamma^k_\lambda(\sigma, \tau) := \mathbb{E}^k_{\sigma,\tau}\left[\sum\nolimits_{m \geq 1} \lambda(1 - \lambda)^{m-1} g(k_m, i_m, j_m)\right],$$

where $\mathbb{E}^k_{\sigma,\tau}$ denotes the expectation with respect to the probability $\mathsf{P}^k_{\sigma,\tau}$.

Stationary strategies. A stationary strategy is a strategy that depends only on the current state. Thus, $x : K \to \Delta(I)$ is a stationary strategy for player 1 while $y : K \to \Delta(J)$ is a stationary strategy for player 2. The sets of stationary strategies are $\Delta(I)^n$ and $\Delta(J)^n$, respectively. A pure stationary strategy is a stationary strategy that is deterministic. The sets of pure stationary strategies are I^n and J^n, respectively, and we refer to pure stationary strategies with the signs in boldface type, $\mathbf{i} \in I^n$ and $\mathbf{j} \in J^n$.

A useful expression. Suppose that both players use stationary strategies x and y in the discounted stochastic game $(K, I, J, g, q, k, \lambda)$ for some $\lambda \in (0, 1]$. The evolution of the state then follows a Markov chain, and the stage rewards depend only on the current state. Let $Q(x, y) \in \mathbb{R}^{n \times n}$ and $g(x, y) \in \mathbb{R}^n$ denote, respectively, the corresponding transition matrix and the vector of expected rewards. Formally, for all $1 \leq \ell, \ell' \leq n$,

$$Q^{\ell,\ell'}(x, y) = \sum_{(i,j) \in I \times J} x^\ell(i) y^\ell(j) q(\ell' \mid \ell, i, j) \qquad \textbf{[2.1]}$$

$$g^\ell(x, y) = \sum_{(i,j) \in I \times J} x^\ell(i) y^\ell(j) g(\ell, i, j). \qquad \textbf{[2.2]}$$

Let $\gamma_\lambda(x, y) = (\gamma^1_\lambda(x, y), \ldots, \gamma^n_\lambda(x, y)) \in \mathbb{R}^n$. Then $Q(x, y)$, $g(x, y)$, and $\gamma_\lambda(x, y)$ satisfy the relations

$$\begin{aligned}\gamma_\lambda(x, y) &= \sum\nolimits_{m \geq 1} \lambda(1 - \lambda)^{m-1} Q^{m-1}(x, y) g(x, y) \\ &= \lambda g(x, y) + (1 - \lambda) Q(x, y) \gamma_\lambda(x, y).\end{aligned}$$

Let Id denote the identity matrix of size n. The matrix $\mathrm{Id} - (1 - \lambda) Q(x, y)$ is invertible because $Q(x, y)$ is a stochastic matrix. Consequently, $\gamma_\lambda(x, y) = \lambda(\mathrm{Id} - (1 - \lambda) Q(x, y))^{-1} g(x, y)$ and, by Cramer's rule,

$$\gamma^k_\lambda(x, y) = \frac{d^k_\lambda(x, y)}{d^0_\lambda(x, y)}, \qquad \textbf{[2.3]}$$

where $d^0_\lambda(x, y) = \det(\mathrm{Id} - (1 - \lambda) Q(x, y))$ and where $d^k_\lambda(x, y)$ is the determinant of the $n \times n$ matrix obtained by replacing the kth column of $\mathrm{Id} - (1 - \lambda) Q(x, y)$ with $\lambda g(x, y)$.

The discounted values. The discounted stochastic game $(K, I, J, g, q, k, \lambda)$ and the zero-sum game $(\Sigma, \mathcal{T}, \gamma^k_\lambda)$ are equal by construction. Thus, the discounted stochastic game has a value whenever

$$\sup_{\sigma \in \Sigma} \inf_{\tau \in \mathcal{T}} \gamma^k_\lambda(\sigma, \tau) = \inf_{\tau \in \mathcal{T}} \sup_{\sigma \in \Sigma} \gamma^k_\lambda(\sigma, \tau).$$

In this case, the value is denoted by v_λ^k and is often referred to as the λ-discounted value of the stochastic game (K, I, J, g, q, k). The following result is due to Shapley (1):

v) Every discounted stochastic game $(K, I, J, g, q, k, \lambda)$ has a value, and both players have optimal stationary strategies. For each $1 \leq \ell \leq n$ and $u \in \mathbb{R}^n$, consider the following matrix of size $|I| \times |J|$:

$$\mathcal{G}_{\lambda,u}^{\ell} := \left(\lambda g(\ell, i, j) + (1-\lambda) \sum\nolimits_{\ell'=1}^{n} q(\ell' \mid \ell, i, j) u^{\ell'}\right)_{i,j}.$$

The vector of values $v_\lambda = (v_\lambda^1, \ldots, v_\lambda^n)$ is then the unique fixed point of the Shapley operator $\Phi(\lambda, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$, which is defined by $\Phi^\ell(\lambda, u) := \mathsf{val}\, \mathcal{G}_{\lambda,u}^{\ell}$, for all $1 \leq \ell \leq n$ and $u \in \mathbb{R}^n$.

Remark. In the model of stochastic games, the discount rate stands for the degree of impatience of the players, in the sense that future rewards are discounted. Alternatively, one can interpret λ as the probability that the game stops after every stage. The more general case of stopping probabilities that depend on the current state and on the players' actions can be handled in a similar way, as already noted by Shapley (1).

The value. The stochastic game (K, I, J, g, q, k) has a value if there exists $v^k \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists M_0 such that player 1 can guarantee that for any $M_0 \leq M \leq +\infty$ the expectation of the average reward per stage in the first M stages of the game is at least $v^k - \varepsilon$, and player 2 can guarantee that this amount is at most $v^k - \varepsilon$. It follows that if the game has a value v^k, then for each $\varepsilon > 0$ there exists a pair of strategies $(\sigma_\varepsilon, \tau_\varepsilon) \in \Sigma \times \mathcal{T}$ such that, for some $\lambda_0 \in (0, 1]$, the following inequalities hold for all $\lambda \in (0, \lambda_0)$:

$$\gamma_\lambda^k(\sigma_\varepsilon, \tau) \geq v^k - \varepsilon \qquad \forall \tau \in \mathcal{T}$$
$$\gamma_\lambda^k(\sigma, \tau_\varepsilon) \leq v^k + \varepsilon \qquad \forall \sigma \in \Sigma.$$

The following result is due to Mertens and Neyman (5):

vi) Every stochastic game (K, I, J, g, q, k) has a value v^k, and $v^k = \lim_{\lambda \to 0} v_\lambda^k$.

C. State of the Art. Since its introduction by Shapley (1), the theory of stochastic games and its applications have been studied in several scientific disciplines. We restrict our brief literature survey to the theory of finite competitive stochastic games and related algorithms.

The discounted values. In 1953, Shapley (1) proved that every discounted stochastic game $(K, I, J, g, q, k, \lambda)$ admits a value v_λ^k and that both players have optimal stationary strategies. Furthermore, the vector of values $v_\lambda = (v_\lambda^1, \ldots, v_\lambda^n)$ is the unique fixed point of an explicit operator.

Existence of the value. Building on Shapley's characterization of the discounted values and on a deep result from real algebraic geometry, the so-called Tarski–Seidenberg elimination theorem, Bewley and Kohlberg (4) proved in 1976 that the discounted values converge as the discount rate tends to zero. Mertens and Neyman (5, 6) strengthened this result in the early 1980s by establishing that every stochastic game (K, I, J, g, q, k) has a value v^k and that the value coincides with the limit of the discounted values. It is worth noting that, unlike discounted stochastic games, where the observation of the past actions is irrelevant, the existence of the value relies on the observation of the stage rewards.

Alternative proofs of convergence. In the late 1990s, Szczechla, Connell, Filar, and Vrieze (11) gave an alternative proof for the convergence of the discounted values as the discount rate goes to zero, using Shapley's characterization of the discounted values and the geometry of complex analytic varieties. Another proof was recently obtained by Oliu-Barton (12), based on the theory of finite Markov chains and on Motzkin's alternative theorem for linear systems.

Robustness of the value. The years 2010 to 2018 have brought many additional results concerning the value of stochastic games. Neyman and Sorin (13) studied stochastic games with a random duration clock. That is, at each stage, the players receive an additional signal which carries information about the number of remaining stages. Assuming that the expected number of remaining stages decreases throughout the game and that the expected number of stages converges to infinity, the values of the stochastic games with a random duration clock converge, and the limit is equal to the value of the stochastic game. Ziliotto (14) considered weighted-average stochastic games, that is, stochastic games where player 1 maximizes in expectation a fixed weighted average of the sequence of rewards, namely $\sum_{m \geq 1} \theta_m g(k_m, i_m, j_m)$. If $\sum_{m \geq 1} |\theta_{m+1}^p - \theta_m^p|$ converges to zero for some $p > 0$, then the values of the weighted-average stochastic games converge, and the limit is equal to the value of the stochastic game. Neyman (15) considered discounted stochastic games in continuous time and proved that their value coincides with the value of the discrete model. Finally, Oliu-Barton and Ziliotto (16) proved that stochastic games satisfy the constant payoff property, as conjectured by Sorin, Venel, and Vigeral (17). That is, for sufficiently small λ, any pair of optimal strategies of the discounted game $(K, I, J, g, q, k, \lambda)$ has the property that, in expectation, the average of the cumulated λ-discounted sum of rewards on any set of consecutive stages of cardinality of order $1/\lambda$ is approximately equal to v^k.

Characterization of the value. The first results on the value of stochastic games go back to the mid-1960s. By adapting the tools developed by Howard (18) for Markov decision problems, Hoffman and Karp (19) obtained a characterization for the limit of the λ-discounted values in the irreducible case (that is, when any pair of stationary strategies induces an irreducible Markov chain), in the spirit of an average cost optimality equation. Soon after, Blackwell and Ferguson (20) determined the value of the "Big Match," an example of a stochastic game whose value depends on the initial state. In the mid-1970s, Kolhberg (21) introduced absorbing games, a class of stochastic games in which there is at most one transition between states and which includes the Big Match as a particular case. Kohlberg proved that these games have a value and provided a characterization using the derivative of Shapley's operator. Two additional characterizations for the value of absorbing games were obtained recently by Laraki (22) and by Sorin and Vigeral (23), respectively.

Algorithms. Whether the value of a finite stochastic game can be computed in polynomial time is a famous open problem in computer science. This problem is intriguing because the class of simple stochastic games is both NP (nondeterministic polynomial time) and co-NP, and several important problems with this property have eventually been shown to be polynomial-time solvable, such as primality testing or linear programming. (A simple stochastic game is one where the transition function depends on one player's action at each state.) The known algorithms fall into 2 categories: decision procedures for the first-order theory of the reals, such as refs. 24–26, and value or strategy iteration methods, such as refs. 27 and 28. All of them are worst-case exponential in the number of states or in the number of actions. Recently, Hansen, Koucký, Lauritzen, Miltersen, and Tsigaridas (29) achieved a remarkable improvement by providing an algorithm which is polynomial in the number of actions, for any fixed number of states. However, the dependence on the number of states is both nonexplicit and doubly exponential. Based on the characterization of the value obtained in the present paper, Oliu-Barton (30) improved the algorithm of Hansen et al. (29) by significantly reducing the dependence on the number of states to an explicit polynomial dependence on the number of pure stationary strategies. Although not polynomial in the number of states, this algorithm is the most efficient algorithm that is known today.

D. Main Results. As already argued, the value is a very robust solution concept for stochastic games. Its existence was proved nearly 40 y ago, and an explicit characterization has been missing since then. The main contribution of the present paper is to provide a tractable formula for the value of stochastic games.

Our result relies on a different characterization of the discounted values, which is obtained by reducing a discounted stochastic game with n states to n independent parameterized matrix games, one for each initial state.

For the rest of this paper, $1 \leq k \leq n$ denotes a fixed initial state. The parameterized game that corresponds to k is simply obtained by linearizing the ratio in Eq. **2.3** for all pairs of pure stationary strategies, as follows:

Definition D.1. *For any $z \in \mathbb{R}$, define the matrix $W_\lambda^k(z)$ of size $|I|^n \times |J|^n$ by setting*

$$W_\lambda^k(z)[\mathbf{i},\mathbf{j}] := d_\lambda^k(\mathbf{i},\mathbf{j}) - z d_\lambda^0(\mathbf{i},\mathbf{j}) \qquad \forall (\mathbf{i},\mathbf{j}) \in I^n \times J^n.$$

Theorem 1 (A Formula for the Discounted Values). *For any $\lambda \in (0,1]$, the value of the discounted stochastic game $(K, I, J, g, q, k, \lambda)$ is the unique solution to*

$$z \in \mathbb{R}, \quad \mathsf{val}\ W_\lambda^k(z) = 0.$$

Theorem 2 (A Formula for the Value). *For any $z \in \mathbb{R}$, the limit $F^k(z) := \lim_{\lambda \to 0} \mathsf{val}\ W_\lambda^k(z)/\lambda^n$ exists in $\mathbb{R} \cup \{\pm\infty\}$. The value of the stochastic game (K, I, J, g, q, k) is the unique solution to*

$$w \in \mathbb{R}, \quad \begin{cases} z > w & \Rightarrow \quad F^k(z) < 0 \\ z < w & \Rightarrow \quad F^k(z) > 0. \end{cases}$$

Comments.

1) *Theorem 1* provides an uncoupled characterization of the discounted values. That is, each initial state is considered separately. This property, which contrasts with Shapley's (1) characterization, provides the key to *Theorem 2*.
2) *Theorem 1* can be extended to stochastic games with compact action spaces and continuous payoff and transition functions, but *Theorem 2* cannot because the discounted values may fail to converge in this case.
3) *Theorem 2* provides a different and elementary proof of the convergence of the λ-discounted values as λ tends to 0.
4) *Theorem 2* captures the characterization of the value for absorbing games obtained by Kohlberg (21).
5) The sign of $F^k(z)$ can be easily computed using linear programming techniques. This is a crucial aspect of the formula of *Theorem 2*.
6) *Theorems 1* and *2* suggest binary search algorithms for computing, respectively, the discounted values and the value, by successively evaluating the sign of $\mathsf{val}\ W_\lambda^k(z)$ and of $F^k(z)$ for well-chosen z. These algorithms are polynomial in the number of pure stationary strategies. The precise description and analysis of these algorithms are the object of a separate paper (30). For completeness, we provide a brief description in *Section 5*.

3. A Formula for the Discounted Values

In this section we prove *Theorem 1*. In the sequel, we consider a fixed discounted stochastic game $(K, I, J, g, q, k, \lambda)$. The proof is based on the following 4 properties:

1) $d_\lambda^0(\mathbf{i},\mathbf{j})$ is positive for all $(\mathbf{i},\mathbf{j}) \in I^n \times J^n$.
2) $(x, y, z) \mapsto d_\lambda^0(x, y) - z d_\lambda^k(x, y)$ is a multilinear map.
3) $z \mapsto \mathsf{val}\ W_\lambda^k(z)$ is a strictly decreasing real map.
4) $\mathsf{val}\ W_\lambda^k(v_\lambda^k) = 0$.

Indeed, *Theorem 1* clearly follows from the last two. The extension of this result to the more general framework of compact-continuous stochastic games (that is, stochastic games with compact metric action spaces and continuous payoff and transition functions) proceeds along the same lines and is postponed to *Section 6B*.

Notation. We use the following notation:

- For any $x = (x^1, \ldots, x^n) \in \Delta(I)^n$ we denote by $\hat{x} \in \Delta(I^n)$ the element that corresponds to the direct product of the coordinates of x. Formally,

$$\hat{x}(\mathbf{i}) := \prod_{\ell=1}^{n} x^\ell(\mathbf{i}^\ell) \qquad \forall\, \mathbf{i} = (\mathbf{i}^1, \ldots, \mathbf{i}^n) \in I^n.$$

  The map $x \mapsto \hat{x}$ is one to one and defines the canonical inclusion $\Delta(I)^n \subset \Delta(I^n)$. The map $y \mapsto \hat{y}$ is defined similarly and gives the canonical inclusion $\Delta(J)^n \subset \Delta(J^n)$.
- The letters **x** and **y** in boldface type refer to elements of $\Delta(I^n)$ and $\Delta(J^n)$, respectively.
- For all $z \in \mathbb{R}$ and all $(\mathbf{x},\mathbf{y}) \in \Delta(I^n) \times \Delta(J^n)$ we set

$$W_\lambda^k(z)[\mathbf{x},\mathbf{y}] := \sum_{(\mathbf{i},\mathbf{j}) \in I^n \times J^n} \mathbf{x}(\mathbf{i})\ W_\lambda^k(z)[\mathbf{i},\mathbf{j}]\, \mathbf{y}(\mathbf{j}).$$

We now prove the 4 properties above. The first one is due to Ostrovski (31), and for completeness we provide a short proof.

Lemma 1. *For any stochastic matrix P of size $n \times n$ and any $\lambda \in (0,1]$, $\det(\mathrm{Id} - (1-\lambda)P) \geq \lambda^n$.*

Proof: Set $M := \mathrm{Id} - (1-\lambda)P$. Because P is a stochastic matrix, $M^{\ell,\ell} - \sum_{\ell' \neq \ell} |M^{\ell,\ell'}| \geq \lambda$ for all $1 \leq \ell \leq n$. Hence, M is strictly diagonally dominant. For any $\mu \in \mathbb{R}$ so that $\mu < \lambda$, the matrix $M - \mu\mathrm{Id}$ is still strictly diagonally dominant, so in particular it is invertible. Consequently, all real eigenvalues of M are larger than or equal to λ. Similarly, for any $\mu = a + bi \in \mathbb{C}$ so that $|\mu| := \sqrt{a^2 + b^2} < \lambda$, the matrix $M - \mu\mathrm{Id}$ is strictly diagonally dominant, so that $M - \mu\mathrm{Id}$ is invertible. Consequently, if $a + bi$ is a complex eigenvalue of M, then $\lambda \leq |a + bi|$, so that $\lambda^2 \leq |a + bi|^2 = a^2 + b^2 = (a + bi)(a - bi)$. Recall that $\det M = \prod_{\ell=1}^{n} \mu_\ell$, where $\mu_1, \ldots, \mu_n$ are the eigenvalues of M counted with multiplicities. Because each real eigenvalue contributes at least λ in the product, and each pair of conjugate eigenvalues contributes at least λ^2, it clearly follows that $\det M \geq \lambda^n$. ■

Lemma 2. *For any $(x, \mathbf{j}) \in \Delta(I)^n \times J^n$ and $z \in \mathbb{R}$,*

i) $d_\lambda^0(x,\mathbf{j}) = \sum_{\mathbf{i} \in I^n} \widehat{x}(\mathbf{i}) d_\lambda^0(\mathbf{i},\mathbf{j})$.
ii) $d_\lambda^k(x,\mathbf{j}) = \sum_{\mathbf{i} \in I^n} \widehat{x}(\mathbf{i}) d_\lambda^k(\mathbf{i},\mathbf{j})$.
iii) $W_\lambda^k(z)[\hat{x},\mathbf{j}] = d_\lambda^k(x,\mathbf{j}) - z d_\lambda^0(x,\mathbf{j})$.

Proof:

i) Let $\mathbf{j} \in J^n$ be fixed. For any $x \in \Delta(I^n)$ set $M(x,\mathbf{j}) := \mathrm{Id} - (1-\lambda)Q(x,\mathbf{j})$, so that $\det M(x,\mathbf{j}) = d_\lambda^0(x,\mathbf{j})$ and, in particular, $\det M(\mathbf{i},\mathbf{j}) = d_\lambda^0(x,\mathbf{j})$ for all $\mathbf{i} \in I^n$. By Eq. **2.1**, the first row of $M(x,\mathbf{j})$ depends on x only through x^1, and the dependence is linear. Write x as a convex combination of the stationary strategies $\{(i,x^2,\ldots,x^n),\ i\in I\}$, and use the multilinearity of the determinant to obtain

$$\det M(x,\mathbf{j}) = \det\left(\sum\nolimits_{i\in I} x^1(i) M((i,x^2,\ldots,x^n),\mathbf{j})\right) = \sum\nolimits_{i\in I} x^1(i) \det M((i,x^2,\ldots,x^n),\mathbf{j}).$$

Using the same argument for the remaining rows, one inductively obtains that $\det M(x,\mathbf{j})$ is equal to

$$\sum_{\mathbf{i}^1\in I} x^1(\mathbf{i}^1) \sum_{\mathbf{i}^2\in I} x^2(\mathbf{i}^2)\cdots \sum_{\mathbf{i}^n\in I} x^n(\mathbf{i}^n) \det M((\mathbf{i}^1,\mathbf{i}^2,\ldots,\mathbf{i}^n),\mathbf{j}),$$

which is equal to $\sum_{\mathbf{i}\in I^n} \widehat{x}(\mathbf{i}) \det M(\mathbf{i},\mathbf{j})$ by the definition of $\hat{x}$.

ii) The proof goes along the same lines as *i*). Fix $\mathbf{j}\in J^n$. For any $x\in\Delta(J)^n$, let $M^k(x,\mathbf{j})$ be the matrix obtained by replacing the kth column of $M(x,\mathbf{j})$ by $\lambda g(x,\mathbf{j})$, so that $\det M^k(x,\mathbf{j}) = d_\lambda^k(x,\mathbf{j})$ and, in particular, $\det M^k(\mathbf{i},\mathbf{j}) = d_\lambda^k(\mathbf{i},\mathbf{j})$ for all $\mathbf{i}\in I^n$. By Eqs. **2.1** and **2.2**, the ℓth row of $M^k(x,\mathbf{j})$ depends on x only through x^ℓ and that the dependence is linear. Like in *i*), these properties imply the desired result, namely $\det M^k(x,\mathbf{j}) = \sum_{\mathbf{i}\in I^n} \widehat{x}(\mathbf{i}) \det M^k(\mathbf{i},\mathbf{j})$.

iii) The result follows directly from *i*), *ii*), and the definition of $W_\lambda^k(z)[\widehat{x},\mathbf{j}]$. Indeed,

$$\begin{aligned} W_\lambda^k(z)[\widehat{x},\mathbf{j}] &= \sum\nolimits_{\mathbf{i}\in I^n} \widehat{x}(\mathbf{i})\, W_\lambda^k(z)[\mathbf{i},\mathbf{j}] \\ &= \sum\nolimits_{\mathbf{i}\in I^n} \widehat{x}(\mathbf{i})\, d_\lambda^k(\mathbf{i},\mathbf{j}) - z\sum\nolimits_{\mathbf{i}\in I^n} \widehat{x}(\mathbf{i})\, d_\lambda^0(\mathbf{i},\mathbf{j}) \\ &= d_\lambda^k(x,\mathbf{j}) - z d_\lambda^0(x,\mathbf{j}). \end{aligned}$$

■

Remark: *Lemma 2* is stated for all $(x,\mathbf{j})$ for convenience, but is also valid for all (x,y). The last property, for instance, can be stated as follows. For all $(x,y,z)\in\Delta(I)^n\times\Delta(J)^n\times\mathbb{R}$,

$$W_\lambda^k(z)[\hat{x},\hat{y}] = d_\lambda^k(x,y) - z d_\lambda^0(x,y).$$

Lemma 3. *For any* $(z_1,z_2)\in\mathbb{R}^2$ *so that* $z_1<z_2$,

$$\mathsf{val}\ W_\lambda^k(z_1) - \mathsf{val}\ W_\lambda^k(z_2) \geq (z_2-z_1)\lambda^n.$$

In particular, $z\mapsto \mathsf{val}\ W_\lambda^k(z)$ *is a strictly decreasing real map.*

Proof: By definition, $Q(\mathbf{i},\mathbf{j})$ is a stochastic matrix of size $n\times n$ for each $(\mathbf{i},\mathbf{j})\in I^n\times J^n$. Hence, by *Lemma 1*,

$$d_\lambda^0(\mathbf{i},\mathbf{j}) = \det(\mathrm{Id}-(1-\lambda)Q(\mathbf{i},\mathbf{j})) \geq \lambda^n \qquad \forall(\mathbf{i},\mathbf{j})\in I^n\times J^n.$$

Therefore, for all $z_1<z_2$ and $(\mathbf{i},\mathbf{j})$,

$$\begin{aligned} W_\lambda^k(z_1)[\mathbf{i},\mathbf{j}] - W_\lambda^k(z_2)[\mathbf{i},\mathbf{j}] &= (z_2-z_1)d_\lambda^0(\mathbf{i},\mathbf{j}) \\ &\geq (z_2-z_1)\lambda^n. \end{aligned}$$

The result follows then from the monotonicity of the value operator, stated in item *ii*) of *Section 2A*. ■

Lemma 4. $\mathsf{val}\ W_\lambda^k(v_\lambda^k)=0$.

Proof: By Lemma 2, *iii*), the relation

$$W_\lambda^k(v_\lambda^k)[\widehat{x},\mathbf{j}] = d_\lambda^k(x,\mathbf{j}) - v_\lambda^k d_\lambda^0(x,\mathbf{j}) \qquad \textbf{[3.1]}$$

holds for all $(x,\mathbf{j})\in\Delta(I)^n\times J^n$. Let $x^*\in\Delta(I)^n$ be an optimal stationary strategy of player 1 in (K,I,J,g,q,k,λ), which exists by Shapley (1) as already noted in item *v*) of *Section 2B*, and let $\widehat{x}^*\in\Delta(I^n)$ denote the direct product of its coordinates. The optimality of x^* implies

$$\gamma_\lambda^k(x^*,\mathbf{j}) = \frac{d_\lambda^k(x^*,\mathbf{j})}{d_\lambda^0(x^*,\mathbf{j})} \geq v_\lambda^k.$$

The matrix $Q(x^*,\mathbf{j})$ is stochastic of size $n\times n$ so that $d_\lambda^0(x^*,\mathbf{j}) = \det(\mathrm{Id}-(1-\lambda)Q(x^*,\mathbf{j}))\geq\lambda^n>0$ by *Lemma 1*. Consequently, the previous relation is equivalent to

$$d_\lambda^k(x^*,\mathbf{j}) - v_\lambda^k d_\lambda^0(x^*,\mathbf{j}) \geq 0. \qquad \textbf{[3.2]}$$

Therefore, $W_\lambda^k(v_\lambda^k)[\widehat{x}^*,\mathbf{j}]\geq 0$ follows from Eqs. **3.1** and **3.2**. For any matrix $M=(m_{a,b})$ of size $p\times q$ and any $s\in\Delta(\{1,\ldots,p\})$, the definition of the value implies that $\mathsf{val}\,M\geq \min_{1\leq b\leq q}\sum_{1\leq a\leq p} s(a)m_{a,b}$. Consequently,

$$\mathsf{val}\ W_\lambda^k(v_\lambda^k) \geq \min_{\mathbf{j}\in J^n} W_\lambda^k(v_\lambda^k)[\widehat{x}^*,\mathbf{j}] \geq 0.$$

By reversing the roles of the players one similarly obtains an analogue of *Lemma 2* for all $(\mathbf{i},y)\in I^n\times\Delta(J)^n$, and then $\mathsf{val}\ W_\lambda^k(v_\lambda^k)\leq 0$, which gives the desired result. ■

Proof of Theorem 1: By *Lemma 3*, $z\mapsto\mathsf{val}\ W_\lambda^k(z)$ is a strictly decreasing real function. Consequently, the set $\{z\in\mathbb{R},\ \mathsf{val}\ W_\lambda^k(z)=0\}$ contains at most one element. By *Lemma 4*, this element is precisely v_λ^k. ■

4. A Formula for the Value

In this section we prove *Theorem 2*. Before we establish this result, we show that the limit $F^k(z):=\lim_{\lambda\to 0}\mathsf{val}\ W_\lambda^k(z)/\lambda^n$ exists in $\mathbb{R}\cup\{-\infty,+\infty\}$ for all $z\in\mathbb{R}$ and that the equation

$$w\in\mathbb{R},\quad \begin{cases} z>w & \Rightarrow\quad F^k(z)<0 \\ z<w & \Rightarrow\quad F^k(z)>0 \end{cases} \qquad \textbf{[4.1]}$$

admits a unique solution. This is shown in the following 2 lemmas.

Lemma 1. *Let* $z\in\mathbb{R}$. *Then, there exists a rational fraction* R *and* $\lambda_0>0$ *so that*

$$\mathsf{val}\ W_\lambda^k(z) = R(\lambda) \qquad \forall\lambda\in(0,\lambda_0).$$

Proof: By construction, the entries of $W_\lambda^k(z)$ are polynomials in λ. By Shapley and Snow (8), the value of a matrix satisfies the formula stated in item *iv*) of *Section 2A*. Consequently, for any $\lambda\in(0,1]$, there exists a rational fraction R so that $\mathsf{val}\ W_\lambda^k(z)=R(\lambda)$. Because the choice of the square submatrix may vary with λ, the corresponding rational fraction may also vary. However, as the number of possible square submatrices is finite, so is the number of possible rational fractions that may satisfy this equality. Consequently, there exists a finite collection $E=\{R_1,\ldots,R_L\}$ of rational fractions so that for each $\lambda\in(0,1]$ there exists $R\in E$ that satisfies $\mathsf{val}\ W_\lambda^k(z)=R(\lambda)$. Hence, for any λ, the point $(\lambda,\mathsf{val}\ W_\lambda^k(z))$ belongs to the union of the graphs of the functions $R_1,\ldots,R_L$. As already noted in item *iii*) of *Section 2A*, the map $\lambda\mapsto\mathsf{val}\ W_\lambda^k(z)$ is continuous on $(0,1]$. Consequently,

ECONOMIC SCIENCES

APPLIED MATHEMATICS

as λ varies on the interval $(0,1]$, the curve $\lambda \mapsto (\lambda, \text{val } W_\lambda^k(z))$ can "jump" from the graph of R to the graph of R' only at points where these 2 graphs intersect. Yet, for any 2 rational fractions, either they are congruent or they intersect finitely many times. Hence, there exists λ_0 so that, for any $R, R' \in E$, either $R(\lambda) = R'(\lambda)$ for all $(0, \lambda_0)$ or $R(\lambda) \neq R'(\lambda)$ for all $(0, \lambda_0)$. In particular, there exists $R \in E$ so that val $W_\lambda^k(z) = R(\lambda)$ for all $(0, \lambda_0)$. ■

Lemma 2. *Eq.* **4.1** *admits a unique solution*.

Proof: By Lemma 1, $\lim_{\lambda\to 0} \text{val } W_\lambda^k(z)/\lambda^n$ exists for all $z \in \mathbb{R}$. Suppose that Eq. **4.1** admits 2 solutions $w < w'$. Then, for any $z \in (w, w')$ one has $F^k(z) < 0$ and $F^k(z) > 0$, which is impossible. Therefore, Eq. **4.1** admits at most one solution. Let $(z_1, z_2) \in \mathbb{R}^2$ satisfy $z_1 < z_2$. Rearranging the terms in *Lemma 3*, dividing by λ^n, and taking λ to 0 yields

$$F^k(z_1) \geq F^k(z_2) + z_2 - z_1. \qquad \textbf{[4.2]}$$

In particular, the following relations hold:

$$\begin{cases} F^k(z) \geq 0 \Rightarrow F^k(z') \geq 0, \ \forall z' \leq z \\ F^k(z) \leq 0 \Rightarrow F^k(z') \leq 0, \ \forall z' \geq z \\ F^k(z) = 0 \Rightarrow F^k(z') \neq 0, \ \forall z' \neq z. \end{cases} \qquad \textbf{[4.3]}$$

We now show that F^k is not constant, which is still compatible with Eq. **4.2** if $F^k \equiv +\infty$ or $F^k \equiv -\infty$. Let $C^- := \min_{k,i,j} g(k,i,j)$ and $C^+ := \max_{k,i,j} g(k,i,j)$. For any $\lambda \in (0,1]$, one clearly has $C^- \leq v_\lambda^k \leq C^+$. Consequently, by *Lemma 3*,

$$\text{val } W_\lambda^k(C^+) \leq \text{val } W_\lambda^k(v_\lambda^k) \leq \text{val } W_\lambda^k(C^-).$$

Dividing by λ^n and taking λ to 0, one obtains

$$F^k(C^+) \leq 0 \leq F^k(C^-). \qquad \textbf{[4.4]}$$

We now define recursively 2 real sequences $(u_m^-)_{m\geq 1}$ and $(u_m^+)_{m\geq 1}$ by setting $u_1^- := C^-$, $u_1^+ := C^+$, and, for all $m \geq 1$,

$$u_{m+1}^- := \begin{cases} \frac{1}{2}(u_m^- + u_m^+) & \text{if } F^k\left(\frac{1}{2}(u_m^- + u_m^+)\right) \geq 0 \\ u_m^- & \text{otherwise,} \end{cases}$$

$$u_{m+1}^+ := \begin{cases} \frac{1}{2}(u_m^- + u_m^+) & \text{if } F^k\left(\frac{1}{2}(u_m^- + u_m^+)\right) \leq 0 \\ u_m^+ & \text{otherwise.} \end{cases}$$

By construction, $F^k(u_m^-) \geq 0$ and $F^k(u_m^+) \leq 0$ for all $m \geq 1$. Moreover, Eqs. **4.3** and **4.4** imply $C^- \leq u_m^- \leq u_m^+ \leq C^+$ for all $m \geq 1$, so that $(u_m^-)_m$ is nondecreasing and $(u_m^+)_m$ is nonincreasing. Furthermore, $u_{m+1}^+ - u_{m+1}^- \leq \frac{1}{2}(u_m^+ - u_m^-)$ for all $m \geq 1$. Hence, the 2 sequences admit a common limit $\bar{u}$. For any $\varepsilon > 0$, let m_ε be such that $u_{m_\varepsilon}^- > \bar{u} - \varepsilon$. By Eq. **4.2**, this implies

$$F^k(\bar{u} - \varepsilon) \geq F^k(u_{m_\varepsilon}^-) + u_{m_\varepsilon}^- - (\bar{u} - \varepsilon) > 0.$$

Similarly, $F^k(\bar{u} + \varepsilon) < 0$ for any $\varepsilon > 0$. Together with Eq. **4.3**, this shows that $\bar{u}$ is a solution to Eq. **4.1**. ■

We are now ready to prove our main result.

Proof of Theorem 2: Let w be the unique solution Eq. **4.1** and fix $\varepsilon > 0$. By the choice of w, $F^k(w - \varepsilon) > 0$. Consequently, there exists $\lambda_0 > 0$ so that

$$\text{val } W_\lambda^k(w - \varepsilon) > 0 \qquad \forall \lambda \in (0, \lambda_0). \qquad \textbf{[4.5]}$$

By *Lemma 3*, the map $z \mapsto \text{val } W_\lambda^k(z)$ is strictly decreasing. By *Lemma 4*, val $W_\lambda^k(v_\lambda^k) = 0$. Therefore, Eq. **4.5** implies

$$v_\lambda^k > w - \varepsilon \qquad \forall \lambda \in (0, \lambda_0). \qquad \textbf{[4.6]}$$

Because ε is arbitrary, $\liminf_{\lambda\to 0} v_\lambda^k \geq w$. By reversing the roles of the players, one obtains in a similar manner $\limsup_{\lambda\to 0} v_\lambda^k \leq w$. Hence, the λ-discounted values converge as λ goes to 0, and $\lim_{\lambda\to 0} v_\lambda^k = w$. The result follows then from item *vi*) of *Section 2B*, namely the existence of the value v^k and the equality $\lim_{\lambda\to 0} v_\lambda^k = v^k$, due to Mertens and Neyman (5). ■

5. Algorithms

The formulas obtained in *Theorems 1* and *2* suggest binary search methods for approximating the λ-discounted values and the value of a stochastic game (K, I, J, g, q, k), based on the evaluation of the sign of the real functions $z \mapsto \text{val } W_\lambda^k(z)$ and $z \mapsto F^k(z)$, respectively. In this section we provide a brief description of these algorithms and discuss their complexity using the logarithmic cost model (a model which accounts for the total number of bits which are involved). We refer the reader to ref. 30 for more technical details and for 2 additional algorithms which provide exact expressions for v_λ^k and v^k within the same complexity class.

Notation. For any $m \in \mathbb{N}$, let $E_m := \{0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m}{m}\}$ and $Z_m := \{0, \frac{1}{2^m}, \frac{2}{2^m}, \ldots, \frac{2^m}{2^m}\}$.

A. Computing the Discounted Values. The following bisection algorithm, which is directly derived from *Theorem 1*, inputs a discounted stochastic game with rational data and outputs an arbitrarily close approximation of its value.
Input. A discounted stochastic game $(K, I, J, g, q, k, \lambda)$ so that, for some $(N, L) \in \mathbb{N}^2$, the functions g and q take values in E_N and $\lambda \in E_L$ and a precision level $r \in \mathbb{N}$.
Output. A 2^{-r} approximation of v_λ^k.
Complexity. Polynomial in n, $|I|^n$, $|J|^n$, $\log N$, $\log L$ and r.

1) Set $\underline{w} := 0$, $\overline{w} := 1$
2) WHILE $\overline{w} - \underline{w} > 2^{-r}$ DO
   2.1) $z := \frac{\underline{w} + \overline{w}}{2}$
   2.2) $v :=$ sign of val $W_\lambda^k(z)$
   2.3) IF $v \geq 0$ THEN $\underline{w} := z$
   2.4) IF $v \leq 0$ THEN $\overline{w} := z$
3) RETURN $u := \underline{w}$.

Clearly, the output u satisfies $|u - v_\lambda^k| \leq 2^{-r}$, and the number of iterations in step 2, the "while" loop, is bounded by r. Also, the complexity of each iteration depends crucially on the complexity of step 2.2. First of all, one needs to determine the matrix $W_\lambda^k(z)$ for some $z \in Z_r$, and this requires the computation of 2 $n \times n$ determinants for each of its $|I|^n \times |J|^n$ entries. Algorithms for computing the determinant of a matrix exist which are polynomial in its size and in the number of bits that which are needed to encode this matrix. Second, the choice of z and Hadamard's inequality imply that the number of bits which are needed to encode $W_\lambda^k(z)$ is polynomial in n, $|I|^n$, $|J|^n$, $\log N$ and $\log L$, and r. Third, computing the value of a matrix can be done with linear programming techniques, and algorithms exist [for example, Karmarkar (32)] which are polynomial in its size and in the number of bits which are needed to encode this matrix. Consequently, the complexity of step 2.2 is polynomial in n, $|I|^n$, $|J|^n$, $\log N$ and $\log L$, and r, and the same is true for the entire algorithm.

B. Computing the Value. The following bisection algorithm, which is directly derived from *Theorem 2*, inputs a stochastic game with rational data and outputs an arbitrarily close approximation of its value.

Input. A stochastic game (K, I, J, g, q, k) so that, for some $N \in \mathbb{N}$, the functions g and q take values in E_N and a precision level $r \in \mathbb{N}$.

Output. A 2^{-r} approximation of v^k.

Complexity. Polynomial in n, $|I|^n$, $|J|^n$, $\log N$ and r.

1) Set $\underline{w} := 0$, $\overline{w} := 1$
2) WHILE $\overline{w} - \underline{w} > 2^{-r}$ DO
   2.1) $z := \frac{\underline{w}+\overline{w}}{2}$
   2.2) $v :=$ sign of $F^k(z)$
   2.3) IF $v \geq 0$ THEN $\underline{w} := z$
   2.4) IF $v \leq 0$ THEN $\overline{w} := z$
3) RETURN $u := \underline{w}$.

Like before, the output u satisfies $|u - v^k| \leq 2^{-r}$, the number of iterations in step 2 (the "while" loop) is bounded by r, and the variable z always takes values in the set Z_r. Unlike before, however, each iteration requires computing the sign of $F^k(z)$ at step 2.2, a computation that might seem problematic due to the limiting nature of the function F^k. However, this difficulty is overcome with the help of proposition 4.1 of ref. 30: "For any $r \in \mathbb{N}$, let $\lambda_r := N^{-10n^2|I|^{2n}r}$. Then, the sign of $F^k(z)$ is equal to the sign of val $W^k_{\lambda_r}(z)$ for all $z \in Z_r$." Consequently, the computation in step 2.2 can be replaced with the computation of val $W^k_{\lambda_r}(z)$. By the choice of λ_r and z, the number of bits which are needed to encode $W^k_{\lambda_r}(z)$ is polynomial in n, $|I|^n$, $|J|^n$, $\log_N$, and r. Hence, the computation of step 2.2 is polynomial in these variables, and the same is true for the entire algorithm.

6. Remarks and Extensions

First, we provide an alternative definition of the parameterized games $W^k_\lambda(z)$. Second, we extend *Theorem 1* to the more general framework of stochastic games with compact metric action sets and continuous payoff and transition function and explain why the extension of *Theorem 2* fails. Finally, we show that the formula obtained by Kohlberg (21) for the value of absorbing games is captured by *Theorem 2*.

A. An Alternative Formulation of the Parameterized Games. The parameterized game $W^k_\lambda(z)$ plays a crucial role in both *Theorems 1* and *2*. We provide an alternative construction of this game which is based on the Kronecker product of matrices. Let U denote a matrix of ones of size $|I| \times |J|$. For each $1 \leq \ell, \ell' \leq n$, consider the matrices $Q^{\ell,\ell'} = (q(\ell' \mid \ell, i, j))_{i,j}$ and $G^\ell = (g(\ell, i, j))_{i,j}$ and use them to form the following $n \times (n+1)$ array of matrices of size $|I| \times |J|$:

$$D_\lambda = \begin{pmatrix} -\lambda G^1 & U - (1-\lambda)Q^{1,1} & \dots & -(1-\lambda)Q^{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ -\lambda G^n & -(1-\lambda)Q^{n,1} & \dots & U - (1-\lambda)Q^{n,n} \end{pmatrix}.$$

For any $0 \leq \ell \leq n$, let D^ℓ_λ be the $n \times n$ array of matrices obtained by removing the $(\ell+1)$ th column of matrices from D. Denote by $\det_\otimes$ the determinant of a square array of matrices, developed along columns and where the products are replaced with the Kronecker product of matrices. By construction one has $\det_\otimes D^0_\lambda = (d^0_\lambda(\mathbf{i},\mathbf{j}))_{\mathbf{i},\mathbf{j}}$ and $(-1)^k \det_\otimes D^k_\lambda = (d^k_\lambda(\mathbf{i},\mathbf{j}))_{\mathbf{i},\mathbf{j}}$, so that

$$W^k_\lambda(z) = (-1)^k \det_\otimes D^k_\lambda - z \det_\otimes D^0_\lambda.$$

The linearity relations established in *Lemma 2* can also be deduced from the properties of the Kronecker product. This alternative expression for $W^k_\lambda(z)$ is reminiscent of (or, rather, inspired by) the theory of multiparameter eigenvalue problems initiated by Atkinson in the 1960s (ref. 33, chap. 6). The interesting connection which exists between stochastic games and multiparameter eigenvalue problems is developed by L.A. and M.O.-B. in a forthcoming paper (34).

B. Compact-Continuous Stochastic Games. Throughout this section we consider stochastic games (K, I, J, g, q), where $K = \{1, \dots, n\}$ is a finite set of states, I and J are 2 compact metric sets, and g and q are continuous functions. These games are referred to as compact-continuous stochastic games, for short. We denote by $\Delta(I)$ and $\Delta(J)$, respectively, the sets of probability distributions over I and J. These sets are compact when endowed with the weak* topology. For any pair of measures $(\alpha, \beta) \in \Delta(I) \times \Delta(J)$, their direct product is denoted by $\alpha \otimes \beta \in \Delta(I \times J)$. For all $1 \leq \ell, \ell' \leq n$ and $u \in \mathbb{R}^n$, we set

$$g(\ell, \alpha, \beta) := \int_{I \times J} g(\ell, i, j)\, d(\alpha \otimes \beta)(i, j)$$
$$q(\ell' \mid \ell, \alpha, \beta) := \int_{I \times J} q(\ell' \mid \ell, i, j)\, d(\alpha \otimes \beta)(i, j),$$
$$\rho^\ell_{\lambda,u}(\alpha, \beta) := \lambda g(\ell, \alpha, \beta) + (1-\lambda) \sum_{\ell'=1}^n q(\ell' \mid \ell, \alpha, \beta).$$

By the minmax theorem stated in item *i*) of *Section 2A*, the zero-sum game $(\Delta(I), \Delta(J), \rho^\ell_{\lambda,u})$ has a value, so one can define the Shapley operator $\Phi(\lambda, \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ like in the finite case. Furthermore, the compact-continuous stochastic game $(K, I, J, g, q, k, \lambda)$ has a value v^k_λ, which is the unique fixed point of $\Phi(\lambda, \cdot)$, and both players have optimal stationary strategies. These results are well known.

Extension of Theorem 1. *Theorem 1* can be extended to compact-continuous stochastic games.

The proof goes along the same lines. Like in the finite case, any pair of stationary strategies $(x, y) \in \Delta(I)^n \times \Delta(J)^n$ induces a Markov chain with state-dependent rewards. Let $Q(x, y) \in \mathbb{R}^{n \times n}$ and $g(x, y) \in \mathbb{R}^n$ denote the transition matrix of this chain and the vector of expected rewards. Formally, they are defined like in Eqs. **2.1** and **2.2**, but replacing, for $1 \leq \ell, \ell' \leq n$, the sum $\sum_{(i,j) \in I \times J} x^\ell(i) y^\ell(j)$ with the corresponding integral $\int_{I \times J} d(x^\ell \otimes y^\ell)(i, j)$. Similarly, let $\gamma_\lambda(x, y) \in \mathbb{R}^n$ be the vector of expected normalized λ-discounted sum of rewards, which is well defined because the state k, the pair (x, y), and the transition function q induce a unique probability measure over $(K \times I \times J)^{\mathbb{N}}$ on the sigma algebra generated by the cylinders, by the Kolmogorov extension theorem. Like in the finite case,

$$\gamma^k_\lambda(x, y) = \frac{d^k_\lambda(x, y)}{d^0_\lambda(x, y)},$$

where $d^0_\lambda(x, y) := \det(\mathrm{Id} - (1-\lambda)Q(x, y)) \neq 0$ and where $d^k_\lambda(x, y)$ is the determinant of the $n \times n$ matrix obtained by replacing the kth column of $\mathrm{Id} - (1-\lambda)Q(x, y)$ with $\lambda g(x, y)$. *Lemma 2* can be extended word for word, by replacing sums with the corresponding integrals and setting $\widehat{x} := x^1 \otimes \dots \otimes x^n \in \Delta(I^n)$.

For each $z \in \mathbb{R}$, the auxiliary game $W^k_\lambda(z)$ can be defined in a similar manner by setting

$$W^k_\lambda(z)[\mathbf{i},\mathbf{j}] := d^k_\lambda(\mathbf{i},\mathbf{j}) - z d^0_\lambda(\mathbf{i},\mathbf{j}) \qquad \forall (\mathbf{i},\mathbf{j}) \in I^n \times J^n.$$

Note that $W^k_\lambda(z)$ is no longer a matrix, but a mapping from the compact metric set $I^n \times J^n$ to $\mathbb{R}$. Like in the finite case, consider

ECONOMIC SCIENCES

APPLIED MATHEMATICS

the mixed extension of this game; that is, the zero-sum game with action sets $\Delta(I^n)$ and $\Delta(J^n)$ and payoff function

$$W_\lambda^k(z)[\mathbf{x},\mathbf{y}] := \int_{I^n \times J^n} W_\lambda^k(z)[\mathbf{i},\mathbf{j}]\, d(\mathbf{x}\otimes\mathbf{y})(\mathbf{i},\mathbf{j}).$$

By the minmax theorem stated in item *i*) of *Section 2A*, this game admits a value, denoted by val $W_\lambda^k(z)$. *Lemmas 3* and *4* can thus be extended word for word as well; it is enough to replace all sums with the corresponding integrals. The extension of *Theorem 1* follows directly from these 2 lemmas.

Extension of Theorem 2. *Theorem 2* cannot be extended to compact-continuous stochastic games.

Indeed, Vigeral (35) provided an example of a stochastic game with compact action sets and continuous payoff and transition functions for which the discounted values do not converge. In this sense, the extension of our result to this framework is not possible. However, we point out that only one point in our proof is problematic. Indeed, the failure occurs in the use of *Lemma 1*, which relies on the formula stated as property *iv*) in *Section 2A*, which holds only in the finite case. For infinite action sets it is no longer true that $\lambda \mapsto$ val $W_\lambda^k(z)$ is a rational fraction in λ in a neighborhood of 0 for all $z\in\mathbb{R}$, which was crucial to prove the existence of the limit $F^k(z) := \lim_{\lambda\to 0}$ val $W_\lambda^k(z)/\lambda^n$.

Determining necessary and sufficient conditions on I, J, g, and q which ensure the convergence of the discounted values or the existence of the value is an open problem. Bolte, Gaubert, and Vigeral (36) provided sufficient conditions, namely that g and q are separable and definable. Without going into a precise definition of these 2 conditions, they hold in particular when the payoff function g and the transition q are polynomials in the players' actions. However, the case where I, J, g, and q are semialgebraic is still unsolved. (A subset E of $\mathbb{R}^d$ is semialgebraic if it is defined by finitely many polynomial inequalities; a function is semialgebraic if its graph is semialgebraic.)

C. Absorbing Games. We now show that Kohlberg's (21) result on absorbing games is captured in *Theorem 2*. An absorbing game is a stochastic game (K, I, J, g, q, k) so that, for some fixed state $k_0 \in K$,

$$q(k \mid k,i,j) = 1 \qquad \forall (i,j)\in I\times J, \quad \forall\, k\neq k_0.$$

For any initial state $k\neq k_0$, the state does not evolve during the game and, as a consequence, v_λ^k is equal to the value of the matrix $(g(k,i,j))_{(i,j)\in I\times J}$ for all $\lambda\in(0,1]$ and $k\neq k_0$. We use the notation v^k to emphasize that v_λ^k does not depend on λ, for all $k\neq k_0$.

Notation. We assume without loss of generality that $k_0 = 1$ and set $u(z) := (z, v^2, \dots, v^n)$ for all $z\in\mathbb{R}$.

Kolhberg's result. Every absorbing game $(K, I, J, g, q, 1)$ has a value, denoted by v^1, which is the unique point where the function $T:\mathbb{R}\to\mathbb{R}\cup\{\pm\infty\}$ changes sign; T is defined using the Shapley operator by

$$T(z) := \lim_{\lambda\to 0} \frac{\Phi^1(\lambda, u(z)) - z}{\lambda}.$$

Comparison to our result. We claim that $F^1 = T$ in the class of absorbing games. First of all, for all $(\mathbf{i},\mathbf{j})\in I^n\times J^n$,

$$d_\lambda^0(\mathbf{i},\mathbf{j}) = \lambda^{n-1}\left(1-(1-\lambda)q(1\mid 1,\mathbf{i}^1,\mathbf{j}^1)\right)$$
$$d_\lambda^1(\mathbf{i},\mathbf{j}) = \lambda^{n-1}\left(\lambda g(1,\mathbf{i}^1,\mathbf{j}^1) + (1-\lambda)\sum_{\ell=2}^n q(\ell\mid 1,\mathbf{i}^1,\mathbf{j}^1)v^\ell\right).$$

Thus, for any $z\in\mathbb{R}$, the $(\mathbf{i},\mathbf{j})$ th entry of $W_\lambda^1(z)$ is equal to

$$\lambda^{n-1}\left(\lambda g(1,\mathbf{i}^1,\mathbf{j}^1) + (1-\lambda)\textstyle\sum_{\ell=1}^n q(\ell\mid 1,\mathbf{i}^1,\mathbf{j}^1)u^\ell(z) - z\right).$$

In particular, $W_\lambda^1(z)$ depends on $(\mathbf{i},\mathbf{j})$ only through $(\mathbf{i}^1,\mathbf{j}^1)\in I\times J$. By eliminating the redundant rows and columns of $W_\lambda^1(z)$ one thus obtains the matrix $\lambda^{n-1}(\mathcal{G}_{\lambda,u}^1 - zU)$, where U denotes a matrix of ones of appropriate size, and $\mathcal{G}_{\lambda,u}^1$ is the matrix game described in item *v*) of *Section 2B*. The affine invariance of the value operator, namely val$(cM+dU) = c$ val $M + d$ for any matrix M and any $(c,d)\in(0,+\infty)\times\mathbb{R}$, gives then

$$\frac{\text{val } W_\lambda^1(z)}{\lambda^n} = \frac{\lambda^{n-1}\text{val }(\mathcal{G}_{\lambda,u(z)}^1 - zU)}{\lambda^n} = \frac{\Phi^1(\lambda,u(z)) - z}{\lambda}.$$

Taking λ to 0 gives the desired equality.

7. Data Availability

There are no data associated with this paper.

ACKNOWLEDGMENTS. We are greatly indebted to Sylvain Sorin, whose comments on an earlier draft led to significant simplifications of our main proofs. We are also very thankful to Abraham Neyman for his careful reading and numerous remarks on a previous version of this paper and to Bernhard von Stengel, the Editor, and the anonymous reviewers for their insightful comments and suggestions at a later stage. M.O.-B. gratefully acknowledges the support of the French National Research Agency for the Project CIGNE (Communication and Information in Games on Networks) (ANR-15-CE38-0007-01), and the support of the Cowles Foundation at Yale University.

1. L. Shapley, Stochastic games. *Proc. Natl. Acad. Sci. U.S.A.* **39**, 1095–1100 (1953).
2. J. von Neumann, Zur theorie der gesellschaftsspiele. *Math. Ann.* **100**, 295–320 (1928).
3. E. Solan, N. Vieille, Stochastic games. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13743–13746 (2015).
4. T. Bewley, E. Kohlberg, The asymptotic theory of stochastic games. *Math. Oper. Res.* **1**, 197–208 (1976).
5. J. F. Mertens, A. Neyman, Stochastic games. *Int. J. Game Theory* **10**, 53–66 (1981).
6. J. F. Mertens, A. Neyman, Stochastic games. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2145–2146 (1982).
7. M. Sion, On general minimax theorems. *Pac. J. Math.* **8**, 171–176 (1958).
8. L. Shapley, R. Snow, "Basic solutions of discrete games" in *Contributions to the Theory of Games, Vol. I*, H. Kuhn, A. Tucker, Eds. (Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 1950), vol. 24, pp. 27–35.
9. S. Sorin, *A First Course on Zero-Sum Repeated Games* (Springer Science & Business Media, 2002), vol. 37.
10. J. Renault, A tutorial on zero-sum stochastic games. ArXiv:1905.06577 (16 May 2019).
11. W. Szczechla, S. Connell, J. Filar, O. Vrieze, On the Puiseux series expansion of the limit discount equation of stochastic games. *SIAM J. Control Optim.* **35**, 860–875 (1997).
12. M. Oliu-Barton, The asymptotic value in stochastic games. *Math. Oper. Res.* **39**, 712–721 (2014).
13. A. Neyman, S. Sorin, Repeated games with public uncertain duration process. *Int. J. Game Theory* **39**, 29–52 (2010).
14. B. Ziliotto, A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. *Math. Oper. Res.* **41**, 1522–1534 (2016).
15. A. Neyman, Stochastic games with short-stage duration. *Dyn. Games Appl.* **3**, 236–278 (2013).
16. M. Oliu-Barton, B. Ziliotto, Constant payoff in zero-sum stochastic games. ArXiv:1811.04518 (11 November 2018).
17. S. Sorin, X. Venel, G. Vigeral, Asymptotic properties of optimal trajectories in dynamic programming. *Sankhya A* **72**, 237–245 (2010).
18. R. Howard, *Dynamic Programming and Markov Processes* (John Wiley, New York, NY, 1960).
19. A. Hoffman, R. M. Karp, On nonterminating stochastic games. *Manag. Sci.* **12**, 359–370 (1966).
20. D. Blackwell, T. Ferguson, The big match. *Ann. Math. Stat.* **39**, 159–163 (1968).
21. E. Kohlberg, Repeated games with absorbing states. *Ann. Stat.* **2**, 724–738 (1974).
22. R. Laraki, Explicit formulas for repeated games with absorbing states. *Int. J. Game Theory* **39**, 53–69 (2010).
23. S. Sorin, G. Vigeral, Existence of the limit value of two person zero-sum discounted repeated games via comparison theorems. *J. Optim. Theory Appl.* **157**, 564–576 (2013).

24. K. Chatterjee, R. Majumdar, T. Henzinger, Stochastic limit-average games are in EXPTIME. *Int. J. Game Theory* **37**, 219–234 (2008).
25. K. Etessami, M. Yannakakis, "Recursive concurrent stochastic games" in *International Colloquium on Automata, Languages and Programming* (Part II, Lecture Notes in Computer Science, Springer, Berlin, Germany, 2006), vol. 4052, pp. 324–335.
26. E. Solan, N. Vieille, Computing uniformly optimal strategies in two-player stochastic games. *Econ. Theor.* **42**, 237–253 (2010).
27. T. Henzinger, L. de Alfaro, K. Chatterjee, "Strategy improvement for concurrent reachability games" in *Third International Conference on the Quantitative Evaluation of Systems (QEST)* (IEEE, 2006), pp. 291–300.
28. S. Rao, R. Chandrasekaran, K. Nair, Algorithms for discounted stochastic games. *J. Optim. Theory Appl.* **11**, 627–637 (1973).
29. K. A. Hansen, M. Koucký, N. Lauritzen, P. B. Miltersen, E. P. Tsigaridas, "Exact algorithms for solving stochastic games" in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing* (ACM, 2011), pp. 205–214.
30. M. Oliu-Barton, New algorithms for solving stochastic games. ArXiv:1810.13019 (15 May 2019).
31. A. Ostrowski, Sur la détermination des bornes inférieures pour une classe des determinants. *Bull. Sci. Math.* **61**, 19–32 (1937).
32. N. Karmarkar, "A new polynomial-time algorithm for linear programming" in *Proceedings of the 16th Annual ACM Symposium on Theory of Computing* (ACM, 1984), pp. 302–311.
33. F. Atkinson, *Multiparameter Eigenvalue Problems* (Academic Press, New York, NY, 1972), vol. I.
34. L. Attia, M. Oliu-Barton, Shapley-Snow kernels, multiparameter eigenvalue problems and stochastic games. ArXiv:1810.08798 (22 May 2019).
35. G. Vigeral, A zero-sum stochastic game with compact action sets and no asymptotic value. *Dyn. Games Appl.* **3**, 172–186 (2013).
36. J. Bolte, S. Gaubert, G. Vigeral, Definable zero-sum stochastic games. *Math. Oper. Res.* **40**, 171–191 (2014).