Poising on Ariadne’s thread: An algorithm for computing a maximum clique in polynomial time

Ioannis Avramopoulos and the Ghost of Helags

In this paper, we present a polynomial-time algorithm for the maximum clique problem, which implies \( P = NP \). Our algorithm is based on a continuous game-theoretic representation of this problem and at its heart lies a discrete-time dynamical system. The rule of our dynamical system depends on a parameter such that if this parameter is equal to the maximum-clique size, the iterates of our dynamical system are guaranteed to converge to a maximum clique.

1 Introduction

“You want forever, always or never.”
— The Pierces

There was a general belief that \( NP \)-complete problems are computationally intractable in that they required exponential time to solve in the worst case. In this paper, we prove a polynomial upper bound on such problems by giving a polynomial-time algorithm for the maximum clique problem \([Pardalos and Xue, 1994, Bomze et al., 1999, Wu and Hao, 2015]\) are introductions to this problem). Briefly, the clique problem is, given a positive integer \( k \), to decide if an undirected graph \( G \) has a clique of size \( k \). This decision problem is \( NP \)-complete \([Karpe, 1972]\). In this paper, we present a polynomial-time algorithm for the corresponding \( NP \)-hard optimization problem (given an undirected graph \( G \), find a maximum clique in \( G \)). Our approach draws on the power of characterizations. The maximum clique problem has many equivalent formulations as an integer programming problem or as a continuous non-convex optimization problem \([Pardalos and Xue, 1994]\). Our approach to computing a maximum clique is based on the latter continuous formulation.

1.1 The maximum clique problem as quadratic optimization

Let \( G \) be an undirected graph and let \( A \) be its adjacency matrix. \([Motzkin and Strauss, 1965]\) relate the solutions of the optimization problem

\[
\max \{ X \cdot AX | X \in \Delta \},
\]

where

\[
\Delta = \left\{ X \in \mathbb{R}^n \left| \sum_{i=1}^{n} X(i) = 1, X(i) \geq 0, i = 1, \ldots, n \right. \right\},
\]

and \( n \) is the number of vertices of \( G \), with the maximum-clique size (clique number) of \( G \). In particular, they show that if \( X^* \) is a global maximizer, then

\[
X^* \cdot AX^* = 1 - \frac{1}{\omega(G)}
\]
where $\omega(G)$ is the clique number. They further showed that the uniform strategy over a maximum clique is a global maximizer. But in their formulation other global maximizers may exist (even further stationary points do not necessarily coincide with maximal cliques [Pelillo and Jagota, 1995]). Bomze [1997] shows that if $X^*$ is a maximizer of the optimization problem

$$\max\{X \cdot CX | X \in \Delta\},$$

where $C = A + 1/2I$ and $I$ is the identity matrix, then

$$X^* \cdot CX^* = 1 - \frac{1}{2\omega(G)}$$

He further shows that uniform strategies over maximum cliques are the unique global maximizers (avoiding the spurious solutions of the Motzkin-Strauss formulation).

### 1.2 The maximum clique problem as equilibrium computation

It is often beneficial to study quadratic programs of the previous form as **doubly symmetric bimatrix games**, that is 2-player games where the payoff matrix of each player is the transpose of that of the other and the payoff matrix is also symmetric. In a doubly symmetric game whose payoff matrix is $C$, there is a one-to-one correspondence between strict local maxima of $X \cdot CX$ over the probability simplex and **evolutionarily stable strategies** (for example, see [Weibull, 1995]).

The evolutionarily stable strategy (ESS) has its origins in mathematical biology [Maynard Smith and Price, 1973; Maynard Smith, 1982] but it admits a characterization [Hofbauer et al., 1979] more prolific in our setting: An ESS is an isolated equilibrium point that exerts an “attractive force” in a neighborhood (especially under the replicator dynamic in continuous [Taylor and Jonker, 1978] or discrete [Baum and Eagon, 1967] form). Starting from an interior strategy, the replicator dynamic in the doubly symmetric bimatrix whose payoff matrix is $C = A + 1/2I$ is ensured to converge to a maximal clique (see [Bomze, 1997; Pelillo and Torsello, 2006]) that is not necessarily maximum.

### 1.3 Our maximum clique computation algorithm

To compute a maximum clique in polynomial time further ideas are needed. Our approach in this paper is based on a game-theoretic construction due to Nisan [2006]. By adapting the aforementioned result of Motzkin and Strauss [1965] and building on Etessami and Lochbihler [2008], Nisan [2006] constructs a game-theoretic **transformation** of the clique problem that receives as input an undirected graph and gives as output a doubly symmetric bimatrix game whose payoff matrix is akin to $C$. We refer to the payoff matrix of the game Nisan designed as the **Nisan-Bomze payoff matrix** and denote it by $C^+$. The primary question driving Nisan’s inquiry is the computational complexity of recognizing an ESS: A distinctive property of the Nisan game is that a certain pure strategy, called strategy 0 in his paper (and also denoted $E_0$ in this paper), is an ESS if and only if a parameter of the Nisan-Bomze payoff matrix (which we call the **Nisan parameter**) exceeds the clique number. Nisan shows that the problem of recognizing if $E_0$ is an ESS is **coNP**-complete.

The Nisan game is the conceptual basis of our maximum-clique computation algorithm.

Our early experience with maximum-clique computation in the Nisan game was negative. If the Nisan parameter is greater than the clique number, $E_0$ is a global ESS (GESS), which implies that $E_0$ is the unique symmetric equilibrium strategy. But if the Nisan parameter is equal to the clique number, $E_0$ remains a global neutrally stable strategy but other equilibria appear, namely, one equilibrium for each maximum clique (which we refer to as **maximum-clique equilibria**) and their corresponding convex combinations with $E_0$. Maximum-clique equilibria and their convex
combinations with $E_0$ are neutrally stable strategies. An approach we followed to compute a maximum clique was to try to compute a neutrally stable strategy other than $E_0$. (See Section 2 for definitions of evolutionary and neutral stability as well as their global versions). The equilibria of the Nisan game when the Nisan parameter is equal to the clique number are an evolutionarily stable set, which implies that such equilibria are neutrally stable strategies that are attractive under the replicator dynamic (in continuous or discrete form). Given any neutrally stable strategy other than $E_0$ a maximum-clique can be readily recovered. Our approach to compute a neutrally stable strategy other than $E_0$ was to try to exploit that neutrally stable strategies attract multiplicative weights. But our effort to provably stay out of the region of attraction of $E_0$ was futile.

We then discovered that the problem of computing a maximum clique has a “backdoor” in the Nisan game. To unlock this backdoor we isolated the attractive force of $E_0$ by intersecting the evolution space of the Nisan game with a hyperplane perpendicular to $E_0$ and using the intersection of this hyperplane with the evolution space of the Nisan game as the evolution space of our equilibrium computation algorithm. This approach can be implemented by restricting the probability mass of strategy $E_0$ to a fixed value $\epsilon \in (0, 1)$ and by adapting the multiplicative weights algorithm. The multiplicative weights algorithm we adapted is Hedge [Freund and Schapire, 1997, 1999]. Denoting the Nisan-Bomze payoff matrix by $C$, Hedge assumes the following expression in the Nisan game:

$$T_i(X) = X(i) \cdot \frac{\exp \{\alpha E_i \cdot CX\}}{\sum_{j=1}^{n} X(j) \exp \{\alpha E_j \cdot CX\}}, \quad i = 0, 1, \ldots, n$$

where $\alpha$ is a parameter called the learning rate, which has the role of a step size in our equilibrium computation setting. If we restrict the probability mass of $E_0$ to the fixed value $0 < \epsilon < 1$ our dynamical system assumes the following expression:

$$T_0(X) = \epsilon$$

$$T_i(X) = X(i) \cdot \frac{(1 - \epsilon) \exp \{\alpha E_i \cdot CX\}}{\sum_{j=1}^{n} X(j) \exp \{\alpha E_j \cdot CX\}}, \quad i = 1, \ldots, n.$$  

The latter dynamical system is not ensured to converge to a maximum-clique equilibrium—instead, it may converge to a maximal clique. Here comes one critical idea in this vein: In the evolution space wherein this system is acting, maximum-clique equilibria have a distinctive property, namely, assuming the Nisan parameter is equal to the clique number, if $X^*$ is a maximum clique equilibrium, then

$$\max_{i=1}^{n} \{(CX^*)_i\} - X^* \cdot CX^* = 0$$

whereas if $X^*$ is a maximal clique equilibrium, then

$$\max_{i=1}^{n} \{(CX^*)_i\} - X^* \cdot CX^* < 0.$$  

To exploit this phenomenon, in an effort to enforce convergence of our dynamical system to a maximum clique we perturbed the components $E_i \cdot CX$ of the gradient $CX$ of the objective function $X \cdot CX$ with the components of the gradient of a logarithmic barrier function, a technique that is akin to interior point optimization methods barring we did not vanish the perturbation as time progresses. But we had trouble capturing the precise effect of the logarithmic barrier function on the behavior of our dynamical system analytically, and we decided to restrict the evolution space of our dynamical system even further. This latter approach is followed in this paper.  

\footnote{A generalization of the discrete-time replicator dynamic.}
In this paper, our dynamical system evolves in a “lower feasibility set” (a slice of the game). This ensures that if the Nisan parameter \( k \) is equal to greater than the clique number (denoted \( \omega(G) \)) no equilibria of \( C \) appear within this feasibility set other than maximum-clique equilibria which appear when \( k = \omega(G) \). Our algorithm is guided by this property starting with a large value of \( k \) and incrementally decreasing \( k \) until a maximum-clique equilibrium can be computed. We have derived a condition that enables us to determine when the search for a particular value of \( k \) should be abandoned, that \( k \) should decrease, and the search should continue using a smaller value.

1.4 Our proof techniques

In our algorithm, which we call Ariadne, the computation of a maximum clique is guided by the iterations of a dynamical system. In fact, there are two versions this dynamical system, one that is discontinuous (but admits a continuous Lyapunov function) and one that is continuous (that also admits a continuous Lyapunov function). The continuous version of the dynamical is activated when the Nisan parameter is equal to the clique number. (Once the secondary dynamical system is activated, we learn the value of the clique number, but execution needs to continue to compute a maximum clique.) To prove that this combination of dynamical systems guides Ariadne toward a maximum clique, we prove asymptotic convergence to a maximum-clique equilibrium. The proof rests on the fact that these dynamical systems are growth transformations (see [Baum and Eagon, 1967], [Baum and Sell, 1968], [Gopalakrishnan et al., 1991]) for corresponding barrier functions.

The growth transformations proposed by previous authors (in the aforementioned references) are based on the discrete-time replicator dynamic. In this paper, our growth transformations are based on Hedge, which facilitates the analysis deriving equilibrium approximation bounds. Our proof that Hedge is a growth transformation makes use of a limiting argument: We derive Hedge as the limit of more elementary maps and use elementary functional analysis for our conclusion.

However, perhaps the most important analytic technique introduced in this paper is the derivation of equilibrium approximation bounds using the Chebyshev order inequality (known also as the generalized Chebyshev sum inequality). To derive equilibrium approximation bounds for our dynamical system, we first derive an order preservation principle for our dynamical system which we then “plug in” the Chebyshev order inequality to obtain polynomial bounds on the number of iterations to approximately converge to a non-equilibrium fixed point or a maximum-clique equilibrium.

Our main result in this latter direction is Theorem 3 in Appendix B which applies to any symmetric bimatrix game and, therefore, its applicability is more general than the doubly symmetric games that are analyzed in this paper. In general symmetric bimatrix games, we cannot expect blanket convergence to a symmetric equilibrium strategy starting from any (interior) initialization: [Daskalakis et al., 2010] show that in Shapley’s 3 × 3 symmetric bimatrix game the dynamics defined by using Hedge in each player position and computing the empirical average of each player’s iterated sequence of strategies that ensues from the interaction diverge (under assumptions on the learning rate) for nonuniform initializations of play. For example, consider the symmetric game \((C, C^T)\), where

\[
C = \frac{1}{2} \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 2 \\
0 & 0 & 0 & 2 & 0 & 1 \\
0 & 0 & 0 & 1 & 2 & 0 \\
0 & 1 & 2 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0
\end{bmatrix},
\]
which an extended form of Shapley’s game. Figure 1 illustrates divergence of the sequence of averages from the uniform equilibrium starting from initial condition (0.1, 0.2, 0.3, 0.2, 0.1, 0.1). This divergence phenomenon cannot manifest in Ariadne (because matrices are doubly symmetric and Lemma 15 ensures convergence of the iterates to a maximum clique equilibrium).

1.5 OVERVIEW OF THE REST OF THE PAPER

The rest of this paper is organized as follows: In Section 2 we present game-theoretic background. In Section 3 we define the archetypical form of our dynamical system and characterize its fixed points—they coincide with the fixed points of the replicator dynamic. In Section 4 we present our algorithm, its various components, and their analysis. A main feature of our algorithm is that we use a barrier function to restrict evolution of a dynamical system inside a desirable subset of the system’s blanket evolution space via a growth transformation. Using such barrier functions we are able to perform non-convex global optimization. To the best of our knowledge this is the first paper where growth transformations are used in this fashion. In Section 5 we prove correctness and that the Ariadne’s complexity is polynomial. Finally, in the Appendix, we prove that Hedge is a growth transformation for positive values of the learning rate parameter in homogeneous polynomials with nonnegative coefficients (subject to constraints on the coefficients). We finally derive an inequality on the fixed-point approximation error of our dynamical system. To the extent of our knowledge, this inequality is the first equilibrium approximation bound in non-convex problems using Hedge (or other multiplicative weight algorithms such as the discrete-time replicator dynamic).

---

2 We would like to thank an anonymous reviewer of a related submission for pointing out this example.
2 Preliminary background on Nash equilibria

2.1 Bimatrix games and symmetric bimatrix games

A 2-player (bimatrix) game in normal form is specified by a pair of \( n \times m \) matrices \( A \) and \( B \), the former corresponding to the row player and the latter to the column player. A mixed strategy for the row player is a probability vector \( P \in \mathbb{R}^n \) and a mixed strategy for the column player is a probability vector \( Q \in \mathbb{R}^m \). The payoff to the row player of \( P \) against \( Q \) is \( P \cdot AQ \) and that to the column player is \( P \cdot BQ \). Denote the space of probability vectors for the row player by \( \mathcal{P} \) and for the column player by \( \mathcal{Q} \). A Nash equilibrium of the bimatrix game \((A,B)\) is a pair of mixed strategies \( P^* \) and \( Q^* \) such that all unilateral deviations from these strategies are not profitable, that is, for all \( P \in \mathcal{P} \) and \( Q \in \mathcal{Q} \), we simultaneously have that

\[
P^* \cdot AQ^* \geq P \cdot AQ^* \quad (1)
\]
\[
P^* \cdot BQ^* \geq P^* \cdot BQ. \quad (2)
\]

(For example, see [von Stengel, 2007].) \( A, B \) are called payoff matrices. We denote the set of Nash equilibria of the bimatrix game \((A,B)\) by \( \text{NE}(A,B) \). If \( B = A^T \), where \( A^T \) is the transpose of matrix \( A \), the bimatrix game is called a symmetric bimatrix game. Let \((C,C^T)\) be a symmetric bimatrix game. We denote the space of symmetric bimatrix games by \( \mathcal{C} \). \( \hat{\mathcal{C}} \) denotes the space of payoff matrices whose entries lie in the range \([0,1]\).

Pure strategies are denoted either as \( i \) or as \( E_i \), where \( E_i \) is a probability vector whose mass is concentrated in position \( i \). \( \mathcal{X}(C) \) denotes the space of mixed strategies of \((C,C^T)\) (a probability simplex). We call \((X^*,X^*)\) a symmetric (Nash) equilibrium strategy if \( X^* \in \mathcal{X}(C) \) satisfies

\[
\forall X \in \mathcal{X}(C) : (X^* - X) \cdot CX^* \geq 0.
\]

\( \text{NE}^+(C) \) denotes the symmetric equilibrium strategies of \((C,C^T)\). We denote the (relative) interior of \( \mathcal{X}(C) \) by \( \hat{\mathcal{X}}(C) \) (every pure strategy in \( \hat{\mathcal{X}}(C) \) has probability mass). Let \( X \in \mathcal{X}(C) \). We define the support or carrier of \( X \) by

\[
\mathcal{C}(X) \equiv \{ i \in \mathcal{K}(C) | X(i) > 0 \}.
\]

A doubly symmetric bimatrix game [Weibull, 1995, p.26] is a symmetric bimatrix game whose payoff matrix, say \( C \), is symmetric, that is \( C = C^T \). Symmetric equilibria in doubly symmetric games are KKT points of a standard quadratic program (cf. [Bomze, 1998]):

\[
\begin{align*}
\text{maximize} & \quad X \cdot CX \\
\text{subject to} & \quad X \in \mathcal{X}(C).
\end{align*}
\]

\( X \cdot CX \) is the potential function of the game.

2.2 Equalizers: Definition and basic properties

**Definition 1.** \( X^* \in \mathcal{X}(C) \) is called an equalizer if

\[
\forall X \in \mathcal{X}(C) : (X^* - X) \cdot CX^* = 0.
\]

We denote the set of equalizers of \( C \) by \( \mathcal{E}(C) \).
Note that $\mathbb{E}(C) \subseteq NE^+(C)$. Equalizers generalize interior symmetric equilibrium strategies, as every such strategy is an equalizer, but there exist symmetric bimatrix games with a non-interior equalizer (for example, if a column of $C$ is constant, the corresponding pure strategy of $C$ is an equalizer of $C$). Note that an equalizer can be computed in polynomial time by solving the linear (feasibility) program (LP)

$$(CX)_1 = \cdots = (CX)_n, \quad \sum_{i=1}^n X(i) = 1, \quad X \geq 0,$$

which we may equivalently write as

$$CX = c1, \quad 1^TX = 1, \quad X \geq 0,$$

where $1$ is a column vector of ones of appropriate dimension. We may write this problem as a standard LP as follows: Let

$$A \doteq \begin{bmatrix} C & -1 \\ 1^T & 0 \end{bmatrix} \quad \text{and} \quad Y \doteq \begin{bmatrix} X \\ c \end{bmatrix},$$

then

$$AY = \begin{bmatrix} C & -1 \\ 1^T & 0 \end{bmatrix} \begin{bmatrix} X \\ c \end{bmatrix} = \begin{bmatrix} CX - c1 \\ 1^TX \end{bmatrix},$$

and the standard form of our LP, assuming $C > 0$, is

$$\begin{bmatrix} CX - c1 \\ 1^TX \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad X \geq 0, c \geq 0 \quad (3)$$

where $0$ is a column vector of zeros of appropriate dimension. We immediately obtain that:

**Lemma 1.** $\mathbb{E}(C)$ is a convex set.

*Proof.* The set of feasible/optimal solutions of a linear program is a convex set. Let

$$\mathbb{Y}^* = \{ [X^T c]^T | [X^T c]^T \text{ is a feasible solution to (3)} \}.$$

Then $\mathbb{Y}^*$ is convex and therefore the set

$$\mathbb{X}^* = \{ X | [X^T c]^T \text{ is a feasible solution to (3)} \}$$

is also convex since $c$ is unique provided the LP is feasible.

We can actually show something stronger:

**Lemma 2.** If $X_1^*, X_2^* \in \mathbb{E}(C)$ then $\Lambda(X_1^*, X_2^*) \subseteq \mathbb{E}(C)$, where

$$\Lambda(X_1^*, X_2^*) = \{ (1 - \lambda)X_1^* + \lambda X_2^* \in \mathbb{X}(C) | \lambda \in \mathbb{R} \}.$$ 

*Proof.* Assume $X_1^*, X_2^* \in \mathbb{E}(C)$. Then, by the definition of an equalizer,

$$\forall X \in \mathbb{X}(C) : X_1^* \cdot CX_1^* = X \cdot CX_1^* \text{ and}$$

$$\forall X \in \mathbb{X}(C) : X_2^* \cdot CX_2^* = X \cdot CX_2^*.$$
Let
\[ Y^* = (1 - \lambda)X_1^* + \lambda X_2^*, \quad \lambda \in \mathbb{R}. \]

Then
\begin{align*}
Y^* \cdot CY^* & = (1 - \lambda) ((1 - \lambda)X_1^* + \lambda X_2^*) \cdot CX_1^* + \lambda ((1 - \lambda)X_1^* + \lambda X_2^*) \cdot CX_2^* \\
& = (1 - \lambda)X_1^* \cdot CX_1^* + \lambda X_2^* \cdot CX_2^* \\
& = (1 - \lambda)X \cdot CX_1^* + \lambda X \cdot CX_2^* \\
& = X \cdot C ((1 - \lambda)X_1^* + \lambda X_2^*) \\
& = X \cdot CY^*.
\end{align*}

Since \( X \) is arbitrary, the proof is complete.

### 2.3 Approximate and Well-Supported Approximate Equilibria

As mentioned earlier, conditions \( \text{(1)} \) and \( \text{(2)} \) simplify as follows for a symmetric equilibrium strategy \( X^* \):
\[
\forall X \in \mathbb{X}(C) : (X^* - X) \cdot CX^* \geq 0.
\]

An \( \epsilon \)-approximate symmetric equilibrium, say \( X^* \), is defined as follows:
\[
\forall X \in \mathbb{X}(C) : (X^* - X) \cdot CX^* \geq -\epsilon.
\]

We may equivalently write the previous expression as
\[
(CX^*)_{\max} - X^* \cdot CX^* \leq \epsilon,
\]
where
\[
(CX^*)_{\max} = \max\{Y \cdot CX^* | Y \in \mathbb{X}(C)\}.
\]

Let us now give an important result on approximate equilibria. To that end, we need a definition:

**Definition 2.** \( (X^*, Y^*) \) is an \( \epsilon \)-well-supported Nash equilibrium of \( (A, B) \) if
\[
E_i \cdot AY^* > E_k \cdot AY^* + \epsilon \Rightarrow X^*(k) = 0 \quad \text{and} \quad X^* \cdot BE_j > X^* \cdot BE_k + \epsilon \Rightarrow Y^*(k) = 0.
\]

Definition 2 is due to Daskalakis et al. [2009]. We note that an \( \epsilon \)-well-supported Nash equilibrium of \( (A, B) \) is necessarily an \( \epsilon \)-approximate equilibrium of \( (A, B) \) but the converse is not generally true. However, given an approximate equilibrium we can obtain a well-supported equilibrium:

**Proposition 1.** Let \( (A, B) \) be such that \( 0 \leq A, B \leq 1 \). Given an \( \epsilon^2/8 \)-approximate Nash equilibrium of \( (A, B) \), where \( 0 \leq \epsilon \leq 1 \), we can find an \( \epsilon \)-well-supported Nash equilibrium in polynomial time.

The previous proposition is due to Chen et al. [2009] motivated by a related result in Daskalakis et al. [2009]. We have the following characterization of well-supported equilibria:
Proposition 2. \((X^*, Y^*)\) is an \(\epsilon\)-well-supported Nash equilibrium of \((A, B)\) if and only if

\[
X^*(i) > 0 \Rightarrow E_i \cdot AY^* \geq \max_{k=1}^m E_k \cdot AY^* - \epsilon \quad \text{and} \\
Y^*(j) > 0 \Rightarrow X^* \cdot BE_j \geq \max_{\ell=1}^n X^* \cdot BE_\ell - \epsilon.
\]

Proof. The statement of the lemma is just the contrapositive of Definition 2. \hfill \Box

In a symmetric bimatrix game, the previous proposition simplifies as:

Proposition 3. \(X^*\) is an \(\epsilon\)-well-supported symmetric equilibrium strategy of \((C, C^T)\) if and only if it is an \(\epsilon\)-approximate symmetric equilibrium strategy and

\[
(\hat{C}X)_{\max} - (\hat{C}X)_{\min} \leq \epsilon
\]

where \(\hat{C}\) is the carrier of \(X^*\).

2.4 Evolutionary stability

An equilibrium notion in symmetric bimatrix games (and, therefore, also in doubly symmetric bimatrix games) of primary interest in this paper is the GESS (global evolutionarily stable strategy), which is a global version of the ESS [Maynard Smith and Price, 1973, Maynard Smith, 1982]. Of primary interest are also related equilibrium notions such as the NSS (neutrally stable strategy) and GNSS (global NSS). These concepts admit the following definitions:

Definition 3. Let \(C \in \mathbb{C}\). We say \(X^* \in \mathbb{X}(C)\) is an ESS, if

\[
\exists O \subseteq \mathbb{X}(C) \forall X \in O/\{X^*\} : X^* \cdot CX > X \cdot CX.
\]

Here \(O\) is a neighborhood of \(X^*\). If \(O\) coincides with \(\mathbb{X}(C)\), we say \(X^*\) is a GESS. If the above inequality is weak we have an NSS and a GNSS respectively.

The aforementioned definition of an ESS was originally obtained as a characterization [Hofbauer et al., 1979]. Note that an NSS, and, therefore, also an ESS, is necessarily a symmetric equilibrium strategy. The ESS and NSS admit the following characterizations. These characterizations correspond to how they were initially defined.

Proposition 4. \(X^*\) is an ESS of \((C, C^T)\) if and only if the following conditions hold simultaneously

\[
X^* \cdot CX^* \geq X \cdot CX^*, \forall X \in \mathbb{X}(C), \quad \text{and} \\
X^* \cdot CX^* = X \cdot CX^* \Rightarrow X^* \cdot CX > X \cdot CX, \forall X \in \mathbb{X}(C) \text{ such that } X \neq X^*.
\]

An NSS correspond to a weak inequality.

The characterization of the ESS and NSS in Proposition 4 does not readily yield the global versions of GESS and GNSS that are of primary interest in this paper. Note finally that:

Lemma 3. If \(X^*\) is an equalizer, then \(X^*\) is an ESS if and only if it is a GESS.

Proof. Straightforward from Proposition 4. \hfill \Box

Lemma 4. If \(X^*\) is an equalizer, then \(X^*\) is an NSS if and only if it is a GNSS.
3 A dynamical systems approach to maximum-clique computation

At the heart of our maximum-clique computation algorithm lies a dynamical system based on Hedge [Freund and Schapire, 1997, 1999] that induces the following map in our setting:

\[ T_i(X) = X(i) \cdot \frac{\exp \{\alpha E_i \cdot CX\}}{\sum_{j=1}^n X(j) \exp \{\alpha E_j \cdot CX\}} \equiv X(i) \cdot \frac{\exp \{\alpha (C X)_i\}}{\sum_{j=1}^n X(j) \exp \{\alpha (C X)_j\}} \quad i = 1, \ldots, n, \]

where \( C \) is the payoff matrix of a symmetric bimatrix game, \( n \) is the number of pure strategies, \( E_i \) is the probability vector corresponding to pure strategy \( i \), and \( X(i) \) is the probability mass of pure strategy \( i \). Parameter \( \alpha \) is called the learning rate, which has the role of a step size. The fixed points of \( T \) coincide with the fixed points of the discrete-time replicator dynamic.

Lemma 5. \( X \) is a fixed point of \( T \) if and only if \( X \) is a pure strategy or otherwise

\[ \forall i, j \in \mathcal{C}(X) : (C X)_i = (C X)_j. \]

Proof. First we show sufficiency, that is, if for all \( i, j \in \mathcal{C}(X), (C X)_i = (C X)_j \), then \( T(X) = X \). Some of the coordinates of \( X \) are zero and some are positive. Clearly, the zero coordinates will not become positive after applying \( T \). Now, notice that, for all \( i \in \mathcal{C}(X) \), \( \exp \{\alpha (C X)_i\} = \sum_{j=1}^n X(j) \exp \{\alpha (C X)_j\} \). Therefore, \( T(X) = X \).

Now we show necessity, that is, if \( X \) is a fixed point of \( T \), then for all \( i \) and for all \( i, j \in \mathcal{C}(X), (C X)_i = (C X)_j \). Let \( \hat{X}(i) = T_i(x) \). Because \( X \) is a fixed point, \( \hat{X}(i) = X(i) \). Therefore,

\[ \hat{X}(i) = X(i) \cdot \frac{X(i) \exp \{\alpha (C X)_i\}}{\sum_j X(j) \exp \{\alpha (C X)_j\}} = X(i) \]

\[ \exp \{\alpha (C X)_i\} = \sum_j X(j) \exp \{\alpha (C X)_j\}, \]

which implies

\[ \exp \{\alpha ((C X)_i - (C X)_j)\} = 1, \quad X(i) > 0, \]

and, thus,

\[ (C X)_i = (C X)_j, \quad X(i) > 0. \]

This completes the proof.

Solving the maximum clique problem using Hedge is an approach also taken by Pelillo and Torsello [2006], where Hedge is referred to as “exponential replicator dynamic” in that paper. Hedge is reported in that paper to be dramatically faster than the discrete-time replicator dynamic and even more accurate. However, a blanket application of this dynamic can compute a maximal (instead of maximum) clique. The techniques considered by Pelillo and Torsello [2006] to enhance the efficacy of the approach do not provably compute a maximum clique (as we do in this paper).

4 Our maximum-clique computation algorithm

“It seems that for the maximum clique problem a good formulation of the problem is of crucial importance in solving the problem.”

— P. M. Pardalos and J. Xue

In this section, we define our maximum-clique computation algorithm and prove some of its
properties. That the algorithm’s computational complexity is polynomial is shown in the next section.

4.1 The Nisan game as the evolution space of our dynamical system

The evolution space of our dynamical system is a subset of the evolution space of the Nisan game, but before defining what this evolution space is, let us start by defining the Nisan game first. Given an undirected graph $G(V,E)$, where $|V| = n$, and an integer $1 < k \leq n$, consider the following $(n + 1) \times (n + 1)$ symmetric matrix $C^+$: $C^+$’s rows and columns correspond to the vertices of $V$, numbered 1 to $n$, with an additional row and column, numbered 0.

- For $1 \leq i \neq j \leq n$: $C^+(i,j) = 1$ if $(i,j) \in E$ and $C^+(i,j) = 0$ if $(i,j) \notin E$.
- For $1 \leq i \leq n$: $C^+(i,i) = 1/2$.
- For $1 \leq i \leq n$: $C^+(0,i) = C^+(i,0) = 1 - \frac{1}{2k}$.
- $C^+(0,0) = 1 - \frac{1}{2k} \equiv C_{00}$.

That is, $C^+$ consists of a symmetric adjacency matrix of 0’s and 1’s with the value $1/2$ on the main diagonal and an extra strategy whose corresponding payoff entries are identical and equal to the potential value of a clique of size $k$. We refer to this matrix as the Nisan-Bomze payoff matrix.

Cliques can be identified with their characteristic vectors, that is, uniform strategies over their corresponding carrier (a property retained from Motzkin and Strauss [1965]). Every characteristic vector (of a clique) is a fixed point of $T$ (cf. Section 3). We call $k$ the Nisan parameter.

Considering the doubly symmetric game whose payoff matrix is $C^+$, one of Nisan’s main results [Nisan, 2006] is that strategy 0 (which we also denote by $E_0$) is an ESS if and only if the maximum clique of $G$ is less than $k$. Note that $E_0$ is an equalizer and, therefore, it is an ESS if and only if it is a GESS (cf. Lemma 3). If the Nisan parameter is such that $E_0$ is a GESS it is easily shown that it is the unique equilibrium of the game. If the Nisan parameter is equal to the clique number, then other equilibria appear, namely, an equilibrium for every maximum clique (which is a global maximizer of the quadratic potential such as $E_0$ is) and a corresponding equilibrium line (of global maximizers) with terminal points $E_0$ and the respective maximum-clique equilibrium.

We use $C$ to denote the $n \times n$ matrix obtained from $C^+$ by excluding strategy 0. We denote the probability simplex of the latter $C$ by $\mathbb{Y}$. The payoff matrix whereby iterates are generated is obtained from $C$ by adding a positive constant matrix (for example, a matrix all of whose entries are equal to one) and scaling with a positive scalar (for example, two) such that the maximum payoff entry over the minimum payoff entry is equal to a constant (for example, two). The proof that our algorithm runs in polynomial time requires this technicality (cf. Sections 4.2.4, 4.2.5, and Appendix B). Under this transformation of $C^+$, $C_{00} \equiv C$ and $C_{\ell}$ (see below) must be accordingly adjusted. To avoid clutter, we are going to assume that this transformation is implicit.

4.2 Evolution inside a desirable “feasibility set”

To ensure the iterates compute a maximum clique when the Nisan parameter becomes equal to the clique number, we restrict the evolution space of our dynamical system. We call the restricted evolution space the “lower feasibility set.” The complement of the lower feasibility set consists of the “upper feasibility set” and the “infeasibility set.” Our algorithm ensures that iterates are initialized and forever remain in the lower feasibility set entirely avoiding the other sets except in the limit as maximum-clique equilibria are located at the boundary of the lower feasibility and upper feasibility sets. Let us first define these sets precisely and identify an elementary property.
4.2.1 Definition of the feasibility and infeasibility sets and an elementary property

We denote the lower feasibility set by \( F_L \), the upper feasibility set by \( F_U \) and the infeasibility set by \( I \) and define them as

\[
\begin{align*}
F_U & \equiv \left\{ X \in Y \mid \max_{i=1}^{n} (CX)_i \geq C_{00} \right\} \\
F_L & \equiv \left\{ X \in Y \mid C_{00} - \varepsilon \leq \max_{i=1}^{n} (CX)_i < C_{00} \right\} \\
I & \equiv \left\{ X \in Y \mid \max_{i=1}^{n} (CX)_i < C_{00} - \varepsilon \right\}.
\end{align*}
\]

All sets are polytopes and the infeasibility set \( I \) is also convex. We will specify a value for \( \varepsilon \) in the sequel. Let us note for now that \( \varepsilon \) is chosen such that only maximum clique equilibria (which are in fact located at the boundary of the lower and upper feasibility sets) and no other fixed points that may the iterates of our dynamical system can be located in the lower feasibility set \( F_L \).

**Definition 4.** \( X \in Y \) is called strictly upper feasible if

\[
\max_{i=1}^{n} (CX)_i > C_{00},
\]

strictly lower feasible if

\[
C_{00} - \varepsilon < \max_{i=1}^{n} (CX)_i < C_{00},
\]

and weakly feasible if

\[
\max_{i=1}^{n} (CX)_i = C_{00}.
\]

**Lemma 6.** Suppose \( X^* \) is a maximal-clique equilibrium and the Nisan parameter \( k \) satisfies \( k > X^* \cdot CX^* = \omega(G) \). Then \( X^* \not\in F_L \). Furthermore if \( k = X^* \cdot CX^* \), \( X^* \) is weakly feasible.

**Proof.** If \( k > X^* \cdot CX^* \), then

\[
X^* \cdot CX^* = \max_{i=1}^{n} (CX^*)_i < C_{00} - \varepsilon
\]

and the definition of the feasibility set implies \( X^* \not\in F_L \) as claimed. If \( k = X^* \cdot CX^* \), then

\[
X^* \cdot CX^* = \max_{i=1}^{n} (CX^*)_i = C_{00},
\]

which implies \( X^* \) is weakly feasible.

4.2.2 The method by which iterates remain within the thin slice of the lower feasibility set

We restrict evolution of the iterates to the interior of the lower feasibility set \( F_L \) by means of a barrier function, namely, a function \( G : Y \to \mathbb{R} \) where

\[
G(X) = \begin{cases} 
\frac{X \cdot CX - C}{\left(\max_{i=1}^{n} (CX)_i - C_0\right) \prod_{i=1}^{n} (C - (CX)_i)} , & X \in F_L \\
0 , & X \in X^* ,
\end{cases}
\]

3See [Bertsekas, 1999, Chapter 4] on Lagrange multiplier algorithms (see also [Bertsekas, 1996]) where barrier functions are discussed in their elementary form in conjunction with interior point algorithms.
where $C = C_0$, $C_\ell = C_0 - \varepsilon$, and $X^*$ is the set of maximum-clique equilibria. Using the product in the denominator of the barrier function (instead of

$$C - \max_{i=1}^n \{(CX)_i\}$$

is a trick that facilitates our subsequent analysis and appears in a related form in [Bertsekas, 1999, Proposition 3.3.10].

We call $G$-feasible any $X \in Y$ which is strictly lower feasible that is

$$C_\ell < \max_{i=1}^n \{(CX)^k_i\} < C.$$ 

We assume that the first iterate is initialized to be $G$-feasible and we prove that the remaining iterates remain so, that is,

$$\forall k \geq 0 : C_\ell < \max_{i=1}^n \{(CX^k)_i\} < C.$$ 

Our algorithm uses, however, a second barrier function, namely,

$$G_S(X) = \begin{cases} 
\frac{X \cdot CX - C}{(X \cdot CX - C_\ell) \prod_{i=1}^n (C - (CX)_i)}, & X \in \mathcal{F}_{LS} \\
0, & X \in X^*,
\end{cases}$$

which is activated in lieu of the primary barrier function on detection of a strategy $X$ such that $X \cdot CX > C_\ell$ in the course of execution of our dynamical system. This can only happen if the Nisan parameter is equal to the clique number. The set $\mathcal{F}_{LS}$ is the set of all $X \in Y$ such that $X \cdot CX > C_\ell$. Since $X \cdot CX > C_\ell$ implies $(CX)_{\text{max}} > C_\ell$, we have $\mathcal{F}_{LS} \subset \mathcal{F}_L$. We note that the secondary barrier function gives rise to a continuous dynamical system unlike the primary barrier function. This facilitates asymptotic convergence to a maximum-clique equilibrium.

Lemma 7. Suppose the Nisan parameter is equal to the clique number. Then the value of our barrier function $G$ at the limit of every strictly lower feasible sequence that converges to a maximum clique equilibrium is equal to zero. Therefore, the restriction of $G$ at any superlevel set

$$\{X \in \mathcal{F}_L \mid G(X) \geq G(X^0)\},$$

where $X^0$ is strictly feasible, is a continuous function.

Proof. Let $X$ be a maximum-clique equilibrium. We claim that $G(Y) \to 0$ as $Y$ approaches $X$ from strictly lower feasible points. That is, considering the restriction of $G$ to the set of strictly lower feasible points, we would like to show that

$$\lim_{Y \to X} \{G(Y)\} = 0.$$ \hspace{1cm} (4)

To prove our claim, note that

$$G(X) \leq \frac{X \cdot CX - C}{\max_{i=1}^n \{(CX)_i\} - C_\ell} \cdot \frac{C - \max_{i=1}^n \{(CX)_i\}}{\left(\max_{i=1}^n \{(CX)_i\} - C_\ell\right)}$$

"The secondary barrier function and corresponding dynamical system are crucial to obtain our main result that $P = NP$ but we should note that without these ideas, the remaining techniques suffice to prove $RP = NP$."

13
and that

\[
\frac{1}{C - \max_{i=1}^{n}(CX)_i} = \frac{1}{C} \frac{1}{1 - \frac{\max_{i=1}^{n}(CX)_i}{C}} \geq \frac{1}{C} \left( 1 + \frac{\max_{i=1}^{n}(CX)_i}{C} \right) \geq \frac{\max_{i=1}^{n}(CX)_i + C}{C^2}
\]

where the inequality follows by the Taylor series expansion of

\[
\frac{1}{1 - \frac{\max_{i=1}^{n}(CX)_i}{C}}
\]

which has only positive terms. Therefore,

\[
G(X) \leq \frac{X \cdot CX - C}{\left( \max_{i=1}^{n}(CX)_i - C_\ell \right)} \left( C - \max_{i=1}^{n}(CX)_i \right)
\]

\[
\leq \frac{X \cdot CX - C}{\max_{i=1}^{n}(CX)_i - C_\ell} \frac{\max_{i=1}^{n}(CX)_i + C}{C^2}
\]

which implies \( G(X) \leq 0 \). Consider now a sequence \( \{Y^k\} \) of strictly lower feasible points converging to \( X \). Since \( G(X) \leq 0 \), the limit of this sequence cannot be positive. So let us assume for the sake of contradiction that the limit of this sequence is strictly negative. This implies that \( X \) is strictly lower feasible, which contradicts Lemma 6. This proves (1), that is, that the limit of every strictly feasible sequence that converges to \( X \) is equal to zero, completing the proof of the lemma’s first part. The second part follows by the definition of \( G \) and the definition of continuity. \( \square \)

4.2.3 Our dynamical system: How iterates remain in the lower feasibility set

The discrete-time replicator dynamic is a growth transformation for any polynomial with nonnegative coefficients [Baum and Eagon, 1967, Baum and Sell, 1968]. That is, the discrete-time replicator dynamic strictly ascends polynomials with nonnegative coefficients except at fixed points wherein the value of the polynomial remains constant. This implies that in a doubly symmetric bimatrix game \( C \), \( \forall X \in \mathbb{X}(C) \) such that \( X \) is not a fixed point, this dynamic ascends the potential function \( P : \mathbb{X}(C) \rightarrow \mathbb{R} \), where \( P(X) = X \cdot CX \). In the Appendix, we prove that \( T \) (cf. Section 3) is a growth transformation for \( P \), \( \forall \alpha > 0 \). We ensure that the iterates of our dynamical system remain \( G \)-feasible and to that end we design a growth transformation for \( G \). To that end, we extend a result by Gopalakrishnan et al. [1991], namely, that rational functions admit growth transformations, which can be obtained by growth transformations for corresponding polynomials. They show that:
Proposition 5. Given a domain

\[ D = \left\{ X \in \mathbb{R} \bigg| \sum_{j=1}^{q_i} X_{ij} = 1 \quad X_{ij} \geq 0 \quad i = 1, \ldots, p \quad j = 1, \ldots, q_i \right\} \]

and a rational function \( R(X) = \frac{S_1(X)}{S_2(X)} \) where \( S_1(X) \) and \( S_2(X) \) are polynomials with real coefficients and \( S_2(X) \) has only positive values in \( D \), for any \( X \in D \), there exists a polynomial \( P_q(X)_{q \leftarrow X} \) parametrized by \( X \) such that

\[ P_q(Y)_{q \leftarrow X} > P_q(X)_{q \leftarrow X} \Rightarrow R(Y) > R(X) \]

and for this it is enough to set

\[ P_q(X)_{q \leftarrow X} = S_1(X) - R(q)_{q \leftarrow X} S_2(X). \]

At the heart of Ariadne lies a (primary) dynamical system based on Hedge that implements a growth transformation for the primary barrier function \( G \). There is also a secondary dynamical system based on the same principles of operation that implements a growth transformation for the secondary barrier function \( G_S \). These dynamical systems are designed such that they ascend the respective barrier function—iterates remain feasible in virtue of this property.

Lemma 8. Given a \( G \)-feasible \( X \), there exists an \( n \times n \) positive matrix \( \bar{C}_X \), which we call the operative matrix at \( X \), that depends on \( X \), such that

\[ T_G(X)_{i} = X(i) \frac{\exp \left\{ \alpha (\bar{C}_X X)_i \right\}}{\sum_{j=1}^{n} X(j) \exp \left\{ \alpha (\bar{C}_X X)_j \right\}} \quad i = 1, \ldots, n \]

is a growth transformation for \( G \). Furthermore if \( X^0 \) is \( G \)-feasible, repeatedly applying \( T_G \), the iterates \( \{X^k\} \) remain \( G \)-feasible.

Proof. To prove the second part of the lemma, we prove that \( \hat{X} \in \mathbb{F}_L \) where \( \hat{X} = T_G(X) \). Along the way, we prove the first part of the lemma. We may write the barrier function as

\[ G(X) = \frac{X \cdot CX - C}{\max_{i=1}^{n} \{(CX)_i - C\ell\} \prod_{i=1}^{n} (C - (CX)_i)} \]

\[ = \frac{X \cdot CX - C}{\max_{Y \in \mathbb{Y}} \{Y \cdot CX - C\ell\} \prod_{i=1}^{n} (C - (CX)_i)} \]

\[ = \frac{X \cdot CX - C}{(Y \cdot CX - C\ell) \prod_{i=1}^{n} (C - (CX)_i)} \]

\[ = G(X|Y) \]

where \( Y \) is a best response to \( X \)[5]. To prove that \( X \in \mathbb{F}_L \) implies \( \hat{X} \in \mathbb{F}_L \) we first prove that

\[ G(\hat{X}) = G(\hat{X}|\hat{Y}) > G(\hat{X}|Y) > G(X|Y) = G(X). \]

[5] If the best response \( Y \) is not unique, a variety of rules can be used to compute a best response such as selecting any one of them—see [Bertsekas et al., 2003, pp. 245-7] for the computation of the subdifferential.
Our proof extends an idea of Gopalakrishnan et al. \cite{1991} (cf. Proposition 5) who reduce the problem of devising a growth transformation for a rational function to one of devising a growth transformation for a corresponding polynomial. In our particular problem, their methodology stipulates that a growth transformation for the polynomial

\[ P_G(X|Y) = (X \cdot CX - C - G(q)|Y)|_{q \leftarrow X} (Y \cdot CX - C |) \prod_{\ell=1}^{n} (C - (CX)_{\ell}), \]

where \( Y \) is a best response to \( X \), is also a growth transformation for \( G(X|Y) \). Growth transformations for \( P_G(X|Y) \) are based on its gradient, which assumes the expression

\[ \frac{\partial P_G(X|Y)}{\partial X(i)} = (CX)_i - G(X) \left( \prod_{\ell=1}^{n} (C - (CX)_{\ell}) \right) (CY)_i + G(X) \left( \sum_{i=1}^{n} \{((CX)_i - C_{\ell}) \sum_{m=1}^{n} \left( \prod_{\ell=1, \ell \neq m}^{n} (C - (CX)_{\ell}) \right) C_{im} \right) \]

and which is equal to the gradient of

\[ Q(X|Y) = X \cdot CX - G(q) \left( \prod_{l=1}^{n} (C - (Cq)_{l}) \right) |_{q \leftarrow X} Y \cdot CX + G(q) \left( \sum_{i=1}^{n} \{((Cq)_i - C_{\ell}) \sum_{m=1}^{n} \left( \prod_{l=1, l \neq m}^{n} (C - (Cq)_{l}) \right) \right) |_{q \leftarrow X} (CX)_m. \]

To find a growth transformation for \( P_G(X|Y) \) it suffices to find a growth transformation for \( Q(X|Y) \). The advantage of this equivalence is that \( Q(X|Y) \) can be expressed in the form of a homogeneous quadratic function using a trick by Bomze \cite{1998}. We may thus write

\[ Q(X|Y) = X \cdot \bar{C}X, \]

where \( \bar{C} \) is a square matrix. We may add a positive constant to the entries of this matrix and then normalize with a positive scalar to obtain a positive matrix \( \bar{C}_X \), which we call the operative matrix at \( X \). In Lemma \ref{19} in the appendix, we prove that applying \( T \) on \( X \) using \( \bar{C}_X \), unless \( X \) is a fixed point corresponding to \( \bar{C}_X \), we obtain \( \tilde{X} \) such that \( \tilde{X} \cdot \bar{C}_X \tilde{X} > X \cdot \bar{C}_X X \). This implies

\[ G(\tilde{X}|Y) = G(TG(\tilde{X}|Y)|Y) > G(X|Y). \tag{6} \]

Keeping now \( \tilde{X} \) fixed and considering the polynomial \( P_G(\tilde{X}|Y) \) in the variable \( Y \) this time, we have

\[ P_G(\tilde{X}|Y) = (\tilde{X} \cdot C \tilde{X} - C - G(\tilde{X}|q)|_{q \leftarrow Y} (Y \cdot C \tilde{X} - C_{\ell}) \prod_{\ell=1}^{n} (C - (C \tilde{X})_{\ell}). \]

Maximizing with respect to \( Y \) weakly increases \( P_G(\tilde{X}|Y) \) and, therefore, also weakly increases \( G(\tilde{X}|Y) \), implying that

\[ G(\tilde{X}|\tilde{Y}) = \max_{Y \in \mathbb{Y}} \{ G(\tilde{X}|Y) \} \geq G(\tilde{X}|Y). \]
and, therefore, combining with (6), proving our claim (5) (which implies that $T_G$ is a growth transformation for the barrier function $G$). Consider now the barrier function

$$G'(X) = \frac{X \cdot CX - C}{\left( \max_{i=1}^{n} \{(CX)_i\} - C_\ell \right) \left( C - \max_{i=1}^{n} \{(CX)_i\} \right) \prod_{i=1, i \neq \max}^{n} |C - (CX)_i|},$$

where $|\cdot|$ is the absolute value. Any growth transformation for $G'$ satisfies the property that if $X$ is $G$-feasible, then $\tilde{X}$ is also $G$-feasible. But $T_G$ is a growth transformation for $G'$, which implies that if $X^0$ is $G$-feasible, repeatedly applying $T_G$, the iterates $\{X^k\}$ remain $G$-feasible, completing the proof.

**Lemma 9.** Given a $G_S$-feasible $X$, there exists an $n \times n$ positive matrix $\bar{C}_X$, which we call the secondary operative matrix at $X$, that depends on $X$, such that

$$T_{G_S}(X)_i = X(i) \frac{\exp \left\{ \alpha(\bar{C}_X X)_i \right\}}{\sum_{j=1}^{n} X(j) \exp \left\{ \alpha(\bar{C}_X X)_j \right\}} \quad i = 1, \ldots, n$$

is a growth transformation for $G_S$. Furthermore if $X^0$ is $G_S$-feasible, repeatedly applying $T_{G_S}$, the iterates $\{X^k\}$ remain $G_S$-feasible.

**Proof.** Analogous to the proof of Lemma 5.

**4.2.4 Fixed points of our dynamical system and a pair of fundamental properties**

**Lemma 10.** The fixed points of $T_G$ are the pure strategies and fixed points of the replicator dynamic. The same property continues to hold for the secondary map $T_{G_S}$.

**Proof.** Let $J : \mathbb{Y} \rightarrow \mathbb{Y}$ be the discrete-time replicator dynamic, that is,

$$J(X)_i = X(i) \frac{(CX)_i}{X \cdot CX} \quad i = 1, \ldots, n.$$ 

Consider the synthesis of $J$ and $T_G$, that is, the map $J \circ T_G$. Then

$$\{X | (J \circ T_G)(X) = X \} \subset \{X | J(X) = X \}$$

and

$$\{X | (J \circ T_G)(X) = X \} \subset \{X | T_G(X) = X \}.$$ 

Therefore

$$\{X | (J \circ T_G)(X) = X \} \subset \{X | J(X) = X \} \cap \{X | T_G(X) = X \}.$$ 

That is, the fixed points of $J \circ T_G$ are the intersection of the fixed points of $J$ and those of $T_G$.

We would like to show that

$$\{X | T_G(X) = X \} \subset \{X | J(X) = X \}.$$
By the first part of the proof it suffices to show that
\[ \{ X | T_G(X) = X \} \subset \{ X | (J \circ T_G)(X) = X \}. \]

Let \( X \) be such that
\[ (J \circ T_G)(X) \neq X. \]

To prove the lemma, it suffices to show that
\[ T_G(X) \neq X. \]

Note that \( J \) is a diffeormorphism [Losert and Akin [1983], Theorem 4] (Theorem 4 in that paper assumes \( C > 0 \)) and, therefore, invertible. Let \( J^{-1}(X) \) denote the inverse of \( X \) under \( T \). If \( X \) is not a fixed point of \( J \), we obtain
\[ (J^{-1} \circ J \circ T_G)(X) \neq J^{-1}(X) \neq X \]
which implies
\[ T_G(X) \neq X \]
as claimed. If \( X \) is a fixed point of \( J \), we obtain \( J^{-1}(X) = X \) and, therefore, that
\[ (J^{-1} \circ J \circ T_G)(X) \neq X \]
which implies
\[ T_G(X) \neq X \]
as claimed. The proof that the same property continues to hold for \( T_G \) is analogous.

**Lemma 11.** In a connected graph \( G \), an upper bound on the maximum potential value of a fixed point of the replicator dynamic that is not a clique is
\[ 1 - \frac{1}{2(\omega(G) - 1)} \]
where \( \omega(G) \) is the clique number of \( G \).

**Proof.** Our proof is by strong induction on the number of edges of \( G \). The basis of the induction corresponds to one edge in a graph of two vertices. The potential value at each of the vertices is 1/2. This proves the induction basis. The induction hypothesis corresponds to the case of a connected graph of \( m \) edges. Let us add one more edge to the graph. Observe now that the potential value of a fixed point of the replicator dynamic that is not a clique depends only on the configuration of edges in its carrier (and it is independent of the configuration of edges outside the carrier): To see this, let \( X \) be a fixed point of the replicator dynamic, let \( C(X) = \{i_1, \ldots, i_k\} \) be the carrier \( X \) and let us renumber pure strategies such that the carrier corresponds to \( \{1, \ldots, k\} \). We may then write
\[ X \cdot C X = \begin{bmatrix} \hat{X} & 0 \end{bmatrix} \begin{bmatrix} \hat{C} & C^{(01)} \\ C^{(10)} & C^{(11)} \end{bmatrix} \begin{bmatrix} \hat{X} \\ 0 \end{bmatrix} = \hat{X} \cdot \hat{C} \hat{X}. \]
In virtue of this observation, strong induction implies the induction step completing the proof. \( \Box \)
Lemma 12. Let $C$ be an arbitrary square payoff matrix. Then the function $F : \mathbb{X}(C) \to \mathbb{R}$, where $F(X) = (CX)_{\text{max}}$, is convex.

Proof. We may write $F$ as

$$F(X) = \max_Y \{ Y \cdot CX \}.$$ 

By a basic property of the maximum function, we obtain for all $X \neq X'$ where $X, X' \in \mathbb{X}(C)$,

$$F((1 - \epsilon)X + \epsilon X') = \max_Y \{ Y \cdot C((1 - \epsilon)X + \epsilon X') \} \leq (1 - \epsilon) \max_Y \{ Y \cdot CX \} + \epsilon \max_Y \{ Y \cdot CX' \}.$$ 

Thus, $F$ is convex as claimed.

Lemma 13. If the Nisan parameter is equal to the clique number, every equilibrium of $C^+$ is a global minimizer of $F : \mathbb{X}(C^+) \to \mathbb{R}$ where $F(X) = (C^+X)_{\text{max}}$.

Proof. Let us first first show that $E_0$ is a global minimizer of $F$. We will show a stronger property that $E_0$ is global minimizer of $(C^+X)_{\text{max}}$ over all $X$ in the hyperplane

$$X(0) + X(1) + \cdots + X(n) = 1.$$ 

Following [Bertsekas, 1999, pp. 331-332], a necessary condition for $E_0$ to be a local minimizer of $(C^+X)_{\text{max}}$ over the previous hyperplane is that there exists a vector $\mu$ such that $\mu \geq 0$ and $\sum \mu_i = 1$ and a scalar $\lambda$ such that

$$C^+ \mu + \lambda 1 = 0.$$ 

Letting $\mu = E_0$ and $\lambda = -C_{00}$ satisfies these conditions. Thus, since, following the proof of Lemma 12, $(C^+X)_{\text{max}}$ is a convex function the aforementioned necessary condition is also sufficient, which implies that $E_0$ is a global minimizer of $(C^+X)_{\text{max}}$ over the previous hyperplane and, therefore, also of $F$. Since any maximum-clique equilibrium, say $X^*$, satisfies $(C^+X^*)_{\text{max}} = C_{00}$, every equilibrium of $C^+$ is a global minimizer of $F$ as claimed. This completes the proof of the lemma.

Lemma 14. Suppose the underlying graph $G$ is connected and that the Nisan parameter $k$ satisfies $k = \omega(G)$. Then, if

$$C_\ell = \frac{1}{2} \left( 1 - \frac{1}{2(k - 1)} + 1 - \frac{1}{2k} \right),$$

we have that

$$\min \left\{ \max_{i=1}^{n} \{(CX)_i \} \mid X \text{ is a fixed point of the replicator dynamic in } \overline{F_L} \right\} \geq C_{00},$$

where $\overline{F_L}$ is the closure of $F_L$.

Proof. Our proof is by strong induction on the number of edges of $G$. The basis of the induction corresponds to one edge in a graph of two vertices, where we have

$$\min \left\{ \max_{i=1}^{n} \{(CX)_i \} \mid X \text{ is a fixed point of the replicator dynamic in } \overline{F_L} \right\} = \min \{1, 1, 3/4\} \geq C_{00} = 3/4.$$
This proves the induction basis. The induction hypothesis corresponds to the case of a connected graph of $m \leq n$ edges. Let us add one more edge to $G$. If $G$ becomes complete, the lemma follows by Lemmas 12 and 13, which imply there are no other fixed points in $F_L$ besides the maximum-clique equilibrium. Otherwise, the lemma follows by eliminating all edges which are not part of a maximum clique, invoking the induction hypothesis, and observing that putting back the edges that were eliminated may only increase the maximum payoff of fixed points that are not maximum-clique equilibria.

4.2.5 Asymptotic convergence to a maximum-clique equilibrium

We do not attempt to identify the minimal assumptions on the learning rate of $T_G$ to enable convergence to an equilibrium. Under a constant learning rate, convergence is certainly attained.

**Proposition 6** ([Losert and Akin 1983]). Suppose a discrete time dynamical system obtained by iterating a continuous map $F : \Delta \to \Delta$ admits a Lyapunov function $G : \Delta \to \mathbb{R}$, i.e., $G(F(p)) \geq F(p)$ with equality at $p$ only when $p$ is an equilibrium. The limit point set $\Omega$ of an orbit $\{p(t)\}$ is then a compact, connected set consisting entirely of equilibria and upon which $G$ is constant.

**Lemma 15.** If the Nisan parameter $k$ is equal to the clique number $\omega(G)$ and

$$C = \frac{1}{2} \left( \left( 1 - \frac{1}{2(k - 1)} \right) + \left( 1 - \frac{1}{2k} \right) \right),$$

then the sequence $\{X^k\}$ of iterates converges to a maximum-clique equilibrium.

**Proof.** [Losert and Akin 1983] show that under the assumption the Lyapunov function $G$ is continuous and strictly (monotonically) increasing every limit point of an orbit under $F$ is a fixed point of $F$ upon which $G$ is constant (even if $F$ is discontinuous as is the case for our primary dynamical system). Our Lyapunov function, $G$, is strictly monotonically increasing and in virtue of Lemma 14, the set of maximum clique equilibria are the unique fixed points of $T_G$ upon which $G$ assumes the value 0. Therefore, had Ariadne been such that the secondary system were not activated, every limit point of the sequence of iterates would have been a maximum-clique equilibrium. The goal of activating the secondary system (in lieu of the primary) is to ensure convergence to a maximum clique. Once $G$ is sufficiently close to 0, the secondary barrier function $G_S$ and dynamical system is activated, giving rise to a continuous system, and in spite of the decrease in the value of the barrier function (corresponding to a single iteration) from that point on we can use the full potential of Proposition 6, which implies that the set of limits points of the secondary dynamical system is connected. Since maximum-clique equilibria are isolated fixed points, we obtain that the sequence of iterates converges to a maximum-clique equilibrium, completing the proof of the lemma.

4.3 Ariadne: Our maximum-clique computation algorithm

Ariadne looks for a maximum clique starting with a large value of the Nisan parameter $k$ (possibly the largest, however, upper bounds on the clique number, e.g., [Pardalos and Philips 1990], can reduce the search space for the appropriate value of the Nisan parameter) and subtracts one from this parameter upon failure to compute a clique of size equal to $k$. If the Nisan parameter is equal to the clique number $\omega(G)$, Ariadne is guaranteed to compute a maximum clique.

20
4.3.1 The process of initialization of our dynamical system

Initializing our dynamical system involves configuring four parameters, namely, the equilibrium approximation error $\epsilon$, $C_\ell$, the initial condition $X^0$, and the learning rate $\alpha$. The equilibrium approximation error $\epsilon$ is set equal to

$$\epsilon = \epsilon_a^2/8$$

where $\epsilon_a = 1/2 \left( \frac{1}{k-1} - \frac{1}{k} \right)$

and $C_\ell$, the lower bound on the maximum payoff, is set equal to

$$C_\ell \leftarrow \frac{1}{2} \left( 1 - \frac{1}{2(k-1)} + 1 - \frac{1}{2k} \right) \equiv \frac{1}{2} \left( 1 - \frac{1}{2(k-1)} + C \right),$$

where $k$ is the Nisan parameter. The initial condition $X^0$ is set such that it satisfies

$$\max_{i=1}^n \{(CX^0)_i\} = \frac{1}{2} (C + C_\ell).$$

Such $X^0$ can be computed by selecting a strategy $E_i$ (any pure strategy is necessarily strictly upper feasible), finding a weakly feasible strategy that minimizes the Euclidean distance between $E_i$ and the boundary of lower and upper feasibility sets, finding a strategy that minimizes the Euclidean distance between $E_i$ and the infeasibility set, and then performing binary search between the corresponding strategies. Thus initialization requires solving two convex quadratic programming problems. The learning rate remains constant throughout the execution of the dynamical system.

4.3.2 Following successful initialization

Upon successful initialization, Ariadne generates an a priori specified number of iterates (a number whose derivation is given in the next section), computes an approximate well-supported equilibrium (cf. Proposition 1) using the empirical average of a corresponding sequence of “multipliers” (which is obtained from the sequence of iterates using a transformation as defined below) and checks if the carrier of the well-supported equilibrium carries a fixed point of the replicator dynamic and if that fixed point is the characteristic vector of a clique of size equal to the Nisan parameter. Upon failure to detect such a fixed point of the replicator dynamic, Ariadne sets $k \leftarrow k - 1$ (that is, it decreases the value of the Nisan parameter by one) and repeats the process. If a fixed point of the replicator dynamic that is a clique of size equal to the Nisan parameter has been detected, Ariadne outputs that clique as a maximum clique and terminates. We should note that, as discussed earlier, if the Nisan parameter is equal to the clique number, Ariadne switches barrier functions as the sequence of iterates assumes values of the primary barrier function close to zero. The mechanism by which the secondary barrier function replaces the primary is simple: Upon detection of an iterate whose potential value is greater than $C_\ell$ the secondary barrier function replaces the primary and the dynamical system is correspondingly replaced from its discontinuous version to the continuous one, which has favorable asymptotic convergence properties. Upon activation of the secondary dynamical system in lieu of the primary, we can infer that the Nisan parameter is equal to $\omega(G)$. We note that had the algorithm stopped here we would have been able to prove $\text{NP} = \text{coNP}$.

4.3.3 The sequence of “multipliers”

For any given value of the Nisan parameter, Ariadne computes an upper bound on the number of iterations that are sufficient to compute a maximum clique so that if this upper bound is exceeded
and a maximum clique has not been computed, the iterations of the dynamical system terminate, the Nisan parameter is decreased by one, and the dynamical system is initialized again for the new lower value of the Nisan parameter. This upper bound does not apply to the number of iterations of $T_G$ directly but rather to the empirical average of a sequence of multipliers, that is, mixed strategies that are obtained by transforming the iterates according to the following process: Let us denote by $X$ the current iterate and by $\hat{X}$ the next iterate. The definition of a multiplier depends on the definition of the operative matrix matrix at $X$ (cf. Lemmas 8 and 9). Using the (primary or secondary) operative matrix, the next iterate our dynamical system generates can be obtained as

$$\hat{X}(i) = X(i) \frac{\exp\{\alpha(\bar{C}_X X)_i\}}{\sum_{j=1}^n X(j) \exp\{\alpha(\bar{C}_X X)_j\}} \quad i = 1, \ldots, n.$$  

The multiplier $Y$ of $X$ is a strategy $Y$ in $Y$ such that $CY = \bar{C}_X X$. Such a strategy can be obtained by either inverting $C$ or equivalently solving a linear feasibility program. Ariadne solves this linear program in every iteration and computes the empirical average of the sequence of multipliers as it is for this sequence that our fixed-point (and equilibrium) approximation bounds apply.

### 4.3.4 A note on the numerical stability of Ariadne

In the course of numerically testing a related algorithm we observed a phenomenon whereby Hedge did not increase the potential function. The iterate was close to a fixed point and the value of the learning rate was large. We believe this phenomenon can be attributed to a numerical roundoff error (possibly to the implementation of the exponential function). Further work is required to understand how our algorithm interacts with commodity software and hardware systems and to make our algorithm backward compatible with the implementation of these systems.

### 5 Computation of a maximum clique requires polynomial time

In this final section, we complete the proof that $P = NP$ by discussing how to configure the approximation error of the dynamical system and analyzing the complexity of its execution.

#### 5.1 On the “minimum positive gap” of a symmetric bimatrix game

Our goal in this section is to define a concept that is able to transform the equilibrium approximation algorithm based on Hedge to a polynomial computation algorithm in the Nisan game. But let us start more generally with the setting of symmetric bimatrix games: Let $C$ be a symmetric bimatrix game and $X \in X(C)$. We may give a preliminary definition of the gap $\Gamma_C(X)$ of $X \in X(C)$ as

$$\Gamma_C(X) \equiv (CX)_{\text{max}} - (CX)_{\text{min}}.$$  

Our motivation for introducing this definition has as follows: Every pure or mixed strategy of a symmetric bimatrix game $C$ has a gap (except for equalizers). One way to define a “minimum gap” is as the minimum over all strategies of $C$. But $C$ has sub-games. The sub-games that are carriers of fixed points have minimum gap of zero. Sub-games that do not carry fixed points also have a positive minimum gap (as sub-games). It is meaningful that in the definition of the minimum gap we take the sub-games into account and here is why: Let us extend the previous preliminary definition and define the extended gap $\Gamma_{CC'}(X)$ of $X \in X(C)$ as

$$\Gamma_{CC'}(X) \equiv (CX)_{\text{max}} - (C'X)_{\text{min}}.$$  

22
where $C'$ is a subgame of $C$ (padded with zeros so that the dimensions of $C$ and $C'$ agree. Furthermore, define the the minimum positive gap of $C$, call is $\gamma_{\text{min}}(C)$ as

$$
\gamma_{\text{min}}(C) \equiv \min_{C'} \left\{ \min_{X \in \mathcal{X}(C)} \{ \Gamma_{CC'}(X) | \Gamma_{CC'}(X) > 0 \} \right\},$

where the first minimization is taken over all subgames of $C$. We claim that a $\gamma_{\text{min}}/2$-well-supported equilibrium, call it $\hat{X}$, lies inside the carrier of an equilibrium (which we can readily compute knowing the carrier). Let us assume for the sake of contradiction that the carrier of $\hat{X}$ does not carry an equilibrium (which is an equalizer of the carrier). Then with Proposition 3 in mind there is a gap equal to or greater than $\gamma_{\text{min}}$ (inside the carrier), which is an impossibility given that $\gamma_{\text{min}}/2$-well-supported equilibrium exists. Hence the claim. In the sequel, we are concerned with the the Nisan game. In this game, a related to the above but more appropriate, in that it simplifies the analysis, definition of gap is as follows:

$$
\hat{\Gamma}(X) \equiv \max \{ X \cdot C X | X \in \mathcal{X}(C) \} - X \cdot C X
$$

where $C$ is the Nisan-Bomze payoff matrix barring strategy 0. We may then define the minimum positive gap as

$$
\hat{\gamma}_{\text{min}} \equiv \min_{C'} \left\{ \min_{X \in \mathcal{X}(C')} \{ \hat{\Gamma}(X) | \hat{\Gamma}(X) > 0 \} \right\}
$$

where the first minimization is taken over all subgames of $C$. As above, we claim that a $\gamma_{\text{min}}/2$-well-supported equilibrium, call it $\hat{X}$, lies inside the carrier of an equilibrium (which we can readily compute knowing the carrier) provided $E_0$ is a GNSS of the Nisan game but not a GESS so that a maximum-clique equilibrium exists. Let us assume for the sake of contradiction that the carrier, say $\hat{C}'$ of $\hat{X}$ does not carry an equilibrium (which is an equalizer of the carrier, where the term equalizer is to be understood with the latest definition of gap). Then, keeping again Proposition 3 in mind, there is a gap equal to or greater than $\gamma_{\text{min}}$ (inside the carrier), that is,

$$
\max \{ X \cdot \hat{C}' X | X \in \mathcal{X}(\hat{C}') \} - X \cdot \hat{C}' X = (\hat{C}' X)_{\text{max}} - X \cdot \hat{C}' X \geq \gamma_{\text{min}}
$$

which is an impossibility given that $\gamma_{\text{min}}/2$-well-supported equilibrium exists implying

$$(\hat{C}' X)_{\text{max}} - (\hat{C}' X)_{\text{min}} \leq \gamma_{\text{min}}/2.$$ 

Hence the claim. We have the following theorem:

**Theorem 1.** The minimum positive gap $\gamma_{\text{min}}$ of $C$ is at least

$$
\left( 1 - \frac{1}{2k} \right) - \left( 1 - \frac{1}{2(k-1)} \right) = \frac{1}{2(k-1)} - \frac{1}{2k}
$$

where $k$ is the size of the maximum clique.

**Proof.** Let $\hat{C}'$ be a subgame of $C$. Then

$$
\min_{X \in \mathcal{X}(\hat{C}')} \{ \hat{\Gamma}(X) \} = \max \{ X \cdot \hat{C} X | X \in \mathcal{X}(\hat{C}) \} - \max \{ X \cdot \hat{C} X | X \in \mathcal{X}(\hat{C}') \}
$$

where

$$
\max \{ X \cdot \hat{C}' X | X \in \mathcal{X}(\hat{C}') \} = 1 - \frac{1}{2k'}
$$

23
where \( k' \) is the maximum clique in the subgraph corresponding to \( \hat{C}' \). The maximum possible clique smaller than the maximum clique has one vertex less. Therefore, in light of Lemma \( \text{[11]} \)

\[
\hat{\gamma}_{min} \equiv \min_{C'} \left\{ \min_{X \in \mathcal{X}(C')} \left\{ \hat{\Gamma}(X) | \hat{\Gamma}(X) > 0 \right\} \right\} \geq \left( 1 - \frac{1}{2k} \right) - \left( 1 - \frac{1}{2(k-1)} \right)
\]

as claimed. \( \square \)

5.2 Computational complexity of the dynamical system’s execution

**Lemma 16.** If the Nisan parameter is equal to or greater than the clique number and the sequence of iterates converges to a fixed point that is not a pure strategy, the sequence of multipliers converges to a maximal clique equilibrium in a polynomial number of iterations in the inverse of the approximation error if the learning rate in each iteration is lower bounded by a constant.

**Proof.** If the Nisan parameter is equal to or greater than the clique number and the sequence of iterates converges to a fixed point that is not a pure strategy, Theorem \( \text{[3]} \) together with Lemma \( \text{[21]} \) imply that

\[
\frac{2\text{nd max}}{p \in \{1, \ldots, n\}} \left\{ (CY^K)_p \right\} - Y^K \cdot CY^K \leq \ln \left( \frac{2\text{nd max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) \frac{2c}{A_K},
\]

where

\[
c = \frac{\max_{ij} C_{ij}}{\min_{ij} C_{ij}},
\]

and Lemma \( \text{[20]} \) gives

\[
\frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ (CY^K)_p \right\} - \frac{2\text{nd max}}{p \in \{1, \ldots, n\}} \left\{ (CY^K)_p \right\} =
\]

\[
= \left( \ln \left( \frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) - \ln \left( \frac{2\text{nd max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) \right) \frac{1}{A_K}.
\]

Summing the previous inequalities, we obtain

\[
\frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ (CY^K)_p \right\} - \bar{Y}^K \cdot CY^K \leq \left( \ln \left( \frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) \right) + (2c - 1) \ln \left( \frac{2\text{nd max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) \frac{1}{A_K},
\]

which implies

\[
\frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ (CY^K)_p \right\} - \bar{Y}^K \cdot CY^K \leq \left( \ln \left( \frac{1\text{st max}}{p \in \{1, \ldots, n\}} \left\{ \frac{X^{K+1}(p)}{X^0(p)} \right\} \right) \right) \frac{2c}{A_K}.
\]

If the learning rate is lower bounded by \( \alpha \) we obtain that \( A_K \geq \alpha(K+1) \) and the lemma follows. \( \square \)

**Lemma 17.** If the sequence of multipliers \( \{Y^K\} \) converges, then the sequence of empirical averages \( \{\bar{Y}^K\} \) also converges to the same limit.
Proof. Assume the sequence \( \{Y^k\} \) converges and let \( X^* \) be its limit. Then

\[
\lim_{k \to \infty} \|X^* - Y^k\| = 0,
\]

where \( \| \cdot \| \) is the Euclidean norm. Then the Stolz-Cesáro theorem implies that

\[
\lim_{K \to \infty} \left\{ \frac{1}{A_K} \sum_{k=0}^{K} \alpha_k \|X^* - Y^k\| \right\} = 0. \tag{9}
\]

The convexity of the Euclidean distance function gives that

\[
\|X^* - \bar{Y}^K\| \leq \frac{1}{A_K} \sum_{k=0}^{K} \alpha_k \|X^* - Y^k\|. \tag{10}
\]

(9) and (10) together imply

\[
\lim_{K \to \infty} \|X^* - \bar{Y}^K\| = 0.
\]

Thus \( \{\bar{Y}^K\} \) also converges to \( X^* \) as claimed. \( \square \)

**Theorem 2.** If the Nisan parameter is equal to the clique number and the sequence of iterates converges to a maximum clique equilibrium, the sequence of multipliers also converges to a maximum clique equilibrium in a polynomial number of iterations in the inverse of the approximation error if the learning rate in each iteration is lower bounded by a constant.

**Proof.** If the Nisan parameter is equal to the clique number, since the sequence of iterates converges, the sequence of multipliers (as well as their average) also converges to the same limit. This is an implication of method that generates the sequence of multipliers from the sequence of iterates (cf. Section 4.3, in particular, Section 4.3.3) and Lemma 7 and as for the convergence of the (weighted) empirical average of the multipliers it is an implication of Lemma 17. Lemma 16 implies the polynomial bound on the number of iterations in the inverse of the approximation error. \( \square \)

If the fixed-point approximation error is set such that

\[
\epsilon = \frac{\epsilon^2_a}{8} \text{ where } \epsilon_a = \frac{1}{2} \left( \frac{1}{\omega(G) - 1} - \frac{1}{\omega(G)} \right),
\]

then, provided the Nisan parameter is equal to the clique number, on attainment of this approximation error, the corresponding approximate well-supported equilibrium (cf. Section 2.3) is in the carrier of a maximum-clique equilibrium, which implies a maximum-clique can be readily computed. Therefore, starting at any value of the Nisan parameter greater than the clique number and following the algorithmic steps of Ariadne, a maximum clique is guaranteed to be computed:

- If the Nisan parameter is equal to \( \omega(G) \) our dynamical system is ensured to generate a sequence of iterates that are able to compute a maximum-clique equilibrium and, thus, compute a maximum clique. As shown above, the number of iterations required to do so is polynomial.

- If the Nisan parameter is greater than \( \omega(G) \), there is an a priori fixed bound on the number iterations of our dynamical system and upon failure to detect a maximum clique (that is, a clique equal to the Nisan parameter) when the number iterations has reached this upper bound, the Nisan parameter is decreased by one, implying that the Nisan parameter eventually becomes equal to the clique number and a maximum clique is computed.
5.3 Closing remarks and future work

In closing, we would like to point out that a feature that is not unique to Ariadne (for example, see [McCreesh and Prosser, 2013]) is that it admits a parallel implementation. We leave the details of such an implementation as future work. We would also like to raise the possibility that the research presented in this paper would benefit from the efficient computation of centroids or barycenters [Shephard and Webster, 1965; Shvartsman, 2004]. Although there exist efficient randomized algorithms for computing barycenters of convex bodies (for example, see [Bertsimas and Vempala, 2004]) it is an interesting question whether such algorithms can be derandomized (cf. [Rademacher, 2007]). We leave this as a question for future work. We would finally like to point out that our result can shed further light on the exact relationship among complexity classes using inapproximability results that have been obtained from the maximum clique problem (see [Wu and Hao, 2015] for a summary of these results). This is also an exciting question for future work.

Acknowledgments

This paper has benefitted from my interaction with my YouTube account and I thank those that are responsible for the configuration of the content in that account.

References

L. E. Baum and J. A. Eagon. An inequality with applications to statistical prediction for functions of Markov processes and to a model of ecology. *Bulletin of the American Mathematical Society*, 73:360–363, 1967.

L. E. Baum and G. R. Sell. Growth transformations for functions on manifolds. *Pacific Journal of Mathematics*, 27(2):211–227, 1968.

D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Athena Scientific, Belmont, Mass., 1996.

D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Mass., second edition, 1999.

D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, Massachusetts, 2003.

D. Bertsimas and S. Vempala. Solving convex programs by random walks. *Journal of the ACM*, 51(4):540–556, 2004.

I. M. Bomze. Evolution towards the maximum clique. *Journal of Global Optimization*, 10:143–164, 1997.

I. M. Bomze. On standard quadratic optimization problems. *Journal of Global Optimization*, 13:369–387, 1998.

I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. The maximum clique problem. In D.-Z. Du and P. M. Pardalos, editors, *Handbook of Combinatorial Optimization*, pages 1–74. Kluwer Academic Publishers, 1999.

X. Chen, X. Deng, and S. Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3), 2009.
C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM J. Comput.*, 39(1):195–259, 2009.

C. Daskalakis, R. Frongillo, C. H. Papadimitriou, G. Pierrakos, and G. Valiant. On learning algorithms for Nash equilibria. In Proc. 3rd International Symposium on Algorithmic Game Theory (SAGT 2010), 2010.

K. Etessami and A. Lochbihler. The computational complexity of evolutionarily stable strategies. *Int J Game theory*, 37:93–103, 2008.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

P. S. Gopalakrishnan, D. Kanevsky, A. Nadas, and D. Namahoo. An inequality for rational functions with applications to some statistical estimation problems. *IEEE Transactions on Information Theory*, 37:107–113, 1991.

J. Hofbauer, P. Schuster, and K. Sigmund. A note of evolutionary stable strategies and game dynamics. *J. theor. Biology*, 81:609–612, 1979.

R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of computer computations*, pages 85–103. Plenum, New York, 1972.

V. Losert and E. Akin. Dynamics of games and genes: Discrete versus continuous time. *J. Math. Biology*, 17:241–251, 1983.

J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

C. McCreesh and P. Prosser. Multi-threading a state-of-the-art maximum clique algorithm. *Algorithms*, 6:618–635, 2013.

T. S. Motzkin and E. G. Strauss. Maxima for graphs and a new proof of a theorem of turan. *Canadian Journal of Mathematics*, 17:533–540, 1965.

N. Nisan. A note on the computational hardness of evolutionary stable strategies. Report no. 76, Electronic Colloquium of Computational Complexity, 2006.

P. M. Pardalos and A. T. Philips. A global optimization approach for solving the maximum clique problem. *Intern. J. Computer Math.*, 33:209–216, 1990.

P. M. Pardalos and J. Xue. The maximum clique problem. *Journal of Global Optimization*, 4:301–328, 1994.

M. Pelillo and A. Jagota. Feasible and infeasible maxima in a quadratic program for maximum clique. *J. Artif. Neural Networks*, 2:411–420, 1995.

M. Pelillo and A. Torsello. Payoff-monotonic game dynamics and the maximum clique problem. *Neural Computation*, 18:1215–1258, 2006.
L. Rademacher. Approximating the centroid is hard. In Proc. of the 23rd Annual Symposium on Computational Geometry (SCG ’07), pages 302–305, 2007.

G. C. Shephard and R. J. Webster. Metrics for sets of convex bodies. Mathematika, 12:73–88, 1965.

P. Shvartsman. Barycentric selectors and a steiner-type point of a convex body in a banach space. Journal of Functional Analysis, 210:1–42, 2004.

P. Taylor and L. Jonker. Evolutionary stable strategies and game dynamics. Mathematical Biosciences, 16:76–83, 1978.

B. von Stengel. Equilibrium computation for two-player games in strategic and extensive form. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, Algorithmic Game Theory, chapter 3, pages 53–78. Cambridge University Press, 2007.

J. W. Weibull. Evolutionary Game Theory. MIT Press, 1995.

Q. Wu and J.-K. Hao. A review on algorithms for maximum clique problems. European Journal of Operational Research, 242:693–709, 2015.

A Hedge as a growth transformation

Our main result in this section is that Hedge is a growth transformation for all $\alpha > 0$. In our proof of this result, we follow [Baum and Eagon 1967]. We will need two auxiliary results:

**Proposition 7** (Hölder’s inequality). For all $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \mathbb{R}^n$, we have

$$
\sum_{k=1}^{n} |x_k| |y_k| \leq \left( \sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}} \left( \sum_{k=1}^{n} |y_k|^q \right)^{\frac{1}{q}}
$$

where

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p, q > 1.$$

**Proposition 8** (Weighted AM-GM inequality). Given nonnegative numbers $x_1, \ldots, x_n$ and non-negative weights $w_1, \ldots, w_n$, set $w = w_1 + \cdots + w_n$. If $w > 0$, then the inequality

$$
\frac{w_1 x_1 + \cdots + w_n x_n}{w} \geq (x_1^{w_1} \cdots x_n^{w_n})^\frac{1}{w}
$$

holds with equality if and only if all the $x_k$ with $w_k > 0$ are equal.

[Baum and Eagon 1967] show that:

**Proposition 9.** Let $P(X) = P(\{x_{ij}\})$ be a polynomial with nonnegative coefficients homogeneous in its variables $\{x_{ij}\}$. Let $x = \{x_{ij}\}$ be any point in the domain $D$, where

$$
D = \left\{ x \cup x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1, i = 1, \ldots, p, j = 1, \ldots, q_i \right\}.
$$
For $x = \{x_{ij}\} \in D$, let $J(X) = J(\{x_{ij}\})$ denote the point of $D$ whose $i,j$ coordinate is

$$J(x)_{ij} = x_{ij} \frac{\partial P}{\partial x_{ij}}(x) \bigg|_x \sum_{j=1}^{q_i} x_{ij} \frac{\partial P}{\partial x_{ij}}(x).$$

Then $J$ is a growth transformation for $P$, that is, $P(J(x)) > P(x)$ unless $J(x) = x$.

Let us summarize their notation (which we will follow in our own results): $\mu$ denotes a doubly indexed array of nonnegative integers: $\mu = \{\mu_{ij}\}$. $x^\mu$ is an abbreviation for

$$x^\mu \equiv \prod_{i=1}^p \prod_{j=1}^{q_i} x_{ij}^{\mu_{ij}}.$$

$c_\mu$ is an abbreviation for $c_{\mu_{ij}}$. Using the previous conventions,

$$P(x) \equiv \sum_\mu c_\mu x^\mu$$

and

$$J(x)_{ij} = \left( \sum_\mu c_\mu x^{\mu_{ij}} \right) \bigg/ \left( \sum_\mu c_\mu \prod_{i=1}^p \prod_{j=1}^{q_i} J(x)_{ij}^{\mu_{ij}} \right).$$

The goal is to prove that

$$P(x) = \sum_\mu c_\mu x^\mu \leq \sum_\mu c_\mu \prod_{i=1}^p \prod_{j=1}^{q_i} J(x)_{ij}^{\mu_{ij}}.$$

**Lemma 18.** Let $P(X) = P(\{x_{ij}\})$ be a polynomial homogeneous in its variables $\{x_{ij}\}$ of degree $d$. Let $x = \{x_{ij}\}$ be any point in the domain $D$, where

$$D = \left\{ x \mid x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1, i = 1, \ldots, p, j = 1, \ldots, q_i \right\}.$$

For $x = \{x_{ij}\} \in D$, let $J^k(X) = J^k(\{x_{ij}\})$ denote the point of $D$ whose $i,j$ coordinate is

$$J^k(x)_{ij} = x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k \sum_{j=1}^{q_i} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k$$

where $k$ is a positive integer. Then provided that

$$\forall i = 1, \ldots, p \forall j = 1, \ldots, q_i : \frac{\partial P}{\partial x_{ij}}(x) \leq 1,$$

$J^k$ is a growth transformation for $P$ for all $\alpha > 0$. Furthermore, for all positive integers $k$, the fixed points of $J^k$ coincide with the fixed points of $J$. 
Proof. We have

\[ P(x) = \sum_{\mu} c_\mu x^\mu \]

which we may equivalently write as

\[ = \sum_{\mu} (c_\mu)^{1/(dk+1)} (c_\mu)^{dk/(dk+1)} x^\mu \]

which we may equivalently write as

\[ = \sum_{\mu} (c_\mu)^{1/(dk+1)} (c_\mu)^{dk/(dk+1)} x^\mu \left( \prod_{i=1}^{p} \prod_{j=1}^{q_i} J^k(x)^{\mu_{ij}} \right)^{1/(dk+1)} \left( \prod_{i=1}^{p} \prod_{j=1}^{q_i} \left( \frac{1}{J^k(x)^{ij}} \right)^{\mu_{ij}} \right)^{1/(dk+1)} \]

and, rearranging terms, we obtain

\[ = \sum_{\mu} \left( c_\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} J^k(x)^{\mu_{ij}} \right)^{1/(dk+1)} \times \left( c_\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} \frac{1}{J^k(x)^{ij}} \right)^{\mu_{ij}/(dk+1)} \]

and using that

\[ x^{\mu + 1} = x^\mu x^{1} = x^\mu x^{1/(dk+1)} = x^{\mu} \prod_{i=1}^{p} \prod_{j=1}^{q_i} x^{\frac{\mu_{ij}}{dk+1}} \]

we further obtain

\[ \leq \left( \sum_{\mu} \left( c_\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} J^k(x)^{\mu_{ij}} \right)^{1/(dk+1)} \right) \left( \sum_{\mu} c_\mu x^\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} \left( \frac{x_{ij}}{J^k(x)^{ij}} \right)^{\mu_{ij}/dk} \right)^{1/(dk+1)} \]

which, using the weighted AM-GM inequality, yields

\[ \leq \left( \sum_{\mu} \left( c_\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} J^k(x)^{\mu_{ij}} \right)^{1/(dk+1)} \right) \left( \sum_{\mu} c_\mu x^\mu \left( \frac{1}{\sum_{i=1}^{p} \sum_{j=1}^{q_i} \frac{\mu_{ij}}{d}} \right) \sum_{i=1}^{p} \sum_{j=1}^{q_i} \frac{\mu_{ij}}{d} \left( \frac{x_{ij}}{J^k(x)^{ij}} \right)^{1/\frac{d}{dk+1}} \right)^{1/(dk+1)} \]

which, by the homogeneity of \( P \), implies

\[ = \left( \sum_{\mu} \left( c_\mu \prod_{i=1}^{p} \prod_{j=1}^{q_i} J^k(x)^{\mu_{ij}} \right)^{1/(dk+1)} \right) \left( \sum_{\mu} c_\mu x^\mu \sum_{i=1}^{p} \sum_{j=1}^{q_i} \frac{\mu_{ij}}{d} \left( \frac{x_{ij}}{J^k(x)^{ij}} \right)^{1/\frac{d}{dk+1}} \right)^{1/(dk+1)} \]  (12)
Let us work with the expression inside the parenthesis in the second product term. We, thus, substituting the expression for $J^k$, have

$$
\sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \frac{\mu_{ij}}{d} \left( \frac{x_{ij}}{J^k(x)_{ij}} \right)^{\frac{1}{k}} =
$$

$$
= \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \frac{\mu_{ij}}{d} \left( \frac{\sum_{\ell=1}^q x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}} \right)^{\frac{1}{k}}
$$

which, cancelling powers in the denominator, implies

$$
= \frac{1}{d} \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \sum_{\ell=1}^q \frac{\mu_{ij} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}}
$$

which, since

$$
\sum_{\ell=1}^q x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k > 1,
$$

implies that

$$
\leq \frac{1}{d} \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \sum_{\ell=1}^q \frac{\mu_{ij} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}}
$$

which further implies by assumption (11) that

$$
\leq \frac{1}{d} \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \sum_{\ell=1}^q \frac{\mu_{ij} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}}
$$

which, by rearranging terms, further implies

$$
= \left( \frac{1}{1 + \frac{1}{k} \alpha} \right) \frac{1}{d} \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \sum_{\ell=1}^q \frac{\mu_{ij} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{\frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}}
$$

which even further implies

$$
= \left( \frac{1}{1 + \frac{1}{k} \alpha} \right) \frac{1}{d} \sum_{\mu} c_{\mu} x^\mu \sum_{i=1}^p \sum_{j=1}^q \sum_{\ell=1}^q \frac{\mu_{ij} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \bigg|_{(x)} \right)^k}{x_{ij} \frac{\partial P}{\partial x_{ij}} \bigg|_{(x)}}
$$

31
which even further implies

\[
\left( \frac{1}{1 + \frac{1}{k} \alpha} \right)^{1/d} \sum_{\mu} c_{\mu} x^{\mu} \sum_{i=1}^{p} \sum_{j=1}^{q_{i}} \mu_{ij} x_{ij} \left( \frac{\sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) }{\sum_{\nu} c_{\nu} x^{\nu} \nu_{ij}} \right)^{k}
\]

which by rearranging the order of summation even further implies

\[
\left( \frac{1}{1 + \frac{1}{k} \alpha} \right)^{1/d} \sum_{i=1}^{p} \sum_{j=1}^{q_{i}} x_{ij} \sum_{\mu} c_{\mu} x^{\mu} \mu_{ij} \left( \sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) \right) ^{k}
\]

which, cancelling terms and summing probability masses, even further implies

\[
\left( \frac{1}{1 + \frac{1}{k} \alpha} \right)^{1/d} \sum_{i=1}^{p} \sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) ^{k}
\]

which, using the binomial theorem, implies

\[
\left( \frac{1}{1 + \frac{1}{k} \alpha} \right)^{1/d} \sum_{i=1}^{p} \sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) ^{m}
\]

which, using assumption (11), implies

\[
\leq \left( \frac{1}{1 + \frac{1}{k} \alpha} \right)^{1/d} \sum_{i=1}^{p} \sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) ^{m}
\]

and using the binomial theorem for a second time, we further obtain

\[
\leq \frac{1}{d} \sum_{i=1}^{p} \sum_{\ell=1}^{q_{i}} x_{i\ell} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{i\ell}} \right) = P(x)
\]

where the equality follows by the Euler theorem for homogeneous functions. Substituting in (12), we obtain

\[
P(x) \leq \left( \sum_{\mu} \left( c_{\mu} \prod_{i=1}^{p} \prod_{j=1}^{q_{i}} J^{k} (x)_{ij}^{\mu_{ij}} \right) \right)^{1/\alpha + 1} \left( P(x) \right)^{\frac{1}{\alpha + 1}}
\]

which implies

\[
(P(x))^{1 - \frac{1}{\alpha + 1}} \leq \sum_{\mu} \left( c_{\mu} \prod_{i=1}^{p} \prod_{j=1}^{q_{i}} J^{k} (x)_{ij}^{\mu_{ij}} \right)^{1/\alpha + 1}
\]

which finally implies

\[
P(x) \leq P(J(x))
\]
The strictness of the inequality if $J^k(x) \neq x$ follows from (12) and the strictness of the weighted arithmetic-geometric inequality in Proposition 8 if all summands are not equal.

Let us proceed with the proof of the second part of the lemma: We will show that $x$ is a fixed point if and only if for all $i = 1, \ldots, p$ and for all $j, \ell$ such that $x_{ij}, x_{i\ell} > 0$, we have that

$$\left. \frac{\partial P}{\partial x_{ij}} \right|_{(x)} = \left. \frac{\partial P}{\partial x_{i\ell}} \right|_{(x)}.$$

First we show sufficiency: Some of the coordinates of $x$ are zero and some are positive. Clearly, the zero coordinates will not become positive after applying the map. Now, notice that, given $i$, for all $j$ such that $x_{ij} > 0$,

$$\left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k = \sum_{j=1}^{q_i} x_{ij} \left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k$$

and, therefore, $J^k(x)_{ij} = x_{ij}$, and this is true for all $i$. Now we show necessity: If $x$ is a fixed point, then for all $i = 1, \ldots, p$ and for all $j, \ell$ such that $x_{ij}, x_{i\ell} > 0$, we have that

$$\left. \frac{\partial P}{\partial x_{ij}} \right|_{(x)} = \left. \frac{\partial P}{\partial x_{i\ell}} \right|_{(x)}.$$

Because $x$ is a fixed point, $J^k(x)_{ij} = x_{ij}$. Therefore,

$$J^k(x)_{ij} = x_{ij}$$

$$\frac{x_{ij}}{\sum_{j=1}^{q_i} x_{ij} \left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k} = x_{ij}$$

$$\left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k = \sum_{j=1}^{q_i} x_{ij} \left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k.$$

Equation (14) implies that, for all $i = 1, \ldots, p$ and for all $j$ such that $x_{ij} > 0$,

$$\left(1 + \frac{1}{k} \frac{\partial P}{\partial x_{ij}} \right)^k = c$$

where $c$ is a constant. Cancelling the power, the constant 1, and the factor $(1/k)\alpha$ yields the claim. Notice that the fixed points are independent of $k$. That these are also the fixed points of the replicator dynamic follows the same pattern (see also [Losert and Akin, 1983]).

**Lemma 19.** Let $P(X) = P(\{x_{ij}\})$ be a polynomial homogeneous in its variables $\{x_{ij}\}$ of degree $d$. Let $x = \{x_{ij}\}$ be any point in the domain $D$, where

$$D = \left\{ x \ | \ x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1, i = 1, \ldots, p, j = 1, \ldots, q_i \right\}.$$
For $x = \{x_{ij}\} \in D$, let $J^\infty(X) = J^\infty(\{x_{ij}\})$ denote the point of $D$ whose $i,j$ coordinate is

$$J^\infty(x)_{ij} = x_{ij} \frac{\exp \left\{ \alpha \frac{\partial P}{\partial x_{ij}}(x) \right\}}{\sum_{j=1}^{q_i} x_{ij} \exp \left\{ \alpha \frac{\partial P}{\partial x_{ij}}(x) \right\}},$$

where $\alpha > 0$. Then, provided that (11) holds, $J^\infty$ is a growth transformation for $P$.

**Proof.** We may equivalently write $J^\infty$ as

$$J^\infty(x)_{ij} = \frac{x_{ij} \lim_{k \to \infty} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k}{\sum_{j=1}^{q_i} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k},$$

which implies that the sequence of maps

$$J^k(x)_{ij} = \frac{x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k}{\sum_{j=1}^{q_i} x_{ij} \left( 1 + \frac{1}{k} \alpha \frac{\partial P}{\partial x_{ij}}(x) \right)^k},$$

converges pointwise to $J^\infty$. Therefore, by the definition of pointwise convergence, for each $x$ and an arbitrarily small $\epsilon > 0$, we can find $\hat{k} > 0$ such that for all $k \geq \hat{k}$, the maps $J^k$ map $x$ to within an $\epsilon$-ball of $J^\infty(x)$. It is easy to show that the fixed points of $J^\infty$ (cf. Lemma 5) coincide with the fixed points of $J$. Lemma 18 then implies that, for all $x$ and $\alpha > 0$, unless $x$ is a fixed point, $J^\infty$ strictly increases the value of the polynomial $P$. This completes the proof. \qed

**B. An inequality on the approximation error of multiplicative weights**

In this section, we analyze the fixed-point approximation error of the general map

$$T_i(X) = X(i) \cdot \frac{\exp \{\alpha (CY)_i\}}{\sum_{j=1}^{n} X(j) \exp \{\alpha (CY)_j\}}, \quad i = 1, \ldots, n,$$

in particular, the fixed point approximation error of the sequence $\{\bar{Y}^K\}_{K=0}^\infty$ of empirical averages of $\{Y^k\}$. The empirical average $\bar{Y}^K$ at iteration $K = 0, 1, 2, \ldots$ is a weighted arithmetic mean

$$\bar{Y}^K = \frac{1}{A_K} \sum_{k=0}^{K} \alpha_k Y^k, \quad \text{where} \quad A_K = \sum_{k=0}^{K} \alpha_k,$$

and $\alpha_k > 0$ is the learning rate parameter used in step $k$. In the case when the learning rate is held constant from round to round, the weighted arithmetic means reduces to a simple arithmetic mean

$$\bar{Y}^K = \frac{1}{K+1} \sum_{k=0}^{K} Y^k.$$
Lemma 20. Suppose $X^0$ is an arbitrary interior strategy. Then

$$\forall i, j \in \mathcal{K}(C) : (E_i - E_j) \cdot CY^K = \frac{1}{A_K} \ln \left( \frac{X^{K+1}(i)}{X^0(i)} \right) - \frac{1}{A_K} \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right).$$

Let $Y \in \mathcal{X}(C)$ be arbitrary. Then, for all $p \in \mathcal{C}(Y)$ the approximation error of the weighted empirical average is

$$(CY^K)_p - Y^K \cdot CY^K = \frac{1}{A_K} \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) - \frac{1}{A_K} \sum_{j=1}^{n} Y^K(j) \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right),$$

an expression which we may equivalently write as follows:

$$(CY^K)_p - Y^K \cdot CY^K = \frac{1}{A_K} \sum_{j=1}^{n} Y^K(j) \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) - \frac{1}{A_K} \sum_{j=1}^{n} Y^K(j) \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right).$$

Proof. Let $T(X) \equiv \hat{X}$. Then straight algebra gives

$$\frac{\hat{X}(i)}{\hat{X}(j)} = \frac{X(i)}{X(j)} \exp \{ \alpha((CY)_i - (CY)_j) \}$$

and taking logarithms on both sides we obtain

$$\ln \left( \frac{\hat{X}(i)}{\hat{X}(j)} \right) = \ln \left( \frac{X(i)}{X(j)} \right) + \alpha((CY)_i - (CY)_j).$$

We may write the previous equation as

$$\ln \left( \frac{X^{k+1}(i)}{X^{k+1}(j)} \right) = \ln \left( \frac{X^k(i)}{X^k(j)} \right) + \alpha_k((CY^k)_i - (CY^k)_j)$$

Summing over $k = 0, \ldots, K$, we obtain

$$\ln \left( \frac{X^{K+1}(i)}{X^{K+1}(j)} \right) = \ln \left( \frac{X^0(i)}{X^0(j)} \right) + \sum_{k=0}^{K} \alpha_k((CY^k)_i - (CY^k)_j)$$

and dividing by $A_K$ and rearranging, we further obtain

$$\frac{1}{A_K} \ln \left( \frac{X^{K+1}(i)}{X^{K+1}(j)} \right) = \frac{1}{A_K} \ln \left( \frac{X^0(i)}{X^0(j)} \right) + (E_i - E_j) \cdot CY^K$$

which implies

$$(E_i - E_j) \cdot CY^K = \frac{1}{A_K} \ln \left( \frac{X^{K+1}(i)}{X^0(i)} \right) - \frac{1}{A_K} \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right)$$

as claimed in the first equation of the lemma. The previous equation further implies

$$(CY^K)_p - E_j \cdot CY^K = \frac{1}{A_K} \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) - \frac{1}{A_K} \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right)$$
which even further implies
\[(C\bar{Y}^K)_p - \bar{Y}^K \cdot CY^K = \frac{1}{A_K} \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) - \frac{1}{A_K} \sum_{j=1}^{n} \bar{Y}^K(j) \ln \left( \frac{X^{K+1}(j)}{X^0(j)} \right)\]
as claimed in the second and third equations of the lemma.

**Lemma 21.** Suppose \(X^0\) is an arbitrary interior strategy. Then, for all \(K \geq 0\), the vectors
\[
\begin{bmatrix}
(C\bar{Y}^K)_1 \\
\vdots \\
(C\bar{Y}^K)_n
\end{bmatrix}
and
\begin{bmatrix}
X^{K+1}(1)/X^0(1) \\
\vdots \\
X^{K+1}(n)/X^0(n)
\end{bmatrix}
\]
have the same ranking in the following sense: For all \(K \geq 0\), if
\[\sigma_K(1), \sigma_K(2), \ldots, \sigma_K(n)\]
is a permutation of the set of pure strategies such that
\[(C\bar{Y}^K)_{\sigma_K(1)} \geq_1 \cdots \geq_{n-1} (C\bar{Y}^K)_{\sigma_K(n)},\]
then
\[
\frac{X^{K+1}(\sigma_K(1))}{X^0(\sigma_K(1))} \geq_1 \cdots \geq_{n-1} \frac{X^{K+1}(\sigma_K(n))}{X^0(\sigma_K(n))}
\]
and, for all \(i = 1, \ldots, n-1\), we have that \(\geq_i\) is an equality if and only if \(\geq_i\) is an equality.

**Proof.** Straightforward implication of Lemma 20.

**Theorem 3.** Let \(C\) be a positive payoff matrix such that
\[
\frac{\max_{ij} C_{ij}}{\min_{ij} C_{ij}} = c
\]
and denote \(X^k \equiv T^k(X^0)\), where \(X^0\) is an interior strategy. Then, if
\[
\liminf_{K \to \infty} \left\{ \ln \left( \max_{p \in C(Y)} \left\{ \frac{X^{k+1}(p)}{X^0(p)} \right\} \right) \right\} > -\infty,
\]
under any sequence of positive learning rates, the weighted empirical average
\[
\bar{Y}^K = \frac{1}{A_K} \sum_{k=0}^{K} \alpha_k Y^k
\]
satisfies, for any \(Y \in \mathbb{X}(C)\), the following inequality
\[
k\text{th} \max_{p \in C(Y)} \{(C\bar{Y}^K)_p\} - \bar{Y}^K \cdot CY^K \leq \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \frac{2c}{A_K}
\]
where \(\max_{p \in C(Y)}\) is the \(k\)th in rank maximum, \(k \in \{2, \ldots, n-1\}\), breaking ties lexicographically.
Proof. Lemma 20 gives that, \( \forall p \in C(Y) \),

\[
(CY^K)_p - \bar{Y}^K \cdot CY^K = \frac{1}{A_K} \sum_{j=1}^{n} \bar{Y}^K(j) \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(j)}{X^{0}(j)} \right).
\]

(15)

Furthermore, from Chebyshev's order inequality and Lemma 21, we have

\[
\sum_{i=1}^{n} (CY^K)_i X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right) \leq \sum_{i=1}^{n} (CY^K)_i X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right).
\]

We may rewrite the previous inequality as the following more concise expression:

\[
\sum_{i=1}^{n} (CY^K)_i X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right) \leq \sum_{i=1}^{n} (CY^K)_i X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right).
\]

(16)

Let us work first with the expression on the left-hand-side of the previous inequality. To that end, we have

\[
\sum_{i=1}^{n} (CY^K)_i X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^K(j) X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \cdot \frac{X^{K+1}(i)}{X^{0}(i)} \right) \right).
\]

37
Assuming \( p \) is any \( k \)th max strategy, \( k \in \{2, \ldots, n - 1\} \), and letting

\[
\ell^* \in \arg \min_{\ell = 1}^{n} \left\{ \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \right\},
\]

and further assuming that \( \sigma_K > 0, \rho_K > 0 \) are such that

\[
\sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \geq 0,
\]

such that

\[
\min_{\ell = 1}^{n} \left\{ \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right\} = 1,
\]

and such that

\[
\sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) = \max_{\ell = 1}^{n} \left\{ \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right\} \leq 2,
\]

and continuing from above, we obtain

\[
\geq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^K(j) X^{K+1}(i) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \left( \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \left( \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right)
\]

which implies

\[
= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^K(j) X^{K+1}(i) \min_{\ell = 1}^{n} \left\{ \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \right\} \left( \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \left( \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right)
\]

which, since

\[
\sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \geq \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right),
\]

38
further implies

$$\geq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \min_{\ell=1}^{n} \left\{ \left( \sigma_{K} + \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \right\} \left( \sigma_{K} - \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies that

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \min_{\ell=1}^{n} \left\{ \left( \sigma_{K}^{2} - \rho_{K}^{2} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \right\} \left( \sigma_{K}^{2} - \rho_{K}^{2} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left\{ \min_{\ell=1}^{n} \left\{ \sigma_{K}^{2} - \rho_{K}^{2} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right\} \left( \min_{\ell=1}^{n} \left\{ \sigma_{K}^{2} - \rho_{K}^{2} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sigma_{K}^{2} - \rho_{K}^{2} \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \sigma_{K}^{2} - \rho_{K}^{2} \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sigma_{K}^{2} - \rho_{K}^{2} \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \sigma_{K}^{2} - \rho_{K}^{2} \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left\{ \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right\} \left( \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left\{ \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right\} \left( \max_{\ell=1}^{\min} \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$

which even further implies

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left\{ \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\} \left( \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right\}$$
which even further implies
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sum_{\ell=1}^{n \min} \left( \sigma_{K} - \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \right) \left( \sigma_{K} + \rho_{K} \sum_{\ell=1}^{n \max} \left( \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \right) = 1,
\]
which, under the previous assumption that
\[
\sum_{\ell=1}^{n \min} \sigma_{K} - \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) = 1,
\]
even further implies that
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sigma_{K} + \rho_{K} \sum_{\ell=1}^{n \max} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) = 1,
\]
which even further implies
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sigma_{K} + \rho_{K} \sum_{\ell=1}^{n \max} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \leq 2 \left( X^{K+1} \cdot C \bar{Y}^{K} \right) \left( \sum_{i=1}^{n} X^{K+1}(i) \right) \left( \sigma_{K} + \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \leq \frac{2}{X^{K+1}(1)} \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(i)}{X^{0}(i)} \right).
\]
Combining the previous inequality with [16], we obtain
\[
\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} \bar{Y}^{K}(j) X^{K+1}(i) \left( \sigma_{K} + \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \leq \frac{2}{X^{K+1}(1)} \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(i)}{X^{0}(i)} \right) \leq 2 \left( X^{K+1} \cdot C \bar{Y}^{K} \right) \left( \sum_{i=1}^{n} X^{K+1}(i) \right) \left( \sigma_{K} + \rho_{K} \ln \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \right) \leq \frac{2}{X^{K+1}(1)} \left( \frac{X^{K+1}(p)}{X^{0}(p)} \right) \left( \frac{X^{K+1}(i)}{X^{0}(i)} \right).
\]
which further implies

\[
(CX^{K+1})_{\text{min}} \sum_{j=1}^{n} \bar{Y}^K(j) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \leq 2 \left( \frac{X^{K+1}}{CX^{K+1}} \right) \sum_{i=1}^{n} X^{K+1} \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right)
\]

which even further implies

\[
\sum_{j=1}^{n} \bar{Y}^K(j) \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) \leq 2 \left( \frac{X^{K+1}}{CX^{K+1}} \right) \sum_{i=1}^{n} X^{K+1} \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right)
\]

which even further implies

\[
\sum_{j=1}^{n} \bar{Y}^K(j) \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \leq 2 \left( \frac{X^{K+1}}{CX^{K+1}} \right) \sum_{i=1}^{n} X^{K+1} \left( \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right) - \sigma_K
\]

which even further implies

\[
\sum_{j=1}^{n} \bar{Y}^K(j) \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \leq 2 \left( \frac{X^{K+1}}{CX^{K+1}} \right) \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right)
\]

Combining the previous inequality with (15) we obtain

\[
(CY^K)_{p} - \bar{Y}^K \cdot CY^K \leq \left( \frac{X^{K+1}}{CX^{K+1}} \right) \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \frac{2}{A_K}
\]
which implies by the assumption $C$ satisfies
\[
\frac{\max_{ij} C_{ij}}{\min_{ij} C_{ij}} = c
\]
that
\[
(C\bar{Y}^K)_p - \bar{Y}^K \cdot C\bar{Y}^K \leq \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \frac{2c}{A_K}
\]
as claimed in the statement of the theorem. Let us now verify that there exist $\sigma_K > 0$ and $\rho_K > 0$ such that
\[
\min_{\ell=1}^n \left\{ \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(\ell)} \right) \right\} \geq 0,
\]
and
\[
\max_{\ell=1}^n \left\{ \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(\ell)} \right) \right\} \leq 2 \iff \min_{\ell=1}^n \left\{ \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(\ell)}{X^0(p)} \right) \right\} = 1,
\]
and
\[
\max_{\ell=1}^n \left\{ \sigma_K - \rho_K \ln \left( \frac{X^{K+1}(p)}{X^0(\ell)} \right) \right\} \leq 2 \iff \max_{\ell=1}^n \left\{ \sigma_K + \rho_K \ln \left( \frac{X^{K+1}(\ell)}{X^0(p)} \right) \right\} \leq 2.
\]
By the assumption
\[
\lim_{K \to \infty} \inf \left\{ \ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right) \right\} > -\infty
\]
the quantity
\[
\ln \left( \frac{X^{K+1}(p)}{X^0(p)} \right)
\]
is bounded away from $-\infty$. Choosing $\rho_K < \gamma \sigma_K$, to satisfy the previous inequalities it suffices to find $\rho_K$, $\sigma_K$, and $\gamma$ such that
\[
1 + \gamma \min_{\ell=1}^n \left\{ \ln \left( \frac{X^{K+1}(p)}{X^0(\ell)} \right) \right\} \geq 0,
\]

\[ \sigma_K + \rho_K \min_{\ell=1}^{n} \left\{ \ln \left( \frac{X^{K+1}(\ell)}{X^0(\ell)} \right) \right\} = 1, \]

and

\[ \sigma_K \left( 1 + \gamma \max_{\ell=1}^{n} \left\{ \ln \left( \frac{X^{K+1}(\ell)}{X^0(\ell)} \right) \right\} \right) \leq 2, \]

Choosing \( \gamma \) arbitrarily close to 0 (but bounded away from 0) satisfies the first inequality, since

\[ \liminf_{K \to \infty} \left\{ \min_{\ell=1}^{n} \left\{ \ln \left( \frac{X^{K+1}(\ell)}{X^0(\ell)} \right) \right\} \right\} > -\infty. \]

Further choosing \( \sigma_K \leq 2 \) also satisfies the third inequality, since

\[ \max_{\ell=1}^{n} \left\{ \ln \left( \frac{X^{K+1}(\ell)}{X^0(\ell)} \right) \right\} \geq 0. \]

Further choosing \( \rho_K \) arbitrarily close to 0 and \( \sigma_K \) arbitrarily close to 1 from above satisfies the second inequality. This completes the proof. \( \square \)