Decompositions and Potentials for Normal Form Games*

William H. Sandholm†

October 13, 2009

Abstract

We introduce a method of decomposing a p-player normal form game into 2^p simultaneously-played component games, each distinguished by the set of "active" players whose choices influence payoffs. We then prove that a normal form game is a potential game if and only if in each of the component games, all active players have identical payoff functions, and that in this case, the sum of these shared payoff functions is the original game's potential function. We conclude by discussing algorithms for deciding whether a given normal form game is a potential game.

1. Introduction

In a normal form potential game (Monderer and Shapley (1996)), all players' incentives are captured by a single scalar-valued function defined on the set of strategy profiles. Potential games therefore possess a number of appealing properties. Most notably, myopic adjustment processes ascend the potential function and converge to Nash equilibria. Potential games appear in a wide range of applications, ranging from genetic competition (Fisher (1930); Hofbauer and Sigmund (1988)) to network congestion (Beckmann et al. (1956); Rosenthal (1973); Sandholm (2001); Roughgarden and Tardos (2002)) to Cournot competition (Slade (1994); Monderer and Shapley (1996)) to externality pricing and evolutionary implementation (Sandholm (2005, 2007)).

In this paper, we introduce a method of decomposing p-player normal form games into 2^p simultaneously-played component games. We then use this decomposition to provide

*I thank Satoru Takahashi, Takashi Ui, two anonymous referees, and an anonymous Advisory Editor for helpful comments, and Katsuhiko Aiba for able research assistance. I also gratefully acknowledge financial support from NSF Grants SES-0092145, SES-0617753, and SES-0851580.

†Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA. e-mail: whs@ssc.wisc.edu; website: http://www.ssc.wisc.edu/~whs.

new characterizations of normal form potential games. Unlike the characterizations of Monderer and Shapley (1996), which consider changes in payoffs along paths through the set of pure strategy profiles, our characterizations are "global" in nature, relying on multilinear transformations of players' entire payoff functions. In addition to providing novel insights into the structure of potential games, one of our characterizations leads to an efficient algorithm for deciding whether a given normal form game is a potential game.

Monderer and Shapley (1996) call a normal form game a potential game if admits a potential function: a scalar-valued function defined on the set of strategy profiles with the property that any unilateral deviation has the same effect on the deviator's payoff as it does on potential. Monderer and Shapley (1996) prove that a normal form game admits a potential function if and only if over any closed path of strategy profiles generated by four unilateral deviations, the sum of the changes in the deviators' payoffs equals zero. When a potential function exists, it can be constructed by traversing a path whose steps are unilateral devisions, adjusting the value of potential by the change in the deviator's payoffs as each step is taken.

In contrast to this "pathwise" approach, our analysis of normal form games views each player's entire payoff function as a discrete algebraic object that can be manipulated as a unit. Using this approach, we introduce a decomposition of a p-player normal form game into 2^p simultaneously-played component games, where each component is distinguished by its set Q of *active players* and the complementary set Q^c of *passive players*.

This decomposition makes use of two orthogonal projections: Φ, which when applied to a vector subtracts off the the vector's mean from each component, and $\Xi = I - \Phi$, which returns the constant vector whose common component is this mean. Thus, for any mixed strategy x^p of player p, the vector Ξx^p is player p's uniform mixed strategy, while $\Phi x^p = x^p - \Xi x^p$ is the vector difference between x^p and the uniform mixed strategy. By applying these transformations directly to player p's payoff function U^p, we create a new payoff function U^p_{Q,Q^c} defined by this property: when mixed strategy profile $(x^1, \ldots, x^p)$ is played, player p's expected payoff under U^p_{Q,Q^c} is the expected payoff he would have obtained under U^p if each active player $q \in Q$ had chosen "mixed strategy" Φx^q and each passive player $r \in Q^c$ had chosen her uniform mixed strategy.[1]

Our characterization theorems for normal form potential games build on this decomposition. In Theorem 3.2, we show that a normal form game is a potential game if and only if in each of its component games, all active players have identical payoff functions; in this case, the sum of these shared payoff functions is the game's potential function.

[1]Of course, the vector Φx^q is not a true mixed strategy, since the sum of its components is zero, but this fact does not prevent us from formally computing payoffs when this "mixed strategy" is employed.

This expression for the potential function provides a link with the work of Ui (2000), who characterizes normal form potential games by means of collections of functions called *interaction potentials*: in Ui's (2000) characterization, the sum of the functions in the interaction potential is a potential function for the game at issue. While our characterization of potential games is thus related to Ui's (2000), it offers some distinct advantages: it provides a systematic method of determining whether or not a given normal form game is a potential game, as well as a mechanical procedure for constructing a game's potential function from its payoff functions. We provide a detailed discussion of the links between Ui's (2000) analysis and ours in Section 3.2.

It is a small step from Theorem 3.2 to a second characterization of potential games, Theorem 3.5, that only requires comparisons of the payoffs of two players at a time. While this approach does not provide a formula for a game's potential function, the characterization it provides is quite simple, in that allows one to determine whether a game is a potential game simply by verifying a collection of matrix equalities. In Section 4, we show how this characterization can be used to define an algorithm for deciding whether a given normal form game is a potential game, and compare the performance of this algorithm to those of others from the literature.

Section 2 begins the analysis by introducing our decomposition of normal form games. Section 3 offers Monderer and Shapley's (1996) definition and characterizations of potential games, and then uses the decomposition to establish our two new characterizations. Section 4 discusses algorithms for deciding whether a given game is a potential game. One proof omitted from the text is provided in the Appendix.

2. Decomposition of Normal Form Games

This section presents our method of representing a p-player normal form game as the sum of 2^p component games. To make the developments to come easier to follow, we first explain how our decomposition works in the simple setting of two-player games.

Consider a two-player game in which players 1 and 2 have n^1 and n^2 strategies, respectively. A player's payoffs in this game can be represented by a matrix $M \in \mathbf{R}^{n^1 \times n^2}$, or by a bilinear function that assigns pairs of vectors $(x^1, x^2) \in \mathbf{R}^{n^1} \times \mathbf{R}^{n^2}$ to real numbers: namely, the function $\mathcal{M}(x^1, x^2) = (x^1)' M x^2$. When (x^1, x^2) is a mixed strategy profile, the function $\mathcal{M}$ is simply the mixed extension of player p's pure-strategy payoff function. Taking the domain of $\mathcal{M}$ to be all of $\mathbf{R}^{n^1} \times \mathbf{R}^{n^2}$ ensures that the map from payoff matrices M to bilinear functions $\mathcal{M}$ is an isomorphism.

Any pair of matrices $T^1 \in \mathbf{R}^{n^1 \times n^1}$ and $T^2 \in \mathbf{R}^{n^2 \times n^2}$ can be used to define a decomposition of the payoff matrix M into the sum of four component matrices. To see how, write $T^p_c = I - T^p$, and use the identity $x^p = T^p x^p + T^p_c x^p$ to express $\mathcal{M}(x^1, x^2)$ as

$$\begin{aligned}\mathcal{M}(x^1, x^2) &= (x^1)' M x^2 \\ &= (T^1 x^1 + T^1_c x^1)' M (T^2 x^2 + T^2_c x^2) \\ &= (T^1 x^1)' M (T^2 x^2) + (T^1 x^1)' M (T^2_c x^2) + (T^1_c x^1)' M (T^2 x^2) + (T^1_c x^1)' M (T^2_c x^2) \\ &= (x^1)'[(T^1)' M T^2] x^2 + (x^1)'[(T^1)' M T^2_c] x^2 + (x^1)'[(T^1_c)' M T^2] x^2 + (x^1)'[(T^1_c)' M T^2_c] x^2 \\ &\equiv (x^1)' M_{\{1,2\},\varnothing}\, x^2 + (x^1)' M_{\{1\},\{2\}}\, x^2 + (x^1)' M_{\{2\},\{1\}}\, x^2 + (x^1)' M_{\varnothing,\{1,2\}}\, x^2.\end{aligned}$$

Then by virtue of the isomorphism, we have the following decomposition of the payoff matrix M:

$$M = M_{\{1,2\},\varnothing} + M_{\{1\},\{2\}} + M_{\{2\},\{1\}} + M_{\varnothing,\{1,2\}}. \tag{1}$$

Now, suppose we choose $T^p_c = I - T^p$ to be the matrix whose entries are all $\frac{1}{n^p}$. Then T^p_c maps each mixed strategy x^p to the uniform mixed strategy, while $T^p x^p = (I - T^p_c) x^p$ maps x^p to its displacement from the uniform mixed strategy. Then as we discuss in Section 2.5, we can interpret each component M_{Q,Q^c} of the decomposition (1) as the payoff matrix obtained when the players in $Q \subseteq \{1, 2\}$ are assigned to "active" roles in the game, while the players in $Q^c = \{1, 2\} - Q$ are assigned to "passive" roles in the game.

In the remainder of this section, we generalize this construction to games with arbitrary numbers of players. Doing so requires us to replace the payoff matrices and bilinear functions from the argument above with payoff arrays and multilinear functions. The last result in this section, Proposition 2.7, presents the p-player version of the decomposition described informally in the previous paragraph. In Section 3, we use this decomposition as the basis for our new characterization of potential games: Theorem 3.2 shows that a normal form game is a potential game if and only if in each of its component games, all active players have identical payoff functions, and that in this case, the sum of these shared payoff functions is a potential function for the game. After presenting this and related results, we revisit the case of two-player games in Example 3.4.

2.1 Normal Form Games

A p-player normal form game is defined by a strategy set $S^p = \{1, \ldots, n^p\}$ and a utility function $U^p : S \to \mathbf{R}$ for each player $p \in \mathcal{P} = \{1, ..., p\}$. The domain of U^p is the set of pure strategy profiles $S = \prod_{q \in \mathcal{P}} S^q$. In what follows, we often use the notation $U^p \in \mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$ to refer to the space of payoff functions.

A mixed strategy for player $p \in \mathcal{P}$ is an element of the simplex $X^p = \{x^p \in \mathbf{R}^{n^p}_+ : \sum_{i \in S^p} x^p_i = m^p\}$. Mixed strategy profiles are elements of the product set $X = \prod_{r \in \mathcal{P}} X^r = \{x = (x^1, \ldots, x^p) \in \mathbf{R}^n_+ : x^r \in X^r\}$. Expected payoffs to mixed strategy profiles are described by the multilinear function $\mathcal{U}^p : \prod_{r \in \mathcal{P}} X^r \to \mathbf{R}$, where

$$\mathcal{U}^p(x) = \sum_{s \in S} U^p(s) \prod_{r \in \mathcal{P}} x^r_{s^r} \text{ for all } x \in X = \textstyle\prod_{r \in \mathcal{P}} X^r. \tag{2}$$

2.2 Arrays and Multilinear Functions

With this motivation, let us call any $M \in \mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$ an *array* of *rank* p.[2] Each such array defines a multilinear function $\mathcal{M} : \prod_{r \in \mathcal{P}} \mathbf{R}^{n^r} \to \mathbf{R}$ via

$$\mathcal{M}(x) = \sum_{s \in S} M(s) \prod_{r \in \mathcal{P}} x^r_{s^r} \text{ for all } x \in \mathbf{R}^n = \textstyle\prod_{r \in \mathcal{P}} \mathbf{R}^{n^r}. \tag{3}$$

The choice of domain $\mathbf{R}^n = \prod_{r \in \mathcal{P}} \mathbf{R}^{n^r}$ rather than X in equation (3) ensures that arrays and multilinear functions are isomorphic. If we let $\iota^p_{s^p}$ denote a standard basis vector in $\mathbf{R}^{n^p}$, then we can write down the map from multilinear functions to arrays that inverts (3):

$$M(s^1, \ldots, s^p) = \mathcal{M}(\iota^1_{s^1}, \ldots, \iota^p_{s^p}). \tag{4}$$

We summarize these points in the following lemma.

Lemma 2.1. *The map* (3) *and its inverse* (4) *define a linear isomorphism between the vector space* $\mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$ *of arrays and the vector space* $\{\mathcal{M} \mid \mathcal{M} : \prod_{r \in \mathcal{P}} \mathbf{R}^{n^r} \to \mathbf{R}\}$ *of multilinear functions.*

We now note a useful consequence of this isomorphism. Let $\mathcal{M} : \prod_{r \in \mathcal{P}} \mathbf{R}^{n^r} \to \mathbf{R}$ be the multilinear function corresponding to array $M \in \mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$, and for each $p \in \mathcal{P}$, let T^p be a matrix in $\mathbf{R}^{n^p \times n^p}$. The function $\hat{\mathcal{M}}$ defined by

$$\hat{\mathcal{M}}(x^1, \ldots, x^p) = \mathcal{M}(T^1 x^1, \ldots, T^p x^p) \tag{5}$$

is clearly multilinear, and so must correspond under the isomorphism to some array $\hat{M}$. Using equations (4), (5), and (3), we can express $\hat{M}$ directly in terms of the original array M:

$$\begin{aligned} \hat{M}(\hat{s}^1, \ldots, \hat{s}^p) &= \hat{\mathcal{M}}(\iota^1_{\hat{s}^1}, \ldots, \iota^p_{\hat{s}^p}) \\ &= \mathcal{M}(T^1 \iota^1_{\hat{s}^1}, \ldots, T^p \iota^p_{\hat{s}^p}) \end{aligned} \tag{6}$$

[2]Thus, an array of rank 1 is a vector, while an array of rank 2 is a matrix.

$$= \sum_{s \in S} M(s) \prod_{r \in \mathcal{P}} (T^r \iota^r_{\hat{s}^r})_{s^r}$$
$$= \sum_{s \in S} M(s) \prod_{r \in \mathcal{P}} T^r_{s^r \hat{s}^r}.$$

Lemma 2.2 summarizes this argument.

Lemma 2.2. *For a given array M and matrices $T^1, \ldots, T^p$, equation* (6) *defines the unique array $\hat{M}$ that corresponds under the isomorphism to the composition $\hat{\mathcal{M}}$ from equation* (5).

2.3 Projections and Decompositions of Arrays

For each $p \in \mathcal{P}$, suppose that $T^p \in \mathbf{R}^{n^p \times n^p}$ is a projection (i.e., that $(T^p)^2 = T^p$). For any subsets $Q, R \subseteq \mathcal{P}$, define the matrix $T^p_{Q,R} \in \mathbf{R}^{n^p \times n^p}$ as follows:

$$T^p_{Q,R} = \begin{cases} I & \text{if } p \in Q^c \cap R^c, \\ T^p & \text{if } p \in Q \cap R^c, \\ I - T^p & \text{if } p \in Q^c \cap R, \\ \mathbf{0} & \text{if } p \in Q \cap R. \end{cases} \tag{7}$$

$T^p_{Q,R}$ can be viewed as a composition of at most two projections. The composition includes the projection T^p whenever $p \in Q$, and it includes the complementary projection $I - T^p$ whenever $p \in R$. If p is in both Q and R, both projections are applied, yielding $T^p_{Q,R} = T^p(I - T^p) = T^p - (T^p)^2 = \mathbf{0}$.

Given an array $M \in \mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$, projections $T^1, \ldots, T^p$, and sets $Q, R \subseteq \mathcal{P}$, let the array $M_{Q,R} \in \mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$ be generated by M and the matrices $T^1_{Q,R}, \ldots, T^p_{Q,R}$ as described in Lemma 2.2. Evidently, the corresponding multilinear functions $\mathcal{M}$ and $\mathcal{M}_{Q,R}$ are related by

$$\mathcal{M}_{Q,R}(x^1, \ldots, x^p) = \mathcal{M}(T^1_{Q,R} x^1, \ldots, T^p_{Q,R} x^p)$$

The next three lemmas note some properties of the map $M \mapsto M_{Q,R}$.

Lemma 2.3 (Distributive property)**.** $(M + N)_{Q,R} = M_{Q,R} + N_{Q,R}$.

Lemma 2.4 (Null property)**.** *If $Q \cap R \neq \emptyset$, then $M_{Q,R} = \mathbf{0}$.*

Lemma 2.5 (Composition property)**.** $(M_{Q,R})_{\hat{Q},\hat{R}} = M_{Q \cup \hat{Q}, R \cup \hat{R}}$.

If $p \in Q \cap R$, then the matrix $T^p_{Q,R}$ equals $\mathbf{0}$; Lemma 2.4 records that in this case, the array $M_{Q,R}$ equals $\mathbf{0}$ as well. Lemma 2.5 observes that the composition $M \mapsto M_{Q,R} \mapsto (M_{Q,R})_{\hat{Q},\hat{R}}$, which employs two rounds of projections, can be expressed as a single round of projections via $M \mapsto M_{Q\cup\hat{Q},R\cup\hat{R}}$. Lemmas 2.3 and 2.4 both follow easily from equation (6). The intuition behind Lemma 2.5 is straightforward, but the proof requires some bookkeeping. We present the proof in the Appendix.

Our last preliminary result, Lemma 2.6, shows how projections can be used to obtain additive decompositions of arrays. This decomposition (with $\hat{\mathcal{P}} = \mathcal{P}$) is a basic ingredient of our characterization of potential games.

Lemma 2.6 (Decomposition of arrays)**.** *Let* $\hat{\mathcal{P}} \subseteq \mathcal{P}$. *Then*

(*i*) $M = \sum_{Q\subseteq\hat{\mathcal{P}}} M_{Q,\hat{\mathcal{P}}-Q}$.

(*ii*) $M = N$ *if and only if* $M_{Q,\hat{\mathcal{P}}-Q} = N_{Q,\hat{\mathcal{P}}-Q}$ *for all* $Q \subseteq \hat{\mathcal{P}}$.

Proof. Suppose without loss of generality that $\hat{\mathcal{P}} = \{1,\ldots,\hat{p}\}$, and let $\mathcal{M}$ be the multilinear function corresponding to M. Then writing T^p_c for $I - T^p$, we obtain

$$\begin{aligned}
\mathcal{M}(x^1,\ldots,x^p) &= \mathcal{M}(T^1x^1,x^2,\ldots,x^p) + \mathcal{M}(T^1_cx^1,x^2,\ldots,x^p)\\
&= \mathcal{M}(T^1x^1,T^2x^2,x^3,\ldots,x^p) + \mathcal{M}(T^1x^1,T^2_cx^2,x^3,\ldots,x^p)\\
&\qquad + \mathcal{M}(T^1_cx^1,T^2x^2,x^3,\ldots,x^p) + \mathcal{M}(T^1_cx^1,T^2_cx^2,x^3,\ldots,x^p)\\
&= \mathcal{M}(T^1x^1,\ldots,T^{\hat{p}-1}x^{\hat{p}-1},T^{\hat{p}}x^{\hat{p}},x^{\hat{p}+1},\ldots,x^p)\\
&\qquad + \mathcal{M}(T^1x^1,\ldots,T^{\hat{p}-1}x^{\hat{p}-1},T^{\hat{p}}_cx^{\hat{p}},x^{\hat{p}+1},\ldots,x^p)\\
&\qquad + \ldots + \mathcal{M}(T^1_cx^1,\ldots,T^{\hat{p}-1}_cx^{\hat{p}-1},T^{\hat{p}}x^{\hat{p}},x^{\hat{p}+1},\ldots,x^p)\\
&\qquad + \mathcal{M}(T^1_cx^1,\ldots,T^{\hat{p}-1}_cx^{\hat{p}-1},T^{\hat{p}}_cx^{\hat{p}},x^{\hat{p}+1},\ldots,x^p)\\
&= \sum_{Q\subseteq\hat{\mathcal{P}}} \mathcal{M}_{Q,\hat{\mathcal{P}}-Q}(x^1,\ldots,x^p).
\end{aligned}$$

Both parts of the lemma follow easily from this equality and the isomorphism. ■

2.4 The Orthogonal Projection Φ

Define the symmetric matrix $\Phi \in \mathbf{R}^{n^p\times n^p}$ by

$$\Phi = I - \tfrac{1}{n^p}\mathbf{1}\mathbf{1}',$$

where I is the identity matrix and $\mathbf{1} \in \mathbf{R}^{n^p}$ is the vector of ones. Φ is the orthogonal projection of $\mathbf{R}^{n^p}$ onto the set $TX^p = \{z^p \in \mathbf{R}^{n^p} : \sum_{i\in S^p} z^p_i = 0\}$. TX^p is the *tangent space* of the simplex X^p; in other words, it contains all directions of motion through X^p. The complement of projection Φ is $\Xi = I - \Phi = \frac{1}{n^p}\mathbf{1}\mathbf{1}'$, the orthogonal projection onto span($\{\mathbf{1}\}$).

Using these projections, we can decompose any mixed strategy $x^p \in X^p$ as

$$\begin{aligned} x^p &= \Phi x^p + \Xi x^p \\ &= (I - \tfrac{1}{n^p}\mathbf{1}\mathbf{1}')x^p + \tfrac{1}{n^p}\mathbf{1}\mathbf{1}'x^p \\ &= (x^p - \tfrac{1}{n^p}\mathbf{1}) + \tfrac{1}{n^p}\mathbf{1}. \end{aligned}$$

Here $\Xi x^p \equiv \frac{1}{n^p}\mathbf{1}$ is the uniform randomization over the strategies in $S^p = \{1, \ldots, n^p\}$, while Φx^p represents the displacement of mixed strategy x^p from this uniform randomization.

It is worth noting that the projection Φ is also useful in studying potential games played by continuous populations of pure strategists; see Sandholm (2009).

2.5 Components and Decompositions of Normal Form Games

We now combine the foregoing ideas to obtain our decomposition of normal form games. Let $U = (U^1, \ldots, U^p)$ be a p-player normal form game, where each payoff function U^p is an array in $\mathbf{R}^{\prod_{r\in\mathcal{P}} n^r}$, and set $T^p = \Phi$, so that $I - T^p = \Xi$. Then for any sets $Q, R \subseteq \mathcal{P}$, each $U^p_{Q,R} \in \mathbf{R}^{\prod_{r\in\mathcal{P}} n^r}$ is a payoff function. We can therefore view $U_{Q,R} = (U^1_{Q,R}, \ldots, U^p_{Q,R})$ as a *derived game* obtained from the original normal form game U. When each player r chooses a mixed strategy $x^r \in X^r$, expected payoffs in $U_{Q,R}$ can be expressed as

$$\text{(8)} \qquad \mathcal{U}^p_{Q,R}(x^1, \ldots, x^p) = \mathcal{U}^p(T^1_{Q,R}x^1, \ldots, T^p_{Q,R}x^p),$$

where $\mathcal{U}^p : \prod_{r\in\mathcal{P}} \mathbf{R}^{n^r} \to \mathbf{R}$ and $\mathcal{U}^p_{Q,R} : \prod_{r\in\mathcal{P}} \mathbf{R}^{n^r} \to \mathbf{R}$ are the multilinear functions corresponding to arrays U^p and $U^p_{Q,R}$, respectively.

Equations (7) and (8) enable us to interpret the derived game $U_{Q,R}$. Suppose that Q and R are disjoint, so that $U_{Q,R}$ is not automatically the null game (cf Lemma 2.4). If $r \in R$, then $T^r_{Q,R} = \Xi$, and so $T^r_{Q,R}x^r = \Xi x^r = \frac{1}{n^r}\mathbf{1}$. Thus, every choice of mixed strategy by player r in the derived game $U_{Q,R}$ corresponds to the choice of the uniform mixed strategy in the original game U. We therefore call each player $r \in R$ a *passive player* in game $U_{Q,R}$.

If instead $q \in Q$, then $T^q_{Q,R} = \Phi$. Thus, the choice of mixed strategy $x^q \in X^q$ in the derived game $U_{Q,R}$ corresponds to the choice of "mixed strategy" $\Phi x^q \in TX^q$ in the original game U. Φx^q represents the displacement of player q's actual mixed strategy from the uniform mixed strategy $\frac{1}{n^q}\mathbf{1}$. Of course, Φx^q is not a true mixed strategy, since its components sum to 0, but because we take the domain of $\mathcal{U}^p$ to be $\prod_{r\in\mathcal{P}} \mathbf{R}^{n^r}$, we can allow "mixed strategies" of this sort. We call each player $q \in Q$ an *active player* in game $U_{Q,R}$.

Finally, if $\hat{p} \in Q^c \cap R^c$, then $T^{\hat{p}}_{Q,R} = I$. By equation (8), the choice of mixed strategy $x^{\hat{p}}$ in game $U_{Q,R}$ corresponds to the choice of the same mixed strategy $x^{\hat{p}}$ in game U. We thus call each player $\hat{p} \in Q^c \cap R^c$ a *full player* in game $U_{Q,R}$.

While it is useful to speak of mixed strategies in interpreting the derived game $U_{Q,R}$, we do not need to do so to compute this game's payoff functions. Assuming still that Q and R are disjoint, we can substitute definition (7) into equation (6) to express $U^p_{Q,R}$ in terms of the original payoff function U^p:

$$(9) \qquad U^p_{Q,R}(\hat{s}) = \sum_{s^Q \in S^Q} \sum_{s^R \in S^R} U^p(s^Q, s^R, \hat{s}^{-(Q \cup R)}) \prod_{q \in Q} \left(1_{\{s^q = \hat{s}^q\}} - \tfrac{1}{n^q}\right) \prod_{r \in R} \tfrac{1}{n^r}.$$

The derived games $U_{Q,R}$ appearing in our first characterization of potential games, Theorem 3.2, all have $R = Q^c$: each player is either active or passive. We call the derived games of this form the *components* of U. Proposition 2.7, a direct consequence of Lemma 2.6, shows that any normal form game can be expressed as the sum of its components.

Proposition 2.7 (Decomposition of normal form games)**.**

Every normal form game U can be expressed as the sum of its components U_{Q,Q^c}. In particular,

$$(10) \qquad U^p = \sum_{Q \subseteq \mathcal{P}} U^p_{Q,Q^c} \;\; \textit{for all } p \in \mathcal{P}.$$

Equation (10) also admits a simple game-theoretic interpretation: it tells us that any play of the normal form game U can be interpreted as the simultaneous play of its component games U_{Q,Q^c}. Simultaneity is important here: for the interpretation to be valid, each player's choice of strategy cannot vary across component games.

3. Characterizations of Potential Games

3.1 Potential Games

The normal form game U is a *potential game* (Monderer and Shapley (1996)) if there is a *potential function* $V : S \to \mathbf{R}$ such that

$$(11) \quad U^p(\hat{s}^p, s^{-p}) - U^p(s^p, s^{-p}) = V(\hat{s}^p, s^{-p}) - V(s^p, s^{-p}) \;\text{ for all } \hat{s}^p, s^p \in S^p, s^{-p} \in S^{-p}, \text{ and } p \in \mathcal{P}.$$

In words, U is a potential game if any unilateral deviation has the same effect on the deviator's payoffs as it has on potential. Equivalently, U is a potential game if there is a potential function V and auxiliary functions $W^p : S^{-p} \to \mathbf{R}$ such that

$$(12) \qquad U^p(s) = V(s) + W^p(s^{-p}) \;\text{ for all } s \in S \text{ and } p \in \mathcal{P},$$

so that each player's payoff is the sum of a common payoff term and a term that only depends on opponents' behavior. It is easy to see that pure strategy profile $s \in S$ is a Nash equilibrium of U if and only if it is a local maximizer of the potential function V, and that evolutionary processes based on payoff-improving changes in strategy lead to Nash equilibria.

In addition to the simple characterization (12), Monderer and Shapley (1996) offer "pathwise" characterizations of potential games and their potential functions. They prove that U is a potential game if and only if over every closed path of strategy profiles generated by unilateral deviations, the sum of the changes in the deviators' payoffs equals zero. To obtain a condition that can be checked more quickly, they show that it is enough to verify the equality on closed paths constructed from exactly four unilateral deviations:

Theorem 3.1 (Monderer and Shapley (1996)). *U is a potential game if and only if for every $s^p, t^p \in S^p$, $s^q, t^q \in S^q$, $s^{-\{p,q\}} \in S^{-\{p,q\}}$, and $p, q \in \mathcal{P}$, we have that*

$$(U^p(\hat{s}) - U^p(s)) + (U^q(t) - U^q(\hat{s})) + (U^p(\hat{t}) - U^p(t)) + (U^q(s) - U^q(\hat{t})) = 0, \tag{13}$$

where $s = (s^p, s^q, s^{-\{p,q\}})$, $\hat{s} = (t^p, s^q, s^{-\{p,q\}})$, $t = (t^p, t^q, s^{-\{p,q\}})$, and $\hat{t} = (s^p, t^q, s^{-\{p,q\}})$.

When a game admits a potential function, this function is unique up to an additive constant. To construct a game's potential function, one can fix the function's value arbitrarily at some strategy profile s_0, and then compute its value at each other profile s_k by adjusting the value of V incrementally along a unilateral deviation path to s_k.[3]

3.2 Characterization of Potential Games and Potential Functions

Theorem 3.2 provides a new characterization of potential games and their potential functions. It is based on the decomposition of normal form games introduced in Section 2. In contrast to Monderer and Shapley's (1996) "pathwise" characterization of potential games, the one presented below is "global", as it relies on transformations of the players' entire payoff arrays.

In array notation, the normal form game $U = (U^1, \dots, U^p)$ is a potential game if there exist arrays V and $W^1, \dots, W^p$ in $\mathbf{R}^{\prod_{r \in \mathcal{P}} n^r}$ such that

$$W^p \text{ is independent of } s^p \text{ for all } p \in \mathcal{P}, \text{ and} \tag{14}$$

$$U^p = V + W^p \text{ for all } p \in \mathcal{P}. \tag{15}$$

[3] Fix $V(s_0)$, and let $\{s_0, s_1, \dots, s_k\}$ be a path through S, where only player $\pi(i)$ chooses different strategies at s_i and s_{i-1}; then the value of potential at profile s_k is $V(s_k) = V(s_0) + \sum_{i=1}^{k} (U^{\pi(i)}(s_i) - U^{\pi(i)}(s_{i-1}))$.

With these conditions in hand, we can state our characterization theorem.

Theorem 3.2. *$U = (U^1, \ldots, U^p)$ is a potential game if and only if*

$$\text{for all } \varnothing \neq Q \subseteq \mathcal{P} \text{ and all } p, q \in Q, \text{ we have that } U^p_{Q,Q^c} = U^q_{Q,Q^c} \equiv Y^Q. \tag{16}$$

If these statements are true, then conditions (14) *and* (15) *are satisfied by*

$$W^p = U^p_{\varnothing,\{p\}} \quad \text{and} \tag{17}$$

$$V = \sum_{\varnothing \neq Q \subseteq \mathcal{P}} Y^Q, \tag{18}$$

and the entries of this potential function V sum to 0.

Proof. To begin, we note that the measurability condition on W^p from condition (14) can be expressed in terms of derivative games: it is easy to verify that

$$W^p \text{ is independent of } s^p \;\Leftrightarrow\; W^p_{\varnothing,\{p\}} = W^p. \tag{19}$$

Now suppose that U is a potential game with potential function V, fix a nonempty set $Q \subseteq \mathcal{P}$, and suppose that $q \in Q$. Then

$$\begin{aligned}
U^q_{Q,Q^c} &= V_{Q,Q^c} + W^q_{Q,Q^c} \\
&= V_{Q,Q^c} + \left(W^q_{\varnothing,\{q\}}\right)_{Q,Q^c} \\
&= V_{Q,Q^c} + W^q_{Q,\{q\}\cup Q^c} \\
&= V_{Q,Q^c},
\end{aligned}$$

where the first equality follows from equation (15) and Lemma 2.3, the second from equation (19), the third from Lemma 2.5, and the fourth from Lemma 2.4. Condition (16) (with $Y^Q = V_{Q,Q^c}$) immediately follows.

To establish the other direction of the equivalence and the characterization of the potential function, suppose that condition (16) holds, and define V as in equation (18). For each nonempty $Q \subseteq \mathcal{P}$, let $\pi(Q)$ be an arbitrary element of Q. Then for each nonempty $P \subseteq \mathcal{P}$, we can compute as follows:

$$\begin{aligned}
V_{P,P^c} &= \sum_{\varnothing \neq Q \subseteq \mathcal{P}} (Y^Q)_{P,P^c} \\
&= \sum_{\varnothing \neq Q \subseteq \mathcal{P}} \left(U^{\pi(Q)}_{Q,Q^c}\right)_{P,P^c}
\end{aligned} \tag{20}$$

$$
\begin{aligned}
&= \sum_{\varnothing \neq Q \subseteq \mathcal{P}} U^{\pi(Q)}_{P \cup Q, P^c \cup Q^c} \\
&= U^{\pi(P)}_{P,P^c}. \\
&= Y^P.
\end{aligned}
$$

Here the third equality follows from Lemma 2.5, while the fourth equality follows from Lemma 2.4 and from the fact that sets $P \cup Q$ and $P^c \cup Q^c$ are disjoint if and only if $P = Q$.

Next, fix $p \in \mathcal{P}$. Applying Lemma 2.6 (twice) and then Lemma 2.5 (setting $P = Q \cup \{p\}$), we can write

$$
\begin{aligned}
(21) \qquad U^p &= U^p_{\{p\},\varnothing} + U^p_{\varnothing,\{p\}} \\
&= \sum_{Q \subseteq \mathcal{P} - \{p\}} \left(U^p_{\{p\},\varnothing}\right)_{Q,(\mathcal{P}-\{p\})-Q} + U^p_{\varnothing,\{p\}} \\
&= \sum_{\{p\} \subseteq P \subseteq \mathcal{P}} U^p_{P,P^c} + U^p_{\varnothing,\{p\}}.
\end{aligned}
$$

Deriving the analogous expression for V, subtracting, and substituting in equation (20), we obtain

$$
\begin{aligned}
(22) \qquad U^p - V &= \left(\sum_{\{p\} \subseteq P \subseteq \mathcal{P}} U^p_{P,P^c} + U^p_{\varnothing,\{p\}} \right) - \left(\sum_{\{p\} \subseteq P \subseteq \mathcal{P}} V_{P,P^c} + V_{\varnothing,\{p\}} \right) \\
&= U^p_{\varnothing,\{p\}} - V_{\varnothing,\{p\}}.
\end{aligned}
$$

If we set $W^p = U^p_{\varnothing,\{p\}} - V_{\varnothing,\{p\}}$, then equation (22) becomes condition (15), and Lemmas 2.3 and 2.5 and equation (19) imply that measurability condition (14) holds. Therefore, U is a potential game with the potential function V defined in (18).

Finally, applying equation (9) to $V_{\varnothing,\mathcal{P}}$ shows that the latter is a constant array whose entries all equal the average of the entries of V. At the same time, repeating calculation (20) with $P = \varnothing$ shows that $V_{\varnothing,\mathcal{P}} = \mathbf{0}$. Together, these statements imply that the entries of V sum to 0, and that $W^p = U^p_{\varnothing,\{p\}}$ as required by condition (17). This completes the proof of the theorem. ■

There are interesting connections between Theorem 3.2 and a characterization of potential games due to Ui (2000).

Theorem 3.3 (Ui (2000))**.** *The normal form game $U = (U^1, \dots, U^p)$ is a potential game if and only if there is a collection of arrays $\{Y^Q\}_{\emptyset \neq Q \subseteq \mathcal{P}}$, called an* interaction potential*, such that*

$$\text{(23)} \qquad \textit{each } Y^Q \textit{ is measurable with respect to } S^Q = \prod_{q \in Q} S^q, \textit{ and}$$

$$\text{(24)} \qquad U^p = \sum_{\{p\} \subseteq Q \subseteq \mathcal{P}} Y^Q.$$

In this event, the potential function for U is the sum of the arrays Y^Q, as in equation (18).

Now consider a new game $\hat{U}$ of the form

$$\text{(25)} \qquad \hat{U}^p = \sum_{\{p\} \subseteq Q \subseteq \mathcal{P}} Y^Q + W^p, \text{ where } W^p \text{ is measurable with respect to } S^{-p}.$$

Evidently, $\hat{U}$ is itself a potential game, again with potential function (18). Combining this observation with equations (19) and (21) yields an alternate proof that condition (16) from Theorem 3.2 is sufficient for a normal form game to be a potential game.

Ui (2000) uses his characterization theorem to establish surprising links among potential functions for normal form games, the potential function on the set of transferable utility cooperative games introduced by Hart and Mas-Colell (1989), and the Shapley value. To compare Ui's (2000) characterization to our Theorem 3.2, note that the arrays Y^Q from Ui's (2000) theorem are not unique, and are taken as exogenous: they are not determined as functions of the payoff arrays U^p. For this reason, Ui's (2000) theorem does not provide a systematic method of determining whether a given normal form game is a potential game, nor does it provide a mechanical procedure for constructing a game's potential function from its payoff functions.

We conclude this section with an application of Theorem 3.2.

Example 3.4. Two-player potential games. Applying our characterization theorem to two-player games is particularly simple, in part because there are only four subsets of $\mathcal{P} = \{1, 2\}$, and in part because arrays of rank $p = 2$ are matrices. In this setting, multilinear transformations of arrays become matrix products: if the rank 2 array $\hat{M} \in \mathbf{R}^{n^1 \times n^2}$ is generated from the rank 2 array $M \in \mathbf{R}^{n^1 \times n^2}$ and matrices $T^1 \in \mathbf{R}^{n^1 \times n^1}$ and $T^2 \in \mathbf{R}^{n^2 \times n^2}$ as in Lemma 2.2, then $\hat{M}$ can be expressed in matrix notation as $\hat{M} = (T^1)' M\, T^2$.

Let $U = (U^1, U^2)$ be a normal form game. Since the only subset of $\mathcal{P} = \{1, 2\}$ with more than one element is $\mathcal{P}$ itself, the necessary and sufficient condition for U to be a potential game is $U^1_{\mathcal{P},\emptyset} = U^2_{\mathcal{P},\emptyset}$: in the component of U in which both players are active, both must have identical payoffs. In matrix notation, the condition becomes

$$\Phi U^1 \Phi = \Phi U^2 \Phi. \tag{26}$$

When this condition holds, the potential function of U whose components sum to zero is easily obtained from equation (18):

$$\begin{aligned} V &= \sum_{\emptyset \neq Q \subseteq \mathcal{P}} Y^Q \\ &= Y^{\mathcal{P}} + Y^{\{1\}} + Y^{\{2\}} \\ &= \Phi U^1 \Phi + \Phi U^1 \Xi + \Xi U^2 \Phi. \end{aligned}$$

It follows immediately that if the game U is symmetric, in the sense that $U = (U^1, U^2) = (A, A')$, then it is a potential game if and only if $\Phi A \Phi$ is symmetric matrix, and that in this case the potential function $V = \Phi A \Phi + \Phi A \Xi + \Xi A' \Phi$ is a symmetric matrix as well. §

3.3 A Pairwise Characterization

To continue, we provide a characterization of normal form potential games that only requires comparisons of the payoffs of two players at a time. This characterization does not yield a formula for the potential function, and it is equivalent to the characterization from Theorem 3.2 in the two-player case (cf Example 3.4). On the other hand, in games with three or more players the new characterization requires fewer calculations to be checked.

Theorem 3.5. *$U = (U^1, \ldots, U^p)$ is a potential game if and only if*

$$U^p_{\{p,q\},\emptyset} = U^q_{\{p,q\},\emptyset} \quad \textit{for all } p, q \in \mathcal{P}. \tag{27}$$

Proof. We show that statement (27) is equivalent to statement (16) from Theorem 3.2.

If condition (27) holds and $\{p, q\} \subseteq Q \subseteq \mathcal{P}$, then $(U^p_{\{p,q\},\emptyset})_{Q,Q^c} = (U^q_{\{p,q\},\emptyset})_{Q,Q^c}$. Lemma 2.5 then implies that $U^p_{Q,Q^c} = U^q_{Q,Q^c}$, and so condition (16) holds.

On the other hand, if condition (16) holds, so that $U^p_{P,P^c} = U^q_{P,P^c}$ whenever $\{p, q\} \subseteq P$, then applying Lemmas 2.6 and 2.5 (and setting $P = Q \cup \{p, q\}$) yields

$$U^r_{\{p,q\},\emptyset} = \sum_{Q \subseteq \mathcal{P} - \{p,q\}} \left(U^r_{\{p,q\},\emptyset}\right)_{Q,(\mathcal{P}-\{p,q\})-Q} = \sum_{\{p,q\} \subseteq P \subseteq \mathcal{P}} U^r_{P,P^c}. \tag{28}$$

It follows that $U^p_{\{p,q\},\emptyset} = U^q_{\{p,q\},\emptyset}$, and so that condition (27) holds. ∎

By viewing $U^p(\cdot,\cdot,\hat{s}^{-\{p,q\}})$ and $U^q(\cdot,\cdot,\hat{s}^{-\{p,q\}})$ as $n^p \times n^q$ matrices and applying equation (9), we can rewrite condition (27) as

$$\text{(29)} \qquad \Phi\, U^p(\hat{s}^{-\{p,q\}})\, \Phi = \Phi\, U^q(\hat{s}^{-\{p,q\}})\, \Phi \quad \text{for all } \hat{s}^{-\{p,q\}} \in S^{-\{p,q\}} \text{ and } p, q \in \mathcal{P}.$$

Thus, while Example 3.4 showed that we can decide whether a two-player game is a potential game by checking a single matrix equality, condition (29) shows that we can decide whether an arbitrary normal form game is a potential game by checking a collection of matrix equalities.

4. Decision Algorithms

Suppose we are given a p-player normal form game U in which each player has a large number n of pure strategies. How can we determine whether U is a potential game?

The set of p-player games with n strategies per player is a linear space of dimension pn^p. Monderer and Shapley (1996, p. 141–142) observe that the set of potential games forms a subspace of dimension $n^p + pn^{p-1} - 1$.[4] It follows that the set of potential games can be characterized by $pn^p - (n^p + pn^{p-1} - 1) = (p-1)n^p - pn^{p-1} + 1$ linear equations.

By using Monderer and Shapley's (1996) definition (11), it is possible to check whether a given game is a potential game by verifying precisely this many linear equations. To determine whether a given function is a potential function, one needs to show that the equality in (11) holds for each strategy pair $\hat{s}^p, s^p \in S^p$, each opponents' strategy profile $s^{-p} \in S^{-p}$, and each player $p \in \mathcal{P}$, yielding $\frac{1}{2}n(n-1) \times p \times n^{p-1} = \frac{1}{2}pn^p(n-1)$ equalities altogether. But since

$$\text{(30)} \qquad U^p(j, s^{-p}) - U^p(i, s^{-p}) = \sum_{k=i}^{j-1} \left(U^p(k+1, s^{-p}) - U^p(k, s^{-p})\right)$$

for $j > i$, it is sufficient to verify the equality in (11) for pairs of adjacent strategies, reducing the number of equalities that must be checked to $pn^{p-1}(n-1)$. To construct

[4] I thank an anonymous referee for pointing out this result. The number $n^p + pn^{p-1} - 1$ can be derived as follows: Condition (12) tells us that the set of potential games can be expressed as the sum of $1 + p$ subspaces. Begin with the subspace of games $U = (V, \dots, V)$ in which all players have the same payoff function; this subspace has dimension n^p. Then for each player p, add the subspace containing the games $U = (\mathbf{0}, \dots, W^p, \dots, \mathbf{0})$ in which only player p obtains a nonzero payoff, and where this payoff does not depend on p's own strategy; each of these p subspaces is of dimension n^{p-1}. Inspection reveals that the intersection of the former subspace and the sum of the latter subspaces is of dimension 1; it consists of the games whose payoff functions U^p are constant and the same for all players. All told, the dimension of the set of potential games is thus $n^p + pn^{p-1} - 1$.

a candidate potential function, one first assigns the function an arbitrary value at an arbitrary initial strategy profile $s_0 \in S$. One then constructs a tree that reaches every other strategy profile in S, and whose edges all correspond to unilateral deviations to adjacent strategies. By adjusting the value of the potential function along the branches of the tree according to the changes in deviators' payoffs (cf footnote 3), one assigns a value to the candidate potential function at every strategy profile in S. This procedure ensures that $n^p - 1$ of the $pn^{p-1}(n-1)$ equations that must be checked are true by construction, leaving $pn^{p-1}(n-1) - (n^p - 1) = (p-1)n^p - pn^{p-1} + 1$ equations to be verified explicitly.

This argument shows that by constructing a candidate potential function, one can determine whether a game is a potential game using an algorithm whose running time is of order $\Theta(p\,n^p)$ in n and p.[5] However, this approach imposes significant storage requirements, as it requires us to record the n^p entries of the candidate potential function.

Using an algorithm based on our characterization from Theorem 3.5, we can achieve a significant reduction in storage requirements, at a running time that is comparable to that of the previous algorithm when p is small. For each pair of players, and for each of the n^{p-2} strategy profiles of the remaining $p-2$ players, this algorithm requires us to check the $n \times n$ matrix equality from condition (29); thus, $\Theta(p^2 n^p)$ scalar equalities must be checked in total. Even after accounting for the computation of the matrix products appearing in (29),[6] this algorithm has a running time of order $\Theta(p^2 n^p)$, while its storage requirements are only of order $\Theta(n^2)$.

One can reduce storage requirements further still by using an algorithm based on Monderer and Shapley's (1996) four-cycle condition, described in Theorem 3.1 above. Since in this case we must check an equality for each strategy pair $s^p, t^p \in S^p$, each strategy pair $s^q, t^q \in S^q$, each profile $s^{-\{p,q\}} \in S^{-\{p,q\}}$, and each pair of players $p, q \in \mathcal{P}$, the number of steps required is of order $\Theta(p^2 n^{p+2})$. However, unpublished work of Hino (2009) shows that it is enough to check a subset of these equalities, leading to an algorithm with a running time of order $\Theta(p^2 n^p)$.

---

[5]A function $g : \mathbf{N} \to \mathbf{R}_+$ has *order of growth* $\Theta(h(k))$ *in* k if there are constants $c, C > 0$ such that $g(k) \in [c\,h(k), C\,h(k)]$ for all sufficiently large k. Similarly, a function $g : \mathbf{N} \times \mathbf{N} \to \mathbf{R}_+$ has *order of growth* $\Theta(h(k,l))$ *in* k *and* l if (i) for each l_0, $g(\cdot, l_0)$ has order of growth $\Theta(h(k, l_0))$ in k, and (ii) if for each k_0, $g(k_0, \cdot)$ has order of growth $\Theta(h(k_0, l))$ in l.

[6]While the basic algorithm for multiplying two $n \times n$ matrices takes $\Theta(n^3)$ steps, the fact that the products appearing in (29) are of the form $\Phi M \Phi$ allows for a computation that requires only $\Theta(n^2)$ steps, so that all of the matrices needed can be computed in $\Theta(p^2 n^p)$ steps in total. To see this, write

$$\Phi M \Phi = \left(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\right) M \left(I - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\right) = M - [M\mathbf{1}]\tfrac{1}{n}\mathbf{1}' - \tfrac{1}{n}[\mathbf{1}'M] + [\mathbf{1}'M\mathbf{1}]\tfrac{1}{n^2}\mathbf{1}\mathbf{1}'.$$

Each expression in brackets can be computed in $\Theta(n^2)$ steps. Once this is done, each of the n^2 terms of $\Phi M \Phi$ can be computed in constant time, since the previous equation can be rewritten as $(\Phi M \Phi)_{ij} = M_{ij} - \frac{1}{n}(M\mathbf{1})_i - \frac{1}{n}(\mathbf{1}'M)_j + \frac{1}{n^2}(\mathbf{1}'M\mathbf{1})_{ij}$.

Theorem 4.1 (Hino (2009)). *U is a potential game if and only if for every $s^p \in S^p - \{n^p\}$, $s^q \in S^q - \{n^q\}$, $s^{-\{p,q\}} \in S^{-\{p,q\}}$, and $p, q \in \mathcal{P}$, the four-cycle equation* (13) *holds with $s = (s^p, s^q, s^{-\{p,q\}})$, $\hat{s} = (s^p + 1, s^q, s^{-\{p,q\}})$, $t = (s^p + 1, s^q + 1, s^{-\{p,q\}})$, and $\hat{t} = (s^p, s^q + 1, s^{-\{p,q\}})$.*

In words, Hino (2009) shows that to determine whether a game is a potential game, it is enough to check that the four-cycle equation holds for all *adjacent* pairs of strategies of players p and q. This is true because the four-cycle equation for arbitrary pairs of strategies of players p and q can be expressed in terms of four-cycle equations involving only adjacent pairs of strategies by means of telescoping sums (cf equation (30)).

It is worth noting that the algorithms based on Theorems 3.5 and 4.1 require the verification of exactly the same number of linear equalities.[7] Nevertheless, because of its lower storage requirements and setup costs, the algorithm based on Hino's (2009) characterization is preferable in practice.

A. Appendix

The Proof of Lemma 2.5

Fix $M \in \mathbf{R}^{\prod_{r\in\mathcal{P}} n^r}$ and $Q, R, \hat{Q}, \hat{R} \subseteq \mathcal{P}$. Three applications of equation (5) yield

$$M_{Q,R}(\tilde{s}) = \sum_{s\in S} M(s) \prod_{r\in\mathcal{P}} (T^r_{Q,R})_{s^r\tilde{s}^r},$$

$$\begin{aligned}(M_{Q,R})_{\hat{Q},\hat{P}}(\hat{s}) &= \sum_{\tilde{s}\in S} M_{Q,R}(\tilde{s}) \prod_{p\in\mathcal{P}} (T^p_{\hat{Q},\hat{R}})_{\tilde{s}^p\hat{s}^p} \\ &= \sum_{\tilde{s}\in S}\left(\sum_{s\in S} M(s) \prod_{r\in\mathcal{P}} (T^r_{Q,R})_{s^r\tilde{s}^r}\right) \prod_{p\in\mathcal{P}} (T^p_{\hat{Q},\hat{R}})_{\tilde{s}^p\hat{s}^p} \\ &= \sum_{s\in S} M(s) \prod_{p\in\mathcal{P}} \sum_{\tilde{s}\in S} (T^p_{Q,R})_{s^p\tilde{s}^p} (T^p_{\hat{Q},\hat{R}})_{\tilde{s}^p\hat{s}^p} \\ &= \sum_{s\in S} M(s) \prod_{p\in\mathcal{P}} (T^p_{Q,R} T^p_{\hat{Q},\hat{R}})_{s^p\hat{s}^p}, \quad \text{and}\end{aligned}$$

$$M_{Q\cup\hat{Q},R\cup\hat{R}}(\hat{s}) = \sum_{s\in S} M(s) \prod_{p\in\mathcal{P}} (T^p_{Q\cup\hat{Q},R\cup\hat{R}})_{s^p\hat{s}^p}.$$

[7] The number of equalities to be verified is $\frac{1}{2}p(p-1)n^{p-2}(n-1)^2$ in both cases. This is easy to see in the case of the algorithm based on Theorem 4.1. For the algorithm based on Theorem 3.5, observe that for each of the $\frac{1}{2}p(p-1)$ pairs of players, and each of the n^{p-2} strategy profiles of their opponents, one must check the $n \times n$ matrix equality from condition (3.5). But because since the matrix Φ is the projection of $\mathbf{R}^{n^p}$ onto the tangent space $TX^p = \{z^p \in \mathbf{R}^{n^p} : \sum_{i\in S^p} z^p_i = 0\}$, checking the equality from (3.5) on any $(n-1) \times (n-1)$ submatrix is enough to establish that the full matrix equality holds.

It follows from the last two equations that if

$$T^p_{Q,R} T^p_{\hat{Q},\hat{R}} = T^p_{Q \cup \hat{Q}, R \cup \hat{R}} \tag{31}$$

for all $p \in \mathcal{P}$, then $(M_{Q,R})_{\hat{Q},\hat{R}} = M_{Q\cup\hat{Q},R\cup\hat{R}}$, proving the lemma.

For any $\tilde{Q}, \tilde{R} \subseteq \mathcal{P}$, equation (7) specifies how the value of $T^p_{\tilde{Q},\tilde{R}}$ depends on the membership of p in $\tilde{Q}$ and $\tilde{R}$. The following table uses equation (7) and the fact that T^p is a projection (i.e., that $(T^p)^2 = T^p$) to determine the value of the product $T^p_{Q,R} T^p_{\hat{Q},\hat{R}}$ for each of the sixteen possible combinations of memberships of p in Q, R, $\hat{Q}$, and $\hat{R}$.

$T^p_{Q,R} T^p_{\hat{Q},\hat{R}}$	$p \in \hat{Q}^c \cap \hat{R}^c$ ($T^p_{\hat{Q},\hat{R}} = I$)	$p \in \hat{Q} \cap \hat{R}^c$ ($T^p_{\hat{Q},\hat{R}} = T^p$)	$p \in \hat{Q}^c \cap \hat{R}$ ($T^p_{\hat{Q},\hat{R}} = I - T^p$)	$p \in \hat{Q} \cap \hat{R}$ ($T^p_{\hat{Q},\hat{R}} = \mathbf{0}$)
$p \in Q^c \cap R^c$ ($T^p_{Q,R} = I$)	I	T^p	$I - T^p$	$\mathbf{0}$
$p \in Q \cap R^c$ ($T^p_{Q,R} = T^p$)	T^p	T^p	$\mathbf{0}$	$\mathbf{0}$
$p \in Q^c \cap R$ ($T^p_{Q,R} = I - T^p$)	$I - T^p$	$\mathbf{0}$	$I - T^p$	$\mathbf{0}$
$p \in Q \cap R$ ($T^p_{Q,R} = \mathbf{0}$)	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$

Meanwhile, equation (7) immediately specifies the values of $T^p_{Q\cup\hat{Q},R\cup\hat{R}}$ as a function of p:

$$T^p_{Q\cup\hat{Q},R\cup\hat{R}} = \begin{cases} I & \text{if } p \in (Q \cup \hat{Q})^c \cap (R \cup \hat{R})^c, \\ T^p & \text{if } p \in (Q \cup \hat{Q}) \cap (R \cup \hat{R})^c, \\ I - T^p & \text{if } p \in (Q \cup \hat{Q})^c \cap (R \cup \hat{R}), \\ \mathbf{0} & \text{if } p \in (Q \cup \hat{Q}) \cap (R \cup \hat{R}). \end{cases}$$

We can now confirm that equation (31) holds for all $p \in \mathcal{P}$. First, we verify that both sides of (31) equal I for the same values of p:

$$\begin{aligned} T^p_{Q,R} T^p_{\hat{Q},\hat{R}} = I &\Leftrightarrow p \in (Q^c \cap R^c) \cap (\hat{Q}^c \cap \hat{R}^c) \\ &\Leftrightarrow p \in (Q^c \cap \hat{Q}^c) \cap (R^c \cap \hat{R}^c) \\ &\Leftrightarrow p \in (Q \cup \hat{Q})^c \cap (R \cup \hat{R})^c \end{aligned}$$

$$\Leftrightarrow T^p_{Q\cup\hat{Q},R\cup\hat{R}} = I.$$

Next, we verify that both sides of (31) equal T^p for the same values of p:

$$\begin{aligned}
T^p_{Q,R}T^p_{\hat{Q},\hat{R}} = T^p &\Leftrightarrow p \in ((Q\cap R^c)\cap(\hat{Q}^c\cap\hat{R}^c)) \cup ((Q\cap R^c)\cap(\hat{Q}\cap\hat{R}^c)) \cup ((Q^c\cap R^c)\cap(\hat{Q}\cap\hat{R}^c)) \\
&\Leftrightarrow p \in (Q\cap R^c\cap\hat{R}^c) \cup ((Q^c\cap R^c)\cap(\hat{Q}\cap\hat{R}^c)) \\
&\Leftrightarrow p \in (Q\cup(Q^c\cap\hat{Q}^c))\cap(R^c\cap\hat{R}^c) \\
&\Leftrightarrow p \in (Q\cup\hat{Q})\cap(R\cup\hat{R})^c \\
&\Leftrightarrow T^p_{Q\cup\hat{Q},R\cup\hat{R}} = T^p.
\end{aligned}$$

The argument that both sides of (31) equal $I - T^p$ for the same values of p is very similar to the previous one, and for all remaining values of p, both sides of (31) must equal $\mathbf{0}$. This establishes (31), and so completes the proof of the lemma.

References

Beckmann, M., McGuire, C. B., and Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Hart, S. and Mas-Colell, A. (1989). Potential, value, and consistency. *Econometrica*, 57:589–614.

Hino, Y. (2009). An improved algorithm for detecting potential games. Unpublished manuscript, UBS.

Hofbauer, J. and Sigmund, K. (1988). *Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge.

Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, 14:124–143.

Rosenthal, R. W. (1973). A class of games possessing pure strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67.

Roughgarden, T. and Tardos, É. (2002). How bad is selfish routing? *Journal of the ACM*, 49:236–259.

Sandholm, W. H. (2001). Potential games with continuous player sets. *Journal of Economic Theory*, 97:81–108.

Sandholm, W. H. (2005). Negative externalities and evolutionary implementation. *Review of Economic Studies*, 72:885–915.

Sandholm, W. H. (2007). Pigouvian pricing and stochastic evolutionary implementation. *Journal of Economic Theory*, 132:367–382.

Sandholm, W. H. (2009). Large population potential games. *Journal of Economic Theory*, 144:1710–1725.

Slade, M. E. (1994). What does an oligopoly maximize? *Journal of Industrial Economics*, 42:45–51.

Ui, T. (2000). A Shapley value representation of potential games. *Games and Economic Behavior*, 31:121–135.