A 'DUAL'-IMPROVED SHORTCUT TO THE LONG RUN

By

Kareen Rozen

March 2008

COWLES FOUNDATION DISCUSSION PAPER NO. 1643

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

A 'Dual'-Improved Shortcut to the Long Run

Kareen Rozen [1]
Yale University

This Version: October 2006

[1]I am grateful to Eric Maskin and Wolfgang Pesendorfer for their support and comments on this paper. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Email kareen.rozen@yale.edu

Abstract

I use the theories of duality and optimal branchings to find a necessary and sufficient characterization of stochastically stable limit sets (SSLS) that helps improve the radius - modified coradius test of Ellison (2000). The improved shortcut I offer may permit the identification of SSLS when Ellison's radius - modified coradius test fails to identify any, or may be able to pinpoint the true SSLS in cases where Ellison's test identifies only a superset. I also demonstrate precisely why the radius - modified coradius test is not universally applicable and illuminate the connection between the modified coradius and the Lagrange multipliers of the optimal branching problem.

Keywords: Evolutionary games; stochastic stability; optimal branchings; extended radius; extended coradius; modified coradius.

JEL classification: C73.

1 Introduction

Since the seminal works of Foster & Young (1990) and Kandori, Mailath & Rob (1993), the solution concept of *stochastic stability* has been a valuable predictor of long run behavior in evolutionary games, especially since the task of identifying stochastically stable limit sets was facilitated by the work of Ellison (2000). Ellison's *radius - modified coradius* condition offers a useful test for checking whether a limit set is stochastically stable. However, the test will not work in all circumstances; that is, Ellison's condition is sufficient but not necessary.

In this paper I find a *necessary and sufficient* characterization of stochastically stable limit sets by using the theories of duality and optimal branchings. This characterization introduces two new measures, the *extended radius* and the *extended coradius*. Using these measures I demonstrate precisely why the radius - modified coradius test is not universally applicable, as well as illuminate the connection between the modified coradius and the Lagrange multipliers of the optimization problem that is solved to find the stochastically stable limit sets. Most importantly, the universally applicable characterization I provide allows me to offer an improved shortcut that may permit the identification of stochastically stable limit sets when Ellison's radius - modified coradius test cannot identify any, or may be able to pinpoint the exact limit set in cases where Ellison's test identifies only a superset.

The remainder of this paper is organized as follows. The formal framework and the solution concept of stochastic stability are discussed in Section 2. I explore the duality approach in Section 3 and introduce the necessary and sufficient characterization using the extended radius and extended coradius (Theorem 3.6). Sections 2 and 3 also discuss some existing methods for finding stochastically stable limit sets. In Section 4, I use the characterization derived in Theorem 3.6 in two ways. First, I derive Ellison's result as a corollary (Theorem 4.1) and explain why the radius - modified coradius test is not universally applicable. Second, I propose the improved shortcut for finding stochastically stable limit sets (Theorem 4.2).

Section 5 concludes.

2 Preliminaries

To encapsulate a wide variety of evolutionary games and dynamics, I employ the abstract framework of Ellison (2000).

Definition 2.1. A *model of evolution with noise* $(S, M, M(\varepsilon))$ is a family of Markov processes indexed by $\varepsilon \in [0, \bar{\varepsilon}]$ and having transition matrices $M(\varepsilon)$ on a finite state space S. The matrices $M(\varepsilon)$ are such that

(i) $M(\varepsilon)$ is ergodic for each $\varepsilon > 0$.

(ii) $M(\varepsilon)$ is continuous in ε and $M(0) = M$.

(iii) $\exists$ a cost function $c_S : S \times S \to \mathbb{R}^+ \cup \{\infty\}$ s.t. $\forall\, s, s' \in S$, $\lim_{\varepsilon \to 0} M_{ss'}(\varepsilon)/\varepsilon^{c_S(s,s')}$ exists and is strictly positive when $c_S(s, s') < \infty$; and $M_{ss'}(\varepsilon) = 0$ for small ε when $c_S(s, s') = \infty$.

This framework is sufficiently general to accommodate the various standard specifications of the underlying evolutionary game and its dynamics. Indeed, assuming that the population is finite, the state space S may be chosen to represent any combination of possible characteristics of play observed over a finite number of periods. Moreover, both the behavioral rules that the players (usually) follow as well as the stochastic shocks to the populations are absorbed into the transition matrices $M(\varepsilon)$.[1]

The cost function $c_S(s, s')$ captures how unlikely the transition from state s to state s' is when the level of noise ε is small. Though the cost function is not an explicit element of the model of evolution with noise, we shall see that it is the most important ingredient.

[1]This restricts the perturbations to stationary ones.

2.1 Stochastic stability

It is well known that any Markov process whose transition matrix is irreducible and aperiodic possesses a unique invariant distribution. Hence for $\varepsilon > 0$, the long run probability of observing a particular state $s \in S$ in the model with ε-noise is given by $\mu^\varepsilon(s)$, where $\mu^\varepsilon = \lim_{t\to\infty} m_0 M^t(\varepsilon)$ and m_0 is an arbitrary initial distribution.

The invariant distribution provides a description of behavior over an extended amount of time, given a particular level of noise. Rather than directly examine the invariant distribution for every level of noise, Foster & Young (1990) introduced the notion of stochastic stability. A stochastically stable set of states is a collection of states $\hat{S} \subseteq S$ to which each invariant distribution μ^ε assigns positive measure for all ϵ sufficiently close to zero.

Definition 2.2. A state (or set of states) is *stochastically stable* if the limiting distribution $\mu^* = \lim_{\varepsilon\to 0} \mu^\varepsilon$ assigns to it positive measure. Such a state (set of states) may also be referred to as a *stochastically stable limit state (set)*, or an *SSLS*.

2.2 The graph-theoretic connection

The task of computing the limiting distribution was simplified by a graph theoretic result given in Freidlin & Wentzell (1984) (henceforth FW). A directed graph is given by the pair (V, A), where V is a set of vertices and $A \subseteq V \times V$ is a set directed arcs. If $(v_1, v_2) \in A$, then the graph contains an arc, or arrow, emanating from v_1 and pointing into v_2. I shall now define a type of graph known as a v-branching.[2]

[2] A v-graph is known less formally as a v-graph in the terminology of FW and a v-tree in Kandori, Mailath & Rob (1993) and Young (1993). Formally speaking, trees require a more general connectedness condition instead of condition (i), while graphs require neither condition given. To be very precise, in graph theory a v-branching would have arrows pointing *away from* the root v rather than *towards* it; however the reversal in direction is important for the purpose of evolutionary games.

Definition 2.3. For $v \in V$, a *v-branching* is a directed graph $(V, \tilde{A})$ satisfying the following two conditions:

(i) **(Degree constraint)** $\forall\, v' \in V \backslash \{v\}$, $\exists!\, v'' \in V$ with $v' \neq v''$ and $(v', v'') \in \tilde{A}$.

(ii) **(No cycles)** If $\{(v_i, v_{i+1})\}_{i=1}^{N}$ is a sequence with $(v_i, v_{i+1}) \in \tilde{A}\ \forall\ i$, then $v_1 \neq v_{N+1}$.

If b_v is a v-branching, then I will often refer to v as the root of the branching b_v. While it is not explicitly stated in Definition 2.3, there is at least one arc entering the root. By condition (i), each node in $V \setminus \{v\}$ has a unique outgoing arc. If none of those arcs enters the root, then there will be a cycle, violating condition (ii).

Let B_v be the set of all possible v-branchings in the graph (V, A) and associate a branching $b \in \cup_{v \in V} B_v$ with the particular arcs it contains. Moreover, identify V with S and A with $S \times S$. Then, Lemma 3.1 in Chapter 6 of FW provides a connection between branchings and invariant distributions:

Lemma 2.4. ***(FW)*** *Consider an irreducible Markov chain with transition matrix $M(\varepsilon)$ on state space S. Then the invariant distribution is given uniquely by*

$$\mu^{\varepsilon}(s) = \frac{\sum_{b \in B_s} \prod_{(s', s'') \in b} M_{s', s''}(\varepsilon)}{\sum_{s' \in S} \sum_{b \in B_{s'}} \prod_{(s', s'') \in b} M_{s', s''}(\varepsilon)} \quad \forall\ s \in S.$$

Recall the cost function c_S defined in item (iii) of Definition 2.1. In their Theorem 1, Kandori, Mailath & Rob (1993) use the graph theoretic connection made by FW to offer the following result.[3]

Theorem 2.5 (Kandori, Mailath & Rob (1993))**.** *Let $C_S^*(s) = \min_{b \in B_s} \sum_{(s', s'') \in b} c_S(s', s'')$. The set of stochastically stable states is given by* $\operatorname{argmin}_{s \in S} C_S^*(s)$.

[3]Their Theorem 1 also contains a formula for the limiting distribution analogous to that of FW.

In other words, a state is stochastically stable if and only if it is the root of a branching of minimal total cost. Young (1993) shows further that one need only consider branchings on the set of recurrent communication classes R of the unperturbed Markov process M. A recurrent communication class r is a set of states such that if $\tilde{s} \notin r$, then $M_{s,\tilde{s}} = 0 \; \forall \; s \in r$, and if $s, s' \in r$, then there exists a finite $t \geq 1$ such that $M^t_{s,s'} > 0$.[4] Young's result means that rather than identifying V with the entire state space S, one may simply identify V with the set of recurrent classes R. To do this it will be necessary to modify the cost to be a set function $\hat{c}_R : R \times R \to \mathbb{R}^+ \cup \{\infty\}$ by defining $\hat{c}_R(r, r') = \min_{s \in r, s' \in r'} c_S(s, s')$ for $r, r' \in R$.

2.3 Ellison's radius - modified coradius

Though often simpler than solving for the limiting distribution directly, the problem of finding an optimal branchings is complicated by the fact that the number of possible branchings increases exponentially in the number of recurrent classes. In an attempt to circumvent the branching problem, Ellison (2000) suggested two new measures, the radius and modified coradius. While not universally applicable, in certain cases these measures help find stochastically stable states and bound the speed of evolutionary change.

Let $\rho \in \mathcal{P}(R)$ be a union of one or more recurrent classes of the unperturbed Markov process $M(0)$.[5] Ellison defines the basin of attraction of ρ to be the set of states from which the unperturbed Markov process will surely converge to ρ. That is, we let the basin of attraction be $D(\rho) = \{r \in R \mid \exists \; s \in r \text{ s.t. } \lim_{t\to\infty} 1_{\{s\}} M^t 1'_\rho = 1\}$.[6]

Let us define a path from a union of recurrent classes ρ to a union of recurrent

[4] Clearly, $M^t_{s,s'}$ refers to the s, s' element of the t-th power of the matrix M, not the value $M_{s,s'}$ raised to the t-th power.

[5] $\mathcal{P}(X)$ denotes the power set of X, minus the empty set.

[6] 1_X is an $|S|$-dimensional row vector with elements equal to 1 for states in X and 0 otherwise. The " ' " denotes transposition.

classes ρ' to be a finite sequence of distinct states $(s_1, s_2, \ldots, s_N)$ such that $s_1 \in \rho$, $s_i \notin \rho'$ for $2 \leq i < N$, and $s_N \in \rho'$. We denote the set of all paths from ρ to ρ' by $P(\rho, \rho')$ and the set of all paths by P. Extending the definition of cost to a path function $c_P : P \to \mathbb{R}^+ \cup \{\infty\}$ by setting $c_P(s_1, s_2, \ldots, s_N) = \sum_{i=1}^{N-1} c_S(s_i, s_{i+1})$, we may now define the radius and modified coradius.

Definition 2.6. The *radius* $\mathcal{R}(\rho)$ of the basin of attraction of ρ is the minimum cost of a path from ρ to $R \setminus D(\rho)$, or

$$\mathcal{R}(\rho) = \min_{(s_1, s_2, \ldots, s_N) \in P(\rho, R \setminus D(\rho))} c_P(s_1, s_2, \ldots, s_N). \tag{1}$$

The *modified coradius* $\mathcal{CR}^*(\rho)$ measures the difficulty of entering ρ when the cost of a path is normalized by the radii of the intermediate recurrent classes through which the path passes. Namely,

$$\mathcal{CR}^*(\rho) = \max_{\rho' \in R \setminus \rho} \min_{(s_1, s_2, \ldots, s_N) \in P(\rho', \rho)} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=2}^{N'-1} \mathcal{R}(r_i), \tag{2}$$

where $(r_1, r_2, \ldots, r_{N'})$ is the sequence of recurrent classes through which the path $(s_1, s_2, \ldots, s_N)$ consecutively passes.[7]

Ellison's Theorem 2 then provides a test for stochastic stability as well as a bound on the speed of evolution.

Theorem 2.7 (Ellison (2000))**.** *Let $(S, M, M(\varepsilon))$ be a model of evolution with noise and suppose that for some union ρ of recurrent sets of M, $\mathcal{R}(\rho) > \mathcal{CR}^*(\rho)$. Then,*

(i) The stochastically stable states are contained in ρ.

(ii) The longest expected wait until a state in ρ is reached is $O(\varepsilon^{-\mathcal{CR}^*(\rho)})$ as $\varepsilon \to 0$.[8]

[7] A recurrent class may appear more than once in the sequence, but not consecutively.

[8] $O(\varepsilon^{-\mathcal{CR}^*(\rho)})$ as $\varepsilon \to 0$ denotes an upper bound of $C\varepsilon^{-\mathcal{CR}^*(\rho)}$ for ε sufficiently small and some uniform constant $C > 0$.

3 An exact characterization

Ellison's radius - modified coradius method is a useful tool for pinpointing stochastically stable recurrent classes. However, as Ellison notes, the method is not universally applicable. In this section, I use results from the theory of optimal branchings to offer a necessary and sufficient characterization of stochastically stable sets from which Ellison's radius - modified coradius theorem follows as a corollary. The characterization I obtain, which follows from the dual of the optimal branching program, demonstrates precisely why the radius - modified coradius measure is not universally applicable. As will be seen in a later section, the characterization paves the way towards an improved shortcut that may work when the radius - modified coradius test cannot.

3.1 Edmonds' branching algorithm

I begin the analysis by considering Edmonds' algorithm for finding optimal branchings. This algorithm is discussed in fuller detail in Korte & Vygen (2002) and Magnanti & Wolsey (1995), to which I will henceforth refer as KV and MW, respectively. To illustrate this algorithm, consider the following example.

Let $V = \{A, B, C, D\}$ and select D as the root for the branching. Suppose that

$$\begin{array}{lll} c(A,B) = 1 & c(A,C) = 10 & c(A,D) = \infty, \\ c(B,A) = \infty & c(B,C) = 2 & c(B,D) = 3, \\ c(C,A) = 2 & c(C,B) = 6 & c(C,D) = 4. \end{array}$$

We shall want to choose the arcs of minimal cost emanating from each of A, B, and C. The arcs selected will not change if we reduce the costs of all the arcs emanating from each of A, B, and C by $\lambda^*(A) = 1$, $\lambda^*(B) = 2$, and $\lambda^*(C) = 2$, respectively. These values are exactly the minimal cost of exiting A, B, and C, respectively. It will be useful for later to note that if we were to identify V with the set of recurrent classes of M in a model $(S, M, M(\varepsilon))$ of evolution with noise, then these values would be radii. Now, let us create the arcs (A, B) (which has reduced cost $1 - 1 = 0$), (B, C) (which has reduced cost $2 - 2 = 0$), and (C, A) (which has reduced cost $2 - 2 = 0$). This forms the cycle ABC, which we shall call a pseudo node.

Definition 3.1. If C is a cycle formed between nodes $v_1, v_2, \ldots, v_N \in V$, then the *pseudo node* $v_C \in \mathcal{P}(V)$ is the set of nodes $\{v_1, v_2, \ldots, v_N\}$.

What arcs may be formed from the pseudo node ABC? Originally, both B and C could form arcs to D. The reduced costs of those arcs would be $3 - 2 = 1$ and $4 - 2 = 2$ respectively. Therefore, we may consider the arc to D emanating from ABC to be of cost 1 (and having B as its origin). Let us then say that $\lambda^*(ABC) = 1$ is the (reduced) cost of the cycle ABC. In order to form a D-branching on the graph composed of the nodes ABC and D, we must choose this arc (ABC, D). Since the arc originally emanates from B, and each node in a branching may have only one outgoing arc, let us remove the previous arc emanating from B, which was (B, C). This leaves us with the branching $C \to A \to B \to D$. The reader may easily check that this is the D-branching of minimal cost. Observe that this procedure defines a function $\lambda^* : \mathcal{P}(V) \to \mathbf{R}$ which measures a *reduced* cost of exiting cycles that are created in the process.

The procedure used in this example is known more formally as Edmonds' branching algorithm and is fully described in the appendix. It is useful to explore the duality theory approach taken by MW in their proof of the validity of this algorithm. Using such an approach, I will be able to obtain an exact characterization of stochastically stable states, as well as derive Ellison's result as a corollary. Troeger (2002) builds on this algorithm to find SSLS, and Hasker (2004) implicity uses this type of algorithm to help provide a characterization of SSLS, as well as tests to pinpoint SSLS. Unlike those papers, we will take a duality theory approach to the problem of finding SSLS.[9]

[9]Further connections between those works and my own will be explored in future versions of this paper.

As is well known, the problem of finding the cheapest v^r-branching in (V, A) is given by the following linear program:

$$\min_{b\in\mathbb{R}^{|A|}} \sum_{(v,v')\in A} c(v,v')\, b_{(v,v')} \tag{3}$$

subject to

$$\sum_{(v,v')\in A} b_{(v,v')} = 1 \qquad \forall\, v \in V \setminus v^r \tag{4}$$

$$\sum_{(v,v')\in A, v\in \hat{V}, v'\notin \hat{V}} b_{(v,v')} \geq 1 \qquad \forall\, \hat{V} \in \mathcal{P}(V \setminus v^r) \tag{5}$$

$$b \geq 0 \qquad \text{and integer.} \tag{6}$$

The constraints of the problem force any feasible vector b to be a vector of 1's (corresponding to edges in the branching) and 0's (corresponding to the rest).

Recall from the theory of linear programming that if $\min\{ya : yH \geq d, y \geq 0\}$ is the primal problem, then the dual problem is given by $\max\{dx : Hx \leq a, x \geq 0\}$.[10] Take H to be an $|A| \times |\mathcal{P}(V \setminus v^r)|$-dimensional matrix with entries $H_{(v',v''),\hat{V}} = 1_{\{v'\in\hat{V}, v''\notin\hat{V}\}}$; and take d to be an $|\mathcal{P}(V \setminus v^r)|$-dimensional row vector of 1's and a to be an $|A|$-dimensional column vector with entries $c(v, v')$ for each $(v, v') \in A$. Dropping just the integrality constraint in (6), it is easy to see that the dual of the primal program (3)-(6) is

$$\max_{\lambda\in\mathbb{R}^{|\mathcal{P}(V\setminus v^r)|}} \sum_{\hat{V}\in\mathcal{P}(V\setminus v^r)} \lambda_{\hat{V}} \tag{7}$$

subject to

$$\sum_{\hat{V}\in\mathcal{P}(V\setminus v^r)} \lambda_{\hat{V}} 1_{v\in\hat{V}, v'\notin\hat{V}} \leq c(v, v') \qquad \forall\, (v, v') \in A. \tag{8}$$

[10]The reader may refer to KV for an overview of linear programming.

A solution of (7)-(8) is precisely a vector of Lagrange multipliers corresponding to the constraints of the primal problem. Moreover, the classic Duality Theorem posits an important relationship between the optimal values of the primal and dual programs.

Theorem 3.2 (Duality Theorem)**.** *A feasible vector y solves the primal problem* $\min\{ya : yH \geq d, y \geq 0\}$ *if and only if there exists a vector x, feasible for the dual problem* $\max\{dx : Hx \leq a, x \geq 0\}$*, for which the objective functions have equal value.*

Consider the function $\lambda^*(\cdot)$, whose values are determined in the execution of Edmonds' algorithm, and set $\lambda^*_{\hat{V}} = \lambda^*(\hat{V})$ for all $\hat{V} \in \mathcal{P}(V \setminus v^r)$. One may check that λ^* is a feasible vector for the dual problem, and that both λ^* and the vector b representing the arcs selected by Edmond's algorithm lead to the same objective value. Therefore, the branching b must be optimal by the Duality Theorem. The Duality Theorem not only reinforces the validity of Edmonds' algorithm, but also motivates my ensuing analysis.

3.2 Extended radii

We will now consider the graph $(R, R \times R)$ equipped with the cost set function c_R. As a convention throughout, we ignore the "arcs" on the diagonal of the product set $R \times R$, that is, we do not permit arcs of the form (r, r) for $r \in R$. Recall that one may arbitrarily resolve multiplicities arising in steps 2 and 3 of Edmonds' branching algorithm. Let us formalize this by supposing that when a single arc must be chosen amongst multiple arcs of minimal cost, a choice function $\tau : \mathcal{P}(R) \times \mathcal{P}(\mathcal{P}(R)) \to \mathcal{P}(R)$ is used as a tie-breaking rule. That is, presented with a node (or set of nodes, corresponding to a pseudo node), an arc will form between this (pseudo) node and the (pseudo) node chosen by τ. Moreover, we denote by $\mathcal{C}_r(\tau, k; (R, R \times R))$ the pseudo nodes formed in the k-th iteration of Edmonds' algorithm when the optimal r-branching in $(R, R \times R)$ is sought; I will suppress the last argument when the graph in question is unambiguous. If the algorithm has reached step 4 by the k-th iteration, then $\mathcal{C}_r(\tau, k') = \emptyset$ for every $k' \geq k$. Finally, I will write $\mathcal{C}_r(\tau) = \cup_{1 \leq k < \infty} \mathcal{C}_r(\tau, k)$.

Definition 3.3. Let λ^* and λ'^* be the solutions of the dual problems (7)-(8) corresponding to rooting at r_1 and r_2, respectively. For any choice rules τ, τ and recurrent classes $r_1, r_2 \in R$, let the *extended radius* from r_1 to r_2 be given by

$$\mathcal{ER}(r_1, r_2) = \mathcal{R}(r_1) + \sum_{C \in \mathcal{C}_{r_2}(\tau) \setminus \mathcal{C}_{r_1}(\tau)} \lambda_C'^*. \tag{9}$$

Define the *extended coradius* of r_1 by

$$\mathcal{ECR}(r_1) = \max_{r_2 \in R \setminus r_1} \left\{ \mathcal{R}(r_2) + \sum_{C \in \mathcal{C}_{r_1}(\tau) \setminus \mathcal{C}_{r_2}(\tau)} \lambda_C^* - \sum_{C \in \mathcal{C}_{r_2}(\tau) \setminus \mathcal{C}_{r_1}(\tau)} \lambda_C'^* \right\}. \tag{10}$$

Given a choice rule τ, $\sum_{C \in \mathcal{C}_{r_2}(\tau) \setminus \mathcal{C}_{r_1}(\tau)} \lambda_C^*$ is the sum of the Lagrange multipliers corresponding to the set of cycles that are formed when rooting at r_2, but are not formed when rooting at r_1. Observe that while the set of cycles formed in the execution of the algorithm may depend on the choice rule employed, the appellations "extended radius" and "extended coradius" offer no such hint of dependence. To assuage any concerns of ambiguity, I prove that the values of the extended radii and coradii are independent of the choice rule. The proof will use the following lemma, which I prove in the appendix.

Lemma 3.4. *Let (V, A) be a graph with cost and choice functions c and τ, and take $v_1, v_2 \in V$. Construct an artificial node z, and set $V' = V \setminus \{v_1, v_2\} \cup z$ and $A' = V' \times V'$. Let the auxiliary graph (V', A') have cost and choice functions $c' : A' \to \mathbb{R}^+ \cup \infty$ and $\tau' : \mathcal{P}(V') \times \mathcal{P}(\mathcal{P}(V')) \to \mathcal{P}(V')$ defined by*

$$c'(v, v') = \begin{cases} c(v, v') & \text{if } v, v' \in V' \setminus \{z\} \\ \min_{k \in \{v_1, v_2\}} c(k, v') & \text{if } v = z \text{ and } v' \in V' \setminus \{z\} \\ \min_{k \in \{v_1, v_2\}} c(v, k) & \text{if } v \in V' \setminus \{z\} \text{ and } v' = z \end{cases} \tag{11}$$

and

$$\tau'(X,Y) = \begin{cases} \tau(X,Y) & \text{if } \tau(X,Y) \notin \{\{v_1\},\{v_2\}\} \\ \{z\} & \text{otherwise.} \end{cases} \tag{12}$$

Then, $\forall\ k \geq 1$, $\{C \in \mathcal{C}_{v_1}(\tau, k; (V, A)) \mid v_2 \notin C\} = \mathcal{C}_z(\tau, k; (V', A'))$*; and the respective dual solutions* λ^* *and* λ'^* *agree on the cycles in* $\mathcal{C}_z(\tau, k; (V', A'))$*.*

Proposition 3.5. *For all choice rules* τ, τ' *and recurrent classes* $r_1, r_2 \in R$*,*

$$\sum_{C \in \mathcal{C}_{r_1}(\tau) \setminus \mathcal{C}_{r_2}(\tau)} \lambda^*_C = \sum_{C' \in \mathcal{C}_{r_1}(\tau') \setminus \mathcal{C}_{r_2}(\tau')} \lambda^*_{C'}. \tag{13}$$

Proof. Applying the values λ^* found by executing Edmonds' algorithm using a choice rule τ and root node r_1, it is clear that the optimal value of the dual problem (7)-(8) is $\sum_{r \in R \setminus \{r_1\}} \mathcal{R}(r) + \sum_{C \in \mathcal{C}_{r_1}(\tau)} \lambda^*_C$. Since the optimal value is independent of τ, so must be $\sum_{C \in \mathcal{C}_{r_1}(\tau)} \lambda^*_C$. Because $\mathcal{C}_{r_1}(\tau) = \big(\mathcal{C}_{r_1}(\tau) \setminus \mathcal{C}_{r_2}(\tau)\big) \cup \big(\mathcal{C}_{r_1}(\tau) \cap \mathcal{C}_{r_2}(\tau)\big)$, it remains to show that $\sum_{C \in \mathcal{C}_{r_1}(\tau) \cap \mathcal{C}_{r_2}(\tau)} \lambda^*_C$ is independent of τ.

To prove this, I construct an auxiliary graph where the nodes r_1 and r_2 are united into a larger limit set ρ. An application of Lemma 3.4 to the graph $(R, R \times R)$ using $v_1 = r_1$ and $v_2 = r_2$ shows that the cycles formed in each stage of the algorithm when finding the optimal ρ-branching in the auxiliary graph $(R', R' \times R')$ are the same as those cycles not containing r_2 that are formed in each stage when finding the optimal r_1-branching in the original graph. In addition, the corresponding coefficients in the optimal vector λ'^* are the same as those in λ^*.

As $\sum_{r \in R' \setminus \{\rho\}} \mathcal{R}(r) + \sum_{C' \in \mathcal{C}_\rho(\tau'; (R', R' \times R'))} \lambda'^*_{C'}$ is the optimal cost of a ρ-branching, we know $\sum_{C' \in \mathcal{C}_\rho(\tau'; (R', R' \times R'))} \lambda'^*_{C'}$ must be independent of τ'. By Lemma 3.4, $\sum_{C' \in \mathcal{C}_\rho(\tau'; (R', R' \times R'))} \lambda'^*_{C'} = \sum_{C \in \mathcal{C}_{r_1}(\tau, k; (R, R \times R)), r_2 \notin C} \lambda^*_C$, so the right-hand side is also independent of τ, from which τ' is defined. Moreover, the symmetry in Lemma 3.4 implies that

$$\{C \in \mathcal{C}_{r_1}(\tau, k; (R, R \times R)) \mid r_2 \notin C\} = \mathcal{C}_\rho(\tau', k; (R', R' \times R')) = \{C \in \mathcal{C}_{r_2}(\tau, k; (R, R \times R)) \mid r_1 \notin C\}.$$

Taking unions over all $k \geq 1$, we obtain the chain of equalities

$$\{C \in \mathcal{C}_{r_1}(\tau; (R, R{\times}R)) \mid r_2 \notin C\} = \mathcal{C}_\rho(\tau'; (R', R'{\times}R')) = \{C \in \mathcal{C}_{r_2}(\tau; (R, R{\times}R)) \mid r_1 \notin C\}.$$

Hence $\mathcal{C}_{r_1}(\tau; (R, R \times R)) \cap \mathcal{C}_{r_2}(\tau; (R, R \times R)) = \{C \in \mathcal{C}_{r_1}(\tau; (R, R \times R)) \mid r_2 \notin C\}$. This observation completes the proof, since then $\sum_{C \in \mathcal{C}_{r_1}(\tau) \cap \mathcal{C}_{r_2}(\tau)} \lambda^*_C$ is independent of τ. □

One may consider the extended radius from the recurrent class r to the class r' to be a measure of the cost reduction incurred by rooting an optimal branching at r rather than at r'. The formulation of the extended coradius of r is intended to emulate Ellison's modified coradius, and provides a worst case measure of the cost incurred by rooting at r. The following theorem uses these measures to offer a precise characterization of stochastically stable classes.

Theorem 3.6. *Let $(S, M, M(\varepsilon))$ be a model of evolution with noise and let $r \in R$. Then, the following two conditions are equivalent:*

(i) $\mathcal{R}(r) \geq \mathcal{ECR}(r)$

(ii) $\mathcal{ER}(r, r') \geq \mathcal{ER}(r', r)\ \forall\ r' \in R \setminus \{r\}$*;*

and either condition holds if and only if r is stochastically stable. Moreover, r is the unique stochastically stable recurrent class if the inequality holds strictly.

Proof. The equivalence between (i) and (ii) is self-evident. Using Theorem 2.5 (KMR, 1993) in conjunction with Young's restriction to recurrent classes, r is stochastically stable if and only if it is the root of a branching of minimal c_R cost on $(R, R \times R)$. Let τ be a choice rule and r' an arbitrary recurrent class in $R \setminus \{r\}$. Also let λ^* and λ'^* be the solutions of the dual problems (7)-(8) corresponding to rooting at r and r', respectively. By the duality theorem, the cost of an optimal branching rooted at r is $\sum_{r \in R \setminus \{r\}} \mathcal{R}(r) + \sum_{C \in \mathcal{C}_r(\tau)} \lambda^*_C$, and analogously for r'.

Hence r is stochastically stable if and only if

$$\sum_{\tilde{r}\in R\setminus\{r\}} \mathcal{R}(\tilde{r}) + \sum_{C\in\mathcal{C}_r(\tau)} \lambda_C^* \leq \sum_{\tilde{r}\in R\setminus\{r'\}} \mathcal{R}(\tilde{r}) + \sum_{C\in\mathcal{C}_{r'}(\tau)} \lambda_C'^*,$$

or equivalently,

$$\mathcal{R}(r) \geq \mathcal{R}(r') + \sum_{C\in\mathcal{C}_r(\tau)} \lambda_C^* - \sum_{C\in\mathcal{C}_{r'}(\tau)} \lambda_C'^*.$$

By Lemma 3.4 and the ensuing analysis in the proof of Proposition 3.5,

$$\sum_{C\in\mathcal{C}_r(\tau)\cap\mathcal{C}_{r'}(\tau)} \lambda_C^* = \sum_{C\in\mathcal{C}_r(\tau)\cap\mathcal{C}_{r'}(\tau)} \lambda_C'^*.$$

Due to the decomposition $\mathcal{C}_r(\tau) = \big(\mathcal{C}_r(\tau)\setminus\mathcal{C}_{r'}(\tau)\big) \cup \big(\mathcal{C}_r(\tau)\cap\mathcal{C}_{r'}(\tau)\big)$, r is stochastically stable if and only if $\mathcal{R}(r) \geq \mathcal{ECR}(r)$, as claimed. Following this line of argument with strict inequality yields the additional claim on uniqueness. $\square$

The following corollary may offer a useful test for ruling out stochastic stability of a particular recurrent class.

Corollary 3.7. *Consider a recurrent class $r \in R$. If $\exists\ r' \in R \setminus \{r\}$ such that every cycle containing r also contains r' and $\mathcal{R}(r') > \mathcal{R}(r)$, then r cannot be stochastically stable.*

It is important to note that to find the dual solution for each potential root node, one need not execute Edmonds' algorithm $|R|$ times. While Edmonds' algorithm finds the optimal branching given a particular root, it is well known that the problems of finding an optimal rooted branching and finding an optimal branching are mathematically equivalent (KV, Proposition 6.6). In fact, a minor transformation of the graph permits Edmonds' algorithm to find the optimal root for a branching. Simply construct an anchoring node a and let $R' = R\cup\{a\}$. Equip the graph $(R', R'\times R')$ with an augmented cost function $c_{R'} : R'\times R' \to \mathbb{R}^+\cup\{\infty\}$, which agrees with c_R on $R \times R$ and has

$$c_{R'}(r, a) = |R| \max_{r_1, r_2 \in R,\ c(r_1, r_2) < \infty} c(r_1, r_2) \ \forall\, r \in R.^{11} \tag{14}$$

Applying Edmonds' algorithm to find the optimal a-branching in $(R', R' \times R')$, it is clear that due to the sizeable cost of each arc to a, only one node r^* will ultimately be connected to the root. Moreover, because the last pseudo node formed is the entire set R, the node r^* selected to connect to the anchor a will be the one that minimizes

$$c_{R'}(r, a) - [\ \mathcal{R}(r) + \sum_{C \in \mathcal{C}_a(\tau) \setminus \{R\}, r \in C} \lambda^*_C\ ],$$

where τ is the choice rule used and λ^* is the dual solution found from the execution of the algorithm when rooting at a. Because the cost of an arc to a is constant, this is equivalent to choosing the node r^* that maximizes $\mathcal{R}(r) + \sum_{C \in \mathcal{C}_a(\tau) \setminus \{R\}, r \in C} \lambda^*_C$. Moreover, the optimality of the a-branching implies that the sub-branching rooted at r^* must also be of minimal cost, making r^* stochastically stable. The following proposition shows we may indeed use this alternative setup to provide an equivalent definition of the extended and modified coradii.

Proposition 3.8. *Let $r_1, r_2 \in R$ and $(R', R' \times R')$ be defined as above with dual solution λ^* when rooting at a under the choice rule τ. The formulation given in Definition 3.3 is equivalent to*

$$\mathcal{ER}(r_1, r_2) = \mathcal{R}(r_1) + \sum_{C \in \mathcal{C}_a(\tau), r_1 \in C, r_2 \notin C} \lambda^*_C \tag{15}$$

and

$$\mathcal{ECR}(r_1) = \max_{r_2 \in R \setminus r_1} \left\{\ \mathcal{R}(r_2) + \sum_{C \in \mathcal{C}_a(\tau), r_2 \in C, r_1 \notin C} \lambda^*_C - \sum_{C \in \mathcal{C}_a(\tau), r_1 \in C, r_2 \notin C} \lambda^*_C\ \right\}. \tag{16}$$

[11] The value of $c_{R'}(a, r)$ for each $r \in R$ will be irrelevant.

Proof. An application of Lemma 3.4 with $v_1 = a$ and $v_2 = r_2$ obtains the auxiliary graph $(\tilde{R}, \tilde{R} \times \tilde{R})$ satisfying $\forall\, k \geq 1$, $\{C \in \mathcal{C}_a(\tau, k) \mid r_2 \notin C\} = \mathcal{C}_{\{a,r_2\}}(\tau', k; (\tilde{R}, \tilde{R} \times \tilde{R}))$; moreover the respective dual solutions λ^* and $\tilde{\lambda}^*$ agree on the identical cycles. Because of the enormous cost of an arc to a, it is clear that $\forall\, k \geq 1$, $\mathcal{C}_{\{a,r_2\}}(\tau', k; (\tilde{R}, \tilde{R} \times \tilde{R})) = \mathcal{C}_{r_2}(\tau, k; (R, R \times R))$;[12] and that once again, the respective dual solutions $\tilde{\lambda}^*$ and λ^{2*} agree on the identical cycles. Similarly, if we apply Lemma 3.4 again with $v_1 = a$ and $v_2 = r_1$ and take into consideration the cost of an arc to a, we obtain that $\forall\, k \geq 1$, $\{C \in \mathcal{C}_a(\tau, k) \mid r_1 \notin C\} = \mathcal{C}_{r_1}(\tau, k; (R, R \times R))$. Combining these results, we have that $\forall\, k \geq 1$,

$$\begin{aligned} \mathcal{C}_{r_2}(\tau, k) \setminus \mathcal{C}_{r_1}(\tau, k) &= \{C \in \mathcal{C}_a(\tau, k) \mid r_2 \notin C\} \setminus \{C \in \mathcal{C}_a(\tau, k) \mid r_1 \notin C\} \\ &= \{C \in \mathcal{C}_a(\tau, k) \mid r_1 \in C, r_2 \notin C\}, \end{aligned}$$

and λ^{2*} and λ^* agree on the identical cycles. The proposition immediately follows. □

4 An improved shortcut

In this section I offer a shortcut which improves on the radius - modified coradius test in Ellison (2000). While neither my shortcut nor the radius - modified coradius test is universally applicable, the improved shortcut may permit the identification of SSLS when Ellison's radius - modified coradius test fails to identify any, or may be able to pinpoint the true SSLS in cases where Ellison's test identifies only a superset. My improved test builds upon the radius - extended coradius characterization presented in Theorem 3.6 as well as Ellison's test itself. Theorem 4.1 demonstrates precisely why Ellison's test follows as a corollary of Theorem 3.6. This information turns out to be useful in constructing the improved shortcut. The content of Theorem 4.1 is therefore not an alternate derivation of Ellison's result (it is already known that the result may be proved using tree arguments, as Ellison

[12] Because $R \subset R'$, the same choice rule τ may apply.

himself demonstrated in the appendix of (2000)); rather, the content of Theorem 4.1 is its role as a building block for the improved shortcut and an explanation of why the radius - modified coradius test is not universally applicable.

Theorem 4.1. *Part (i) of Ellison's result in Theorem 2.7 is a corollary of Theorem 3.6. In particular, this follows for each* $r \in R$ *because* $\max_{r' \in R \setminus \{r\}} \mathcal{ER}(r', r) \leq \mathcal{CR}^*(r)$*, hence* $\mathcal{ECR}(r) \leq \mathcal{CR}^*(r)$*. It follows for a union* ρ *of recurrent classes because* $\mathcal{R}(\rho) > \mathcal{CR}^*(\rho)$ *implies that for each* $r' \notin \rho$ *there is some* $r^*_{r'} \in \rho$ *such that* $\mathcal{ER}(r^*_{r'}, r') > \mathcal{ER}(r', r^*_{r'})$.

Proof. The proof is relegated to the appendix. □

To gain some intuition for the proof of Theorem 4.1, note that Ellison's modified coradius is derived from a shortest path problem. Finding a shortest path and finding an optimal branching are distinct mathematical problems; and while both problems may be solved by a greedy algorithm (Edmonds' in the case of branchings, and Dijkstra's, for example, in the case of a shortest path), the optimality criteria in the greedy steps differ. The branching algorithm myopically chooses arcs of minimal cost, while the shortest path algorithms take into account both the cost to an intermediate node and the cost of the shortest path from the intermediate node to the destination. As a result of these differing optimality criteria, the modified coradius ends up overcompensating as a measure of the cost of rooting at a particular node.

Why is this? Recall that the extended radius and coradius are composed of sums of Lagrange multipliers, each having the form of the cost of a transition minus the radii of certain recurrent classes; this bears some resemblance to the modified coradius. Because of the particular greedy criterion used in the branching algorithm, the Lagrange multipliers will be smaller than similar terms present in the modified coradius. In fact, Theorem 4.1 shows that the modified coradius always overestimates the extended coradius when the object in question is a single recurrent class. $\mathcal{CR}^*(r)$ disregards the negative term in the extended coradius that corresponds to the benefit of rooting at r. Secondly, the shortest path calculation overestimates the positive term in the modified coradius, which measures the persistence of r relative to another recurrent class r'. Moreover, for a union ρ of recurrent classes, it shows that the force behind Ellison's condition $\mathcal{R}(\rho) > \mathcal{CR}^*(\rho)$ is that each recurrent class not in ρ is "dominated" in the sense of the extended radius by some recurrent class in ρ. All that is going on behind Ellison's statement on unions of recurrent classes is that when ρ satisfies $\mathcal{R}(\rho) > \mathcal{CR}^*(\rho)$, then for each $r' \notin \rho$ there is some $r^*_{r'} \in \rho$ such that $\mathcal{ER}(r^*_{r'}, r') > \mathcal{ER}(r', r^*_{r'})$.

This permits the construction of the following test for stochastic stability. As usual, we identify a cycle with the union of the recurrent classes that it contains. We say that a first-phase cycle is one created in the first pass of Edmond's algorithm. This includes singleton "cycles." Essentially, the set of first-phase cycles is the set of nodes used at the beginning of the second phase of the algorithm. Finally, we say that Ellison's test selects some set of recurrent classes $\rho \subset R$ when $\mathcal{R}(\rho) > \mathcal{CR}^*(\rho)$.

Theorem 4.2. *Let C be a first-phase cycle. Set $r_C \in \arg\min_{r \in C} [c(r, R \setminus \mathcal{D}(C)) - \mathcal{R}(r)]$ and $C^* = \arg\max_{r \in C} \mathcal{R}(r)$. Then,*

(i) A recurrent class $r \in C^*$ is a SSLS whenever it satisfies

$$\mathcal{R}(r) + c(r_C, R \setminus \mathcal{D}(C)) - \mathcal{R}(r_C) \geq \mathcal{CR}^*(r), \tag{17}$$

and is the unique SSLS if the inequality is satisfied strictly.

(ii) If C^* is non-singleton and Ellison's test cannot select C^*, then every element of C^* is a SSLS if it satisfies

$$\mathcal{R}(C^*) + c(r_C, R \setminus \mathcal{D}(C)) - \mathcal{R}(r_C) = \mathcal{CR}^*(C^*). \tag{18}$$

Moreover, C^* satisfies (18) iff every $r \in C^*$ satisfies (17) with equality.

Note from Corollary 3.7 that if $r \in C \setminus C^*$ then r cannot be a SSLS. Before proving Theorem 4.2, let us discuss condition (17). Note that $c(r_C, R \setminus \mathcal{D}(C)) \geq \mathcal{R}(C)$ and that $\mathcal{R}(C^*) - \mathcal{R}(r_C) \geq 0$ because r_C is an element of the cycle. Therefore, the LHS of (17) is at least as large as the LHS of Ellison's condition given in part (i) of Theorem 2.7, while the RHS in both conditions is the same. This indicates that condition (17) could hold for a particular first-phase cycle even when $\mathcal{R}(C) \leq \mathcal{CR}^*(C)$, i.e., it may hold even when Ellison's condition would not be able to identify C as containing the SSLS. Moreover, because C is a first-phase cycle rather than an arbitrary union of recurrent classes ρ, we can pinpoint precisely the subset of SSLS in C: the elements of maximal radius within C. That is, even if Ellison's condition would identify C as a superset of the SSLS, this condition could sharpen the prediction.

Theorem 4.2 follows from the dual-based representation in Theorem 3.6, the result that $\mathcal{CR}^*(r) \geq \mathcal{ECR}(r)$ from Theorem 4.1, and the following lemma, which is proved in the appendix.

Lemma 4.3. *Let C be a first-phase cycle and $C^* = \arg\max_{r \in C} \mathcal{R}(r)$ a non-singleton set. Then exactly one of the following holds:*

(i) $\mathcal{R}(C^*) > \mathcal{CR}^*(C^*)$, so Ellison's test selects C^*;

(ii) $\mathcal{CR}^*(r) = \mathcal{CR}^*(C^*)$ for all $r \in C^*$.

The proof of this lemma is rather simple.

Proof. We prove that if $\mathcal{CR}^*(r) > \mathcal{CR}^*(C^*)$ for some $r \in C^*$ then Ellison's test can be used to prove that C^* contains all the SSLS. Because the maximum is taken over a larger set, it is clear that $\mathcal{CR}^*(r) \geq \mathcal{CR}^*(C^*)$ for every r. In fact, because C^* is a first-phase cycle and transitions between the elements of the cycle occur at radial cost,

$$\mathcal{CR}^*(r) = \max\{\mathcal{CR}^*(C^*), \max_{r' \in C^* \setminus \{r\}} \mathcal{R}(r')\}.$$

Therefore, if $\mathcal{CR}^*(r) > \mathcal{CR}^*(C^*)$ for some r then there is some $r' \in C^* \setminus \{r\}$ such that $\mathcal{R}(r') > \mathcal{CR}^*(C^*)$. But by construction of C^*, $\mathcal{R}(r') = \mathcal{R}(C^*)$, implying that Ellison's test selects C^*. □

5 Discussion

In this paper I have taken a duality-based approach to the problem of calculating SSLS in the analysis of evolutionary games. In doing so, I have found a necessary and sufficient characterization of SSLS which illuminates the connection between the modified coradius of Ellison (2000) and the Lagrange multipliers of the optimal branching problem, and reveals why the radius - modified coradius test is not universally applicable. Using my characterization I have proposed an alternate test that may be able to either identify the SSLS when Ellison's radius - modified coradius cannot or pinpoint the true SSLS in cases where Ellison's test identifies only a superset.

Appendix

Edmond's Branching Algorithm

Consider a digraph (V, A) equipped with a cost function $c : V \times V \to \mathbb{R}^+ \cup \{\infty\}$ and fix a root node $v^r \in V$.

Step 0: Initialize. Define $\lambda^* : \mathcal{P}(V) \to \mathbb{R}^+ \cup \{\infty\}$ by setting $\lambda^*(v) = \min_{v' \in V \setminus v} c(v, v')$ for each $v \in V \setminus v^r$ and zero otherwise. Define $c' : \mathcal{P}(V) \times \mathcal{P}(V) \to \mathbb{R}^+ \cup \{\infty\}$, setting $c'(v, v') = c(v, v') - \lambda^*(v)$ for every $v, v' \in V \setminus v^r$ and c' equal to infinity otherwise. Go directly to Step 2, letting $((\tilde{V}, \tilde{A}), \tilde{c})$ be $((V, A), c')$.

Step 1: Reduce costs. For each new pseudo node[13] $v_C \in \tilde{V}$, let $\lambda^*(v_C) = \min_{v' \in \tilde{V} \setminus v_C} \tilde{c}(v_C, v')$ and $c'(v_C, v') = \tilde{c}(v_C, v') - \lambda^*(v_C)$ for every $v' \in \tilde{V} \setminus v_C$. Proceed to Step 2, letting $((\tilde{V}, \tilde{A}), \tilde{c})$ be $((\tilde{V}, \tilde{A}), c')$.

Step 2: Find the node greedy solution. For each $v \in \tilde{V} \setminus v^r$, choose one $v'_v \in \operatorname{argmin}_{v' \in \tilde{V} \setminus v} \tilde{c}(v, v')$ and let $A' = \bigcup_{v \in \tilde{V} \setminus v^r} \{v, v'_v\}$. Proceed to Step 3, letting $((\tilde{V}, \tilde{A}), \tilde{c})$ be $((\tilde{V}, A', \tilde{c})$.

Step 3: Contract. Let $C(\tilde{V}, \tilde{A})$ be the set of directed cycles in $(\tilde{V}, \tilde{A})$. If $C(\tilde{V}, \tilde{A}) = \emptyset$, then skip directly to Step 4, using this same $(\tilde{V}, \tilde{A})$, and ignore the rest of this step. Otherwise, if $C(\tilde{V}, \tilde{A}) \neq \emptyset$, then replace each $C \in C(\tilde{V}, \tilde{A})$ with a pseudo node v_C to obtain V'. To obtain A' and c', let arcs in $\tilde{A}$ incident to $v \in \tilde{V}$ remain arcs (of the same $\tilde{c}$ cost) incident[14] to the pseudo node v_C containing v; for parallel arcs,[15] allow only a single arc of minimal $\tilde{c}$ cost. Go to Step 1, letting $((\tilde{V}, \tilde{A}), \tilde{c})$ be $((V', A'), c')$.

Step 4: Expand. To obtain A', for each pseudo node v_C in $\tilde{V}$ and arc $(v_C, v') \in \tilde{A}$, remove the arc in C that emanates from the source (pseudo) node of arc (v_C, v'). To obtain V', replace the pseudo node v_C with the (pseudo) nodes constituting

[13] That is, a pseudo node which did not exist in the previous iteration of the algorithm.

[14] An arc is incident to v if it has the form (v, v') or (v', v) for some $v' \in V$

[15] Arcs (v, v') and (v'', v''') are parallel if v, v'' are contained in the same pseudo node and v', v''' are contained in the same pseudo node; it could also be that exactly one of the origin or destination is a node rather than a pseudo node, so for example $v' = v'''$ and v' is not contained in a pseudo node.

the cycle C. Repeat Step 4, letting $(\tilde{V}, \tilde{A})$ be (V', A'), until $V' = V$.

Proof of Lemma 3.4

By strong induction on k. The statement is clearly true when $k = 1$, since c' and τ' have been defined so that in the first pass of the algorithm, if $v, v' \in V \setminus \{v_1, v_2\}$, then arc (v, v') is selected in (V, A) if and only if arc (v, v') is selected in (V', A'). The dual solutions corresponding to pseudo nodes formed when $k = 1$ are clearly identical. Now assuming the result is true up through some general k, I prove it for $k + 1$. There are three major cases to examine.

(i) $\mathcal{C}_{v_1}(\tau, k; (V, A)) = \emptyset$. By the induction hypothesis, $\mathcal{C}_z(\tau', k; (V', A')) = \emptyset$ too. The algorithm in both graphs is complete, hence so is the proof.

(ii) $\{C \in \mathcal{C}_{v_1}(\tau, k; (V, A)) \mid v_2 \notin C\} = \emptyset$, but a pseudo node containing v_2 is formed. By the induction hypothesis, $\mathcal{C}_z(\tau', k; (V', A')) = \emptyset$. Since any new cycles formed in the $(k + 1)$-st pass for the graph (V, A) would have to contain v_2, the claim is also valid for $k + 1$.

(iii) $\{C \in \mathcal{C}_{v_1}(\tau, k; (V, A)) \mid v_2 \notin C\} \neq \emptyset$. By the induction hypothesis, $\{C \in \mathcal{C}_{v_1}(\tau, k; (V, A)) \mid v_2 \notin C\} = \mathcal{C}_z(\tau', k; (V', A'))$. Take $C \in \mathcal{C}_z(\tau', k; (V', A'))$. Three subcases arise in the $(k + 1)$-st iteration of the algorithm in (V', A').

   (a) At outgoing arc is formed from C to a single node $v \in V' \setminus \{z\}$ such that an arc exists from v to a node in C. This creates a larger pseudo node $C \cup \{v\}$. By the case $k = 1$, r is also a singleton node in (V, A). Due to the corresponding definitions of c and c' and τ and τ', $C \cup \{v\}$ must also form in (V, A).

   (b) At outgoing arc is formed from C to a node $v \in V' \setminus \{z\}$ such that there does not exist an arc from v to a node in C. This does not create a new pseudo node in (V', A'). If v is not in a pseudo node containing v_2 in (V, A), no cycle forms there either; but if v is in a pseudo node containing v_2 and if a cycle forms, that cycle contains v_2. In either case the statement is valid for $k+1$.

(c) At outgoing arc is formed from C to another pseudo node C' not containing v_2. If a cycle is formed, C' is also present in (V, A) and the same cycle forms due to the corresponding definitions of c and c' and τ and τ'.

It is clear by the strong inductive step and the definitions of c and c' that the dual solutions again agree on the corresponding cycles. This completes the proof of the lemma.

Proof of Theorem 4.1

For each $r' \in R$, fix the shortest path and let $(r_1, r_2, \ldots, r_{N_{r'}-1}, r_{N_{r'}})$ be the sequence of recurrent classes through which the shortest path from $r' = r_1$ to $r = r_{N_{r'}}$ consecutively passes. I shall also fix the choice rule τ and keep it in mind implicitly in what follows. Let $\bar{k}$ be the final iteration of the algorithm when rooting at a, and for each $r' \in R$ and $1 \le k \le \bar{k}$, let $C(r', k)$ be the new pseudo node containing r' that forms in the k-th iteration of Edmonds' algorithm; if no such pseudo node forms, then $C(r', k) = \emptyset$. Recall from the steps of the algorithm that $\lambda^*_{C(r',k)} = 0$ if $C(r', k) = \emptyset$. For the sake of notational simplicity, I will also write $\mathcal{C}_a(r', r) = \{C \in \mathcal{C}_a \mid r' \in C, r \notin C\}$, and in a slight abuse of notation, will let $r' \in \mathcal{C}_a(r', r)$ mean that $\exists\ C \in \mathcal{C}_a(r', r)$ such that $r' \in C$. Finally, for each $r' \in R$, let $k_{r'} = \min_{1 \le k \le \bar{k}, C(r',k) \in \mathcal{C}_a(r',r)} k$ if $\{1 \le k \le \bar{k}, C(r', k) \in \mathcal{C}_a(r', r)\} \neq \emptyset$, and $k_{r'} = \bar{k}$ otherwise.

Recall the equivalent formulation of the extended coradius offered in Proposition 15. To prove the theorem, I will in fact prove something stronger, that $\forall$ $r, r' \in R$ and an arbitrary fixed choice rule τ,

$$\sum_{C \in \mathcal{C}_a(\tau), r' \in C, r \notin C} \lambda^*_C \le \min_{(s_1, s_2, \ldots, s_N) \in P(r', r)} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=1}^{N_{r'}-1} \mathcal{R}(r_i), \qquad (19)$$

where λ^* is the dual solution obtained when rooting at the anchor a. Step 1 proves Equation (19). Given Equation (19), one may show that $\max_{r' \in R\setminus\{r\}} \mathcal{ER}(r',r) \leq \mathcal{CR}^*(r)$ simply by adding $\mathcal{R}(r') = \mathcal{R}(r_1)$ to both sides and taking the maximum over all $r' \in R\setminus\{r\}$ on both sides of (19). Since $\sum_{C \in \mathcal{C}_a(\tau), r \in C, r' \notin C} \lambda^*_C$ is nonnegative, it is clear that $\max_{r' \in R\setminus\{r\}} \mathcal{ER}(r',r) \geq \mathcal{ECR}(r)$. The extension to the case when ρ is a union of two or more recurrent classes is dealt with in Step 2.

Step 1: Let us assume that $\mathcal{C}_a(r',r) \neq \emptyset$, else Equation (19) holds trivially. Consider the first transition (r_1, r_2) in the shortest path. Clearly $r_1 = r \in \mathcal{C}_a(r',r)$. If $r_2 \notin \mathcal{C}_a(r',r)$, then the proof is complete. To see this, note that by the definition of λ^*, $c_R(r',r_2) - \mathcal{R}(r') - \sum_{j=1}^{k-1} \lambda^*_{C(r',j)} \geq \lambda^*_{C(r',k)}$ $\forall\, 1 \leq k \leq \bar{k} - 1$ (since new cycles are not formed in the $\bar{k}$-th iteration). In general, consider the transition (r_i, r_{i+1}) for $1 \leq i \leq N-1$ and recall that $k_{r_i} = \bar{k}$ if $r_i \notin \mathcal{C}_a(r',r)$. If $r_i \in \mathcal{C}_a(r',r)$, then the definition of λ^* implies the relation

$$c_R(r_i, r_{i+1}) - \mathcal{R}(r_i) - \sum_{j=1}^{k_{r_i}-1} \lambda^*_{C(r_i,j)} - \sum_{j=k_{r_i}}^{\hat{k}_{r_i}-1} \lambda^*_{C(r',j)} \geq \lambda^*_{C(r',\hat{k}_{r_i})} \quad \forall\, k_{r_i} \leq \hat{k}_{r_i} \leq k_{r_{i+1}} - 1.$$

Drop the first (nonnegative) summation in the above equation and rearrange to obtain

$$c_R(r_i, r_{i+1}) - \mathcal{R}(r_i) \geq \sum_{j=k_{r_i}}^{\hat{k}_{r_i}} \lambda^*_{C(r',j)} \quad \forall\, k_{r_i} \leq \hat{k}_{r_i} \leq k_{r_{i+1}} - 1. \tag{20}$$

Denote $I = \{1 \leq i < N_{r'} \mid k_{r_i} < k_{r_{i+1}}\}$. I will need to construct a few sequences as follows. Initialize $j_1 = \max_{i \in I, r_i \in \mathcal{C}_a(r',r), r_{i+1} \notin \mathcal{C}_a(r',r)} i$ and $\hat{k}_{r_{j_1}} = \bar{k}$. For each $l \geq 1$ until $k_{r_{j_l}} = 1$, I inductively define $j_l = \max_{i \in I, i < j_{l-1}, k_{r_i} \leq k_{r_{j_{l-1}}} \leq k_{r_{i+1}}} i$ and $\hat{k}_{r_{j_l}} = k_{r_{j_{l-1}}} - 1$. Since $k_{r'} = 1$, the sequence will eventually terminate at some $\bar{l}$. For each $1 \leq l \leq \bar{l}$, apply the values $k_{r_{j_l}}$ and $\hat{k}_{r_{j_l}}$ in the inequality in (20). Summing the resulting inequalities yields (19), as desired.

Step 2: Now take ρ to be a union of two or more recurrent classes and

$\rho' \subseteq R \setminus \rho$. Note that if $r \in \rho$ and $r' \in \rho'$, then $P(r', r) \supseteq \{(s_1, \ldots, s_N) \in P(\rho', \rho) \mid s_1 \in r', s_N \in r\}$. This means

$$\min_{P(r',r)} c_P(s_1, \ldots, s_N) - \sum_{i=1}^{N_{r'}-1} \mathcal{R}(r_i) \leq \min_{P(\rho',\rho), s_1 \in r', s_N \in r} c_P(s_1, \ldots, s_N) - \sum_{i=1}^{N_{r'}-1} \mathcal{R}(r_i).$$

Therefore, using (19)

$$\sum_{C \in \mathcal{C}_a(\tau), r' \in C, r \notin C} \lambda_C^* \leq \min_{(s_1, s_2, \ldots, s_N) \in P(\rho', \rho), s_1 \in r', s_N \in r} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=1}^{N_{r'}-1} \mathcal{R}(r_i). \tag{21}$$

Bounding the LHS below by adding the negative term $-\sum_{C \in \mathcal{C}_a(\tau), r \in C, r' \notin C} \lambda_C^*$ and taking the minimum over all $r \in \rho$ on both sides of (21) gives

$$\begin{aligned} &\min_{r \in \rho} \sum_{C \in \mathcal{C}_a(\tau), r' \in C, r \notin C} \lambda_C^* - \sum_{C \in \mathcal{C}_a(\tau), r \in C, r' \notin C} \lambda_C^* \\ &\leq \Big[\min_{r \in \rho} \min_{(s_1, s_2, \ldots, s_N) \in P(\rho', \rho), s_1 \in r', s_N \in r} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=2}^{N_{r'}-1} \mathcal{R}(r_i) \Big] - \mathcal{R}(r'). \end{aligned} \tag{22}$$

Clearly, $\min_{r \in \rho} \min_{(s_1, s_2, \ldots, s_N) \in P(\rho', \rho), s_1 \in r', s_N \in r} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=2}^{N_{r'}-1} \mathcal{R}(r_i)$ is bounded above by

$$\max_{\rho' \subseteq R \setminus \rho} \max_{r' \in \rho'} \min_{r \in \rho} \min_{(s_1, s_2, \ldots, s_N) \in P(\rho', \rho), s_1 \in r', s_N \in r} c_P(s_1, s_2, \ldots, s_N) - \sum_{i=2}^{N_{r'}-1} \mathcal{R}(r_i), \tag{23}$$

and a moment's reflection shows that the object in (23) is none other than $\mathcal{CR}^*(\rho)$. Ellison's hypothesis is that $\mathcal{CR}^*(\rho) < \mathcal{R}(\rho)$. By definition, $\mathcal{R}(\rho) \leq \mathcal{R}(r)\ \forall\ r \in \rho$. Therefore, for any $r \in \rho$, $\mathcal{R}(r)$ offers a strict upper bound for the bracketed term in (22). For each $r' \in \rho'$, let $r_{r'}^*$ be the minimizer of the LHS of (22). Coupling

(22) with this upper bound for the bracketed term gives

$$\sum_{C \in \mathcal{C}_a(\tau), r' \in C, r^*_{r'} \notin C} \lambda^*_C - \sum_{C \in \mathcal{C}_a(\tau), r^*_{r'} \in C, r' \notin C} \lambda^*_C < \mathcal{R}(r^*_{r'}) - \mathcal{R}(r'). \tag{24}$$

With a slight rearrangement, Equation (24) says that $\mathcal{ER}(r^*_{r'}, r') > \mathcal{ER}(r', r^*_{r'})$, thereby violating the necessary and sufficient condition for the stochastic stability of r'. Since r' was an arbitrary element of $R \setminus \rho$, the stochastically stable classes must be contained in ρ.

References

Ellison, G. 2000. "Basins of Attraction, Long Run Stochastic Stability, and the Speed of Step-by-Step Evolution." *Review of Economic Studies* 67:17–45.

Foster, D. & H.P. Young. 1990. "Stochastic Evolutionary Game Dynamics." *Theoretical Population Biology* 38:219–232.

Freidlin, M. & A. Wentzell. 1984. *Random Perturbations of Dynamical Systems*. New York: Springer Verlag.

Hasker, D. 2004. "The Emergent Seed: Simplifying the Analysis of Dynamic Evolution." *Working Paper* .

Kandori, M., G. Mailath & R. Rob. 1993. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica* 61:29–56.

Korte, B. & J. Vygen. 2002. *Combinatorial Optimization*. 2nd ed. New York: Springer.

Magnanti, T.L. & L.A. Wolsey. 1995. Optimal Trees. In *Network Models*, ed. M.O. Ball, T.L. Magnanti, C.L. Monma & G.L. Nemhauser. Vol. 7 of *Handbooks in Operations Research and Management Science* Amsterdam, The Netherlands: Elsevier Science pp. 503–615.

Troeger, T. 2002. "Why Sunk Costs Matter for Bargaining Outcomes: An Evolutionary Approach." *Journal of Economic Theory* 102:375–402.

Young, H. P. 1993. "The Evolution of Conventions." *Econometrica* 61:57–84.