An Efficient Algorithm for Testing Epsilon-Approximate Core Membership in Negotiation Games

by

Greg Bodwin

An honors thesis for the Department of Computer Science

Tufts University, 2013

Abstract

We study a game theoretic model of many-party negotiations, in which each player can perform work that benefits the rest of the group. We examine the *core* [8] of the game, which is the set of potential outcomes that leave no group of negotiators with an incentive to break from the agreement and pick a new outcome. More specifically, we are interested in whether or not core membership of an arbitrary point can be tested in polynomial time. This is motivated by the recent economic movement towards *bounded rationality* [7], which holds that a game theoretic solution concept is poor if its justification relies on economic agents solving computationally infeasible problems.

We first show that the problem is NP-Hard in a basic model of negotiations. We then add to our model the standard economic assumption of diminishing marginal returns, and show that this assumption is sufficient to push the problem into P.

Contents

Chapter 1

Introduction

1.1 Introduction

We will begin with a motivating example, inspired by the work done in [5].

There are three neighbouring countries named Left, Middle, and Right. A river runs through the three countries, with Left upstream of Middle and Middle upstream of Right. Additionally, the climate of the surrounding area is such that there is a constant prevailing wind, with Right upwind of Middle and Middle upwind of Left.

The countries each have a healthy amount of economic production, and this production produces some pollution as a side effect. The sludge dumped into the river will flow downstream, and the smog released into the air will flow downwind, so each country feels the effects of each other country's pollution. The kings of the three countries realize that perhaps all three countries would be happier if each were willing to restrict its personal levels of production - damaging the country's economic growth, but improving the air and water quality of all three countries - in exchange for a promise that the other countries will do the same. The kings convene at the negotiating table in an effort to write a pollution reduction treaty.

As economists, we might be interested in trying to predict whether the treaty written by the kings is *stable*, or if some of its members have an incentive to abandon the agreement. In this three-country game, we will define stability as the intersection of the following three properties:

1. There is no other possible treaty that every country prefers to the treaty they picked. If there were a "better" treaty, then we would expect our countries to abandon the current treaty in favor of the better one.

2. The treaty should leave no country worse off than it was before the treaty was signed. Otherwise, this country will choose to ignore the treaty and return to their individually rational behavior, and its citizens will be better off.

3. No *two* countries should have a course of action that, regardless of the third country's behavior, leaves these countries happier than they are under the current treaty. If they did, the countries would have an incentive to jointly abandon the treaty in favor of this superior course of action.

In game theory language, the set of treaties that obey these conditions is called the *core* [8].

But perhaps these conditions are a bit presumptuous. What if condition (1) is *not* satisfied - there exists a superior treaty to the one that was signed - but our countries lack the ability to figure out the specifics of the superior treaty? Then perhaps a treaty can be stable despite violating the first condition. We must concede that the conditions we have described only make economic sense if there exists a strategy (algorithm) to test for them in a reasonable (polynomial) amount of time.

This sort of application of computational complexity to economic equilibria is called *bounded rationality* [7], and it is precisely what we will study in this paper. We will adapt a game created by Elliott & Golub in [5], using our game to describe arbitrary n-player negotiations with any network of benefit flows. We will similarly generalize the three criteria for stability listed above to arbitrary n-player games. We will then attempt to determine whether or not the criteria can be checked in time polynomial in n.

History is not on our side. We are aware of three other games in which the core membership problem has been studied, and the problem was shown to be NP-Complete twice [6, 3] and Co-NP-Complete once [10] in the three different types of game. Each time, the culprit is the exponential blowup in the definition of the core: with n players in the game, we need to make sure that every possible coalition of these players (of which there are $2^n - 1$) cannot favorably change their behavior from the current outcome. A successful algorithm will therefore require more finesse than a simple brute-force search of every available subset of players.

Our first result falls in line with these prior results: we show that core membership testing, in a basic model of negotiation games, is NP-Hard. However, we show that with a simple economic assumption, *diminishing marginal returns*, the problem admits a polynomial-time algorithm.

In part two, we will give the details of our generalized model of negotiations, we state the corresponding computational problem, we give our proof of the NP-Hardness of this problem, and we introduce the new assumption of diminishing marginal returns.

In part three, we will develop an economic structure underlying our model of negotiations. In part four, we will use this economic structure to build an algorithm that checks the result of a negotiation for stability in polynomial time.

1.2 Notation Conventions

Throughout this paper, the following conventions will be used:

1. The max operator on vectors is meant termwise: if $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, then $\max(\mathbf{a}, \mathbf{b}) = \langle \max(a_1, b_1), \ldots, \max(a_n, b_n) \rangle$.
2. When $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we write $\mathbf{a} = \mathbf{b}, \mathbf{a} \geq \mathbf{b}, \mathbf{a} > \mathbf{b}$ to mean that $a_i = b_i, a_i \geq b_i, a_i > b_i$ (respectively) for all $i \in N$. We write $\mathbf{a} \gneq \mathbf{b}$ to mean that $a_i \geq b_i$ for all $i \in N$, *and* $a_j > b_j$ for some $j \in N$.
3. Suppose $\mathbf{a} \in \mathbb{R}^n$ and $C \subset \{1, \ldots, n\}$. Then $\mathbf{a}_C$ denotes the $|C|$-length vector induced by deleting the entries of $\mathbf{a}$ whose indices do not appear in C. The remaining entries of $\mathbf{a}_C$ are ordered as in $\mathbf{a}$. For example, $\langle 10, 11, 12, 13, 14 \rangle_{\{2,3,5\}} = \langle 11, 12, 14 \rangle$.
4. The set $\mathbb{R}^n_{\geq \mathbf{0}}$ denotes $\{r \in \mathbb{R}^n \,|\, r \geq \mathbf{0}\}$.
5. We write $\mathbf{e}^j$ to refer to the j^{th} basis vector: the element of $\mathbb{R}^n$ with a 1 in the j^{th} place and 0 elsewhere (n will be clear from context).
6. We use $\mathbf{J}(\mathbf{v} \,|\, \mathbf{u})$ to notate the Jacobian[1] of the function $\mathbf{u}$ at the vector $\mathbf{v}$. Most often $\mathbf{u}$ will be clear from context, so we will simply write $\mathbf{J}(\mathbf{v})$.

[1]For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian is the $n \times m$ matrix whose $(j,k)^{th}$ entry is given by $\frac{\partial f_k}{\partial x_j}(\mathbf{x})$.

Chapter 2

Preliminaries and Problem Statement

2.1 Introduction to Cooperative Game Theory

We will begin with a short introduction to cooperative game theory. Our objective is not to comprehensively survey the study of cooperative games; rather, it is to supply just enough background for a reader unfamiliar with game theory to understand negotiation games and the role of the core in cooperative game theory.

Classical game theory models strategic interactions between self-interested agents. It assumes a cut-throat environment in which deals are unenforceable: a player would gladly deviate from a plan, ruining it for everyone else, just to achieve a slightly more favorable outcome for himself.

Recently, the study of *cooperative games* [4] has emerged in an effort to model interactions that are not so cut-throat. In these strategic interactions, players have the ability to make enforceable agreements before actions are taken. For example, negotiations might be backed up by treaties or contracts such that players must stick to the agreed-upon course of action under penalty of law.

Game theorists use *solution concepts* to sort through the many possible outcomes of a game and select a small handful that seem particularly likely to occur. The most important and widely applied solution concept in cooperative games is called *the core* [8]. The purpose of this part is to introduce the core, discuss its importance, and demonstrate how it can predict the outcome of a certain simple type of cooperative game. We will also introduce *negotiation games*, another type of cooperative game that will serve as the central object of study in this paper. The model of negotiation games is original, but it is adapted from a similar one introduced by Elliott & Golub [5].

2.1.1 The General Form of Games

We use the following original definition of a game. This definition is *not* sufficient for every type of game studied in game theory, but it is sufficient for the games that we will consider in this paper.

A game is a tuple composed of the following components:

1. A set N of players in the game

2. A set Ω of possible outcomes of the game

3. A set Γ_i of available *pure strategies* for each player $i \in N$. A pure strategy is a complete contingency plan: it tells a player the exact strategic decision they should make in every situation they might find themselves in, given the information they have about the game's history so far.

4. A *rules* function $R : \Gamma_1 \times \cdots \times \Gamma_n \to \Omega$. This function decides which outcome the game is in, given the behavior chosen by the players.

5. A total ordering for each player on the set of outcomes Ω for each player i. If player i prefers outcome a to b, we write $a >_i b$. We similarly write $a <_i b$ or $a =_i b$ if player i prefers b to a or is indifferent between the two. We write $a > b$ if every player prefers a to b. The notations $=$, $\geq$, and $\gneq$ are defined similarly.[1]

Here is a basic example. Suppose you and I decide to play three consecutive rounds of rock-paper-scissors, and whoever wins the most rounds wins the game. Then:

1. (Set of Players) $N = \{\text{me, you}\}$.

2. (Set of Outcomes) Ω is the set of ordered triples of ordered pairs of elements of {rock, paper, scissors}. For example, in the outcome {(rock, paper), (scissors, rock), (paper, scissors)}, you have defeated me three times in a row.

3. (Pure Strategies) In round 1, my pure strategies are simply $\Gamma^1_{me} = \{\text{rock, paper, scissors}\}$. In round 2, my pure strategies Γ^2_{me} are the functions from ordered pairs of {rock, paper, scissors} (representing the outcome of round 1) to {rock, paper, scissors} (representing my choice in round 2). In round 3, my pure strategies Γ^3_{me} are the functions from pairs of ordered pairs of {rock, paper, scissors} (representing the outcome of rounds 1 and 2) to

[1]Some games require that players choose between an average outcome or a probabilistic mix over a superior and inferior outcome. In these games, a total ordering is not sufficient - we need a *utility function* to accurately predict strategic behavior. However, in this paper, a total ordering will suffice.

{rock, paper, scissors} (representing my choice in round 3). Then $\Gamma_{me} = \Gamma^1_{me} \times \Gamma^2_{me} \times \Gamma^3_{me}$. Since the game is symmetric, Γ_{you} is identical.

4. (Rules) $R(\Gamma_{me}, \Gamma_{you})$ is evaluated as follows: suppose $\Gamma_{me} = (\gamma^1_{me}, \gamma^2_{me}, \gamma^3_{me})$ and $\Gamma_{you} = (\gamma^1_{you}, \gamma^2_{you}, \gamma^3_{you})$. Let $g_1 = (\gamma^1_{me}, \gamma^1_{you})$. Let $g_2 = (\gamma^2_{me}(g_1), \gamma^2_{you}(g_1))$. Let $g_3 = (\gamma^3_{me}(g_1, g_2), \gamma^3_{you}(g_1, g_2))$. Then $R(\Gamma_{me}, \Gamma_{you}) = (g_1, g_2, g_3)$.

5. (Preference Ordering) Suppose $g = (g_1, g_2, g_3), h = (h_1, h_2, h_3)$ are two possible outcomes of the game. If g contains more elements of $\{$(rock, scissors), (scissors, paper), (paper, rock)$\}$ than it does elements of $\{$(scissors, rock), (paper, scissors), (rock, paper)$\}$, then call it *won*; if it has the same number of elements of the two sets call it *drawn*, if it has more of the second set than the first then call it *lost*. Classify h similarly. Then my preference function is as follows: $g >_{me} h$ if I won g but lost or drew h or if I drew g and lost h. Similarly, $g =_{me} h$ if I won both, drew both, or lost both, and otherwise I have $g <_{me} h$. Your preference function is defined symmetrically.

2.1.2 Pareto Efficiency

Rock-paper-scissors is an unusual game in that it is fully competitive: I can only win at your expense, and vice-versa. Other games are less adversarial; there might be outcomes a, b such that we both prefer a to b. In this sort of game, we should intuitively reject the idea that the game might end up at outcome b, since there is pressure - and no opposition - to switch the agreement from b to a. This is the most fundamental solution concept in cooperative games, and it is called *Pareto efficiency*.

Definition 2.1. An outcome a is called a *Pareto improvement* on an outcome b if $a \gneq b$.[2] If no Pareto improvement on outcome a exists, then outcome a is called *Pareto efficient*. The set of Pareto efficient outcomes is called the *Pareto frontier*.

In rock-paper-scissors, every outcome is Pareto efficient. This is because if I prefer outcome a to b, you necessarily prefer outcome b to outcome a, so no outcome is a Pareto improvement on another. Of course this is not true in general; we will later see some games that contain non-Pareto efficient outcomes.

2.1.3 The Core

Pareto efficiency is an appealing solution concept because it implements our intuitive notion of stability. Outcome b is not *stable* if there is an incentive to switch to

[2]This means that $a \geq_i b$ for all $i \in N$ and $a >_i b$ for some $i \in N$.

outcome a. The core is a refinement of Pareto efficiency that applies this idea to every coalition of players, not just the entire group.

Definition 2.2. Let $C = \{1, \ldots, c\}$, and let $\omega \in \Omega$. If there exist pure strategies $\gamma_1 \in \Gamma_1, \ldots, \gamma_c \in \Gamma_C$ such that for every set of pure strategies $\gamma_{c+1} \in \Gamma_{c+1}, \ldots, \gamma_n \in \Gamma_n$ we have $R(\gamma_1, \ldots, \gamma_n) \gneq_C \omega$, then we say $(\gamma_1, \ldots, \gamma_c)$ is a *deviation* for C from ω.

An outcome $\omega \in \Omega$ is in the *core* of the game if no nonempty coalition has a deviation from ω.

In plain English: suppose our players agree to outcome ω, but coalition C threatens to change its behavior to reach a new outcome that C will prefer. If C will succeed even when $N - C$ does its best to behave destructively and spite C, then we should believe that ω is an unstable outcome since it leaves C with an incentive to split from the group. The core is the solution concept that implements this notion of stability.

Note that the definition of Pareto efficiency is subsumed in the definition of the core; that is, every core outcome is Pareto efficient. This can be seen easily by considering the coalition $C = N$.

2.1.4 An Example: Allocation Games

One simple type of cooperative game is an *allocation game*, also known as a *transferable utility game*.

Definition 2.3. An *allocation game* is described by a pair (N, v). It has the following properties:

1. The set of players in the game is given by $N = \{1, \ldots, n\}$.
2. The element v is a function from $2^N \to \mathbb{R}$ with $v(\emptyset) = 0$. It represents the value that a given subset of players (a *coalition*) is capable of creating. Further, v is monotonic: whenever $C' \subset C$ we have $v(C') \leq v(C)$.
3. An outcome is any pair of the form $(C, \mathbf{d})$, where (1) $C \subset N$, (2) $\sum_{i \in N} d_i = v(C)$, and (3) $\mathbf{d}_{N-C} = \mathbf{0}$. Intuitively, this represents a division of the wealth created among the players who participated in its creation.
4. The pure strategies are as follows: first, each player simultaneously chooses between {participate, don't participate}. Then, $v(C)$ amount of wealth is divided evenly among the participating players. Finally, each participating player can donate any amount of their newly acquired wealth to any other participating player.

5. The rules are such that if C is the coalition of players that choose *participate*, then the outcome is $(C, \mathbf{d})$, where $\mathbf{d}$ is the initial allocation $(\frac{v(C)}{|C|})\langle 1, \ldots, 1\rangle$ adjusted for the donations made by C.

6. The preference ordering is defined such that $(C, \mathbf{d}) >_i (C', \mathbf{d}')$ iff $d_i > d'_i$.

Here are some examples of montonoic allocation games:

Example 2.1. *(Diamond in the Rough) Peter has a pickaxe, Larry has a lantern, and Chris has a cart. Using all three items, the group can recover a $300 diamond from a mine, but no two items are sufficient to navigate the mine. If the group pools its resources and goes mining, how should they divide the $300 of wealth?*

In this game, $v(C) = 0$ *whenever* $C \neq \{Peter, Larry, Chris\}$*, but* $v(\{Peter, Larry, Chris\}) = 300$.

Example 2.2. *(Two to Tango) A tango troupe consists of 10 men and 11 women. When the troupe puts on a show, the ticket revenue is $100 per male/female dancing pair. How should they divide up the revenue the show generates?*

In this game, $v(C) = 100 \cdot \min(\textit{men in } C,\ \textit{women in } C)$.

Example 2.3. *(Too Many Cooks) Ten chefs can choose whether or not to participate in making a meal. Each additional chef up to five will improve the meal by a fixed amount (worth $10), but each chef beyond the fifth will crowd the kitchen and add no additional value.*

In this game, $v(C) = \min(10|C|, 50)$.

Pareto Efficiency

We will briefly discuss the Pareto Frontier of the three allocation games discussed above.

Solution 1. (Diamond in the Rough) Any division of the money is Pareto efficient as long as all three men participate and go mining. This is because if not all men go mining, then the sole division $(0, 0, 0)$ of $0 is Pareto dominated by every possible division of $300.

Solution 2. (Two to Tango) Any division is Pareto efficient as long as all ten men and at least ten women participate in the tango show, generating $1000 (the maximum possible amount of revenue per show).

Solution 3. (Too Many Cooks) Any division is Pareto efficient as long at least five chefs enter the kitchen. This generates $50, the maximum possible amount of revenue per meal.

The Core

We will next give the cores of our sample allocation games.

Solution 4. (Diamond in the Rough) The core of this game is precisely the set of solutions in which the participating coalition is all of N. No smaller coalition can deviate, since it can't produce any money to divide up among its players.

Solution 5. (Two to Tango) Surprisingly, the only core allocation of this game is one that gives \$100 to all men and \$0 to all women.

1. First, we will show that this allocation is in the core. If some coalition attempts a deviation from this solution, it must pay at least \$100 to each of its men to incentivize their participation in the deviation. This precisely exhausts the money that the coalition can generate, so the coalition cannot pay its women more than the \$0 they would get in our proposed solution. Therefore, no deviation will be successful, so this allocation is in the core.

2. Second, we will show that no other allocation is in the core. Any other allocation must leave some (male, female) pair with a total of less than \$100; therefore, this two-person coalition can deviate, make \$100, and divide the money such that both see an increase in profits.

Intuitively, male dancing is a scarce resource in the Tango troupe while female dancing is oversupplied, so in accordance with standard economic principles the first will be priced and the second will be free.

Solution 6. (Too Many Cooks) This game has an empty core. At any solution, pick any five chefs who do not jointly receive \$50. These five chefs can form a coalition, deviate, make \$50, and improve each of their payoffs.

2.1.5 Negotiation Games

Negotiation games are another type of cooperative game. The definition of these games is original to this paper, although it is adapted from a similar model used by Elliott and Golub [5]. These games will be our central object of study in this paper.

Our model is as follows:

Definition 2.4. A *negotiation game* is a pair $(N, \mathbf{u})$ with the following properties:

1. The set $N = \{1, \ldots, n\}$ represents the set of players.

2. The element $\mathbf{u}$ is a *utility function* from $\mathbb{R}^n_{\geq \mathbf{0}}$ to $\mathbb{R}^n$.

3. Each player, as a pure strategy, chooses any real number (an *action level*) on the interval $[0, \infty)$.

4. The outcome of the game is an *action vector* $\mathbf{a}$, where a_i is the action level chosen by player i.

5. Each player i prefers action vector $\mathbf{a}$ to $\mathbf{a}'$ if and only if $u_i(\mathbf{a}) > u_i(\mathbf{a}')$.

Additionally, the utility function $\mathbf{u}$ in negotiation games is assumed to satisfy the following criteria:

1. (Well-Behaved) We assume that $\mathbf{u}$ is continuous and differentiable.

2. (Positive Externalities) For any $\lambda > \mathbf{0}$, we have $u_i(\mathbf{a} + \lambda \cdot \mathbf{e}^j) > u_i(\mathbf{a})$ whenever $i \neq j$.

3. (Bounded Externalities) Let $\mathbf{v} \in \mathbb{R}^n$ with $\|\mathbf{v}\|_\infty \leq 1$. There exists a value m such that $\sup_{\mathbf{a} \in \mathbb{R}^n_{\geq \mathbf{0}}, i \in N} (\mathbf{J}(\mathbf{a})\mathbf{v})_i < m$.

The model used by Elliott & Golub differs only in the assumptions placed on $\mathbf{u}$. They do not assume bounded externalities, and their assumption of positive externalities carries only a weak inequality (they say $u_i(\mathbf{a} + \lambda \cdot \mathbf{e}^j) \geq u_i(\mathbf{a})$). They do make the following additional assumptions:

1. (Costly Actions) For any action vector $\mathbf{a}$, we have $\mathbf{J}_{ii}(\mathbf{a}) < 0$ for any $i \in N$.

2. (Connectedness of Benefit Flows) Whenever C is a nonempty proper subset of N, there exist players $i \in C$ and $j \notin C$ such that $\mathbf{J}_{ij}(\mathbf{a}) > 0$.

3. (Bounded Improvements) The set $\{\mathbf{a} \in \mathbb{R}^n_{\geq \mathbf{0}} \,|\, \text{there is a scalar } s > 1 \text{ such that } s\mathbf{a} \text{ is a Pareto improvement on } \mathbf{a}\}$ is bounded.

None of these three assumptions will be adopted in our work.

Here are two types of strategic interaction that negotiation games are well-suited to model:

Example 2.4. *(Pollution Reduction) [5] A group of neighbouring countries each produce pollution, which reduces the air quality of the region. All countries would prefer that the other countries restrict their industrial production in order to reduce pollution, but there are pros (less pollution) and cons (less production) to restricting* *their own production. The countries convene to discuss a treaty in which each country agrees to restrict its industrial production in an effort to reduce its pollution output. By how much should each country be responsible for restricting its industrial production, taking into account (1) how much that country benefits from other countries' restrictions, and (2) how costly a restricted economy is to that country's development?*

Example 2.5. *(Housework) A group of housemates must routinely sweep the floors of their house. The housemates value cleanliness by varying amounts and dislike sweeping by varying amounts. How often should each be responsible for sweeping?*

2.1.6 The Relationship Between the Core and Pareto Efficiency

Just as we defined a Pareto frontier for the grand coalition (N), we can define a Pareto frontier for each sub-coalition.

Definition 2.5. If $(N, \mathbf{u})$ is a negotiation game and $C \subset N$, then we say an action vector $\mathbf{a}$ is *attainable by* C if $\mathbf{a}_{N-C} = \mathbf{0}$. If action vectors $\mathbf{d}, \mathbf{a}$ are both attainable by C, we say that $\mathbf{d}$ is a *Pareto improvement for* C on $\mathbf{a}$ if $\mathbf{u}_C(\mathbf{d}) \gneq \mathbf{u}_C(\mathbf{a})$. An action vector is *Pareto efficient for* C if it is attainable by C and no Pareto improvements for C on it exist. The set of action vectors that are Pareto efficient for C is called the *coalitional Pareto frontier* for C. An action vector $\mathbf{a}$ is *maximally Pareto efficient* for C if it is Pareto efficient for no action vector $\mathbf{a}'$ exists satisfying $\mathbf{a}' \gneq \mathbf{a}$ and $\mathbf{u}_C(\mathbf{a}') = \mathbf{u}_C(\mathbf{a})$.

In other words, the coalitional Pareto frontier for C is the set of optimal action vectors for C, given that no help will be given from $N - C$. This allows for another characterization of the core in terms of coalitional Pareto efficiency:

Proposition 2.6. *An action vector* $\mathbf{a}$ *is in the core of a negotiation game* $(N, \mathbf{u})$ *if and only if there does not exist a coalitionally Pareto efficient action vector* $\mathbf{d}$ *for a coalition* C *satisfying* $u_i(\mathbf{d}) \geq u_i(\mathbf{a})$ *for all* $i \in C$ *and* $u_i(\mathbf{d}) > u_i(\mathbf{a})$ *for some* $i \in C$.

Proof. Suppose that $\mathbf{d}$ is a deviation for a coalition C from an action vector $\mathbf{a}$ but $\mathbf{d}$ is not coalitionally Pareto efficient for C. Then there exists some $\mathbf{d}'$ that is a coalitionally Pareto efficient coalitional Pareto improvement on $\mathbf{d}$, and this $\mathbf{d}'$ must also be a deviation for C from $\mathbf{a}$. Therefore, it suffices to only check the coalitional Pareto frontiers for deviations from $\mathbf{a}$. □

This characterization should not be very surprising, but it is useful because it lends itself nicely to a graphical interpretation. Consider the negotiation game where

$N = \{1,2\}$ and $\mathbf{u}(a_1, a_2) = \langle 1 + a_2 - \frac{a_1^2}{2}, 2 + a_1 - \frac{a_2^2}{2}\rangle$. Player 1 can attain a utility of 1 on her own and player 2 can attain a utility of 2 on her own (because $\mathbf{u}(0,0) = \langle 1, 2\rangle$). The Pareto frontier in this game is the set of action vectors of the form $(a, \frac{1}{a})$, which correspond to the set of utility vectors described by the parametric curve $\langle 1 + \frac{1}{a} - \frac{a^2}{2}, 2 + a - \frac{1}{2a^2}\rangle$. Graphically, here are these three curves on one set of axes:

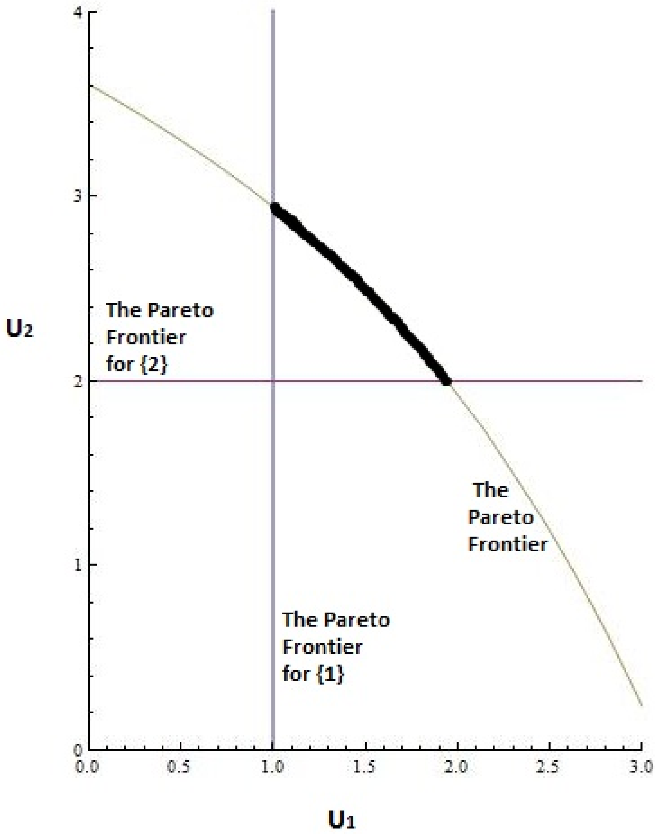

FIGURE 2.1: The Coalitional Pareto Frontiers of the Game

The bolded region - that is, the Pareto Frontier cut off at the coalitional Pareto frontiers - is precisely the set of utilities attainable under core action vectors.

2.2 The Computational Problem of Testing Core Membership in Negotiation Games

2.2.1 Bounded Rationality

Hopefully by now we have provided enough support for the core to convince an economist that it is a sensible solution concept to apply to cooperative game theory. But a computer scientist might still raise an objection.

We have defined the core on the premise that if a deviation exists, then our players will capitalize on it. But how can we be sure that, just because a deviation exists, our players are capable of finding it? Perhaps the problem of testing core membership lies outside P, and it would take our players a lifetime to decide whether or not a deviation exists from a given solution. The idea of treating economic players as Turing machines, rather than idealized computational oracles, is a common implementation of the economic idea of *bounded rationality* [7]. Ultimately, *we should only respect the core as a solution concept if there exists an algorithm that tests membership of an action vector in polynomial time.*

At a glance, things look perilous for the core. An action vector is in the core if no coalition has a deviation from it. Even if we can very quickly decide whether or not a given coalition has a deviation, a brute-force algorithm is impractical: there are $2^n - 1$ coalitions to consider. For games of size $n = 20$ or larger, it would take far more than a lifetime for a computer to brute-force check every coalition for a deviation.

As one might expect from this argument, our first result is negative: we show that the computational problem of core membership testing is NP-Hard. But first, we will give a more rigorous statement of the problem we aim to solve.

2.2.2 Epsilon-Indifference

It should be intuitively clear at this point that we are doomed to fail in our task of deciding whether or not a point is in the core of a negotiation game. Without knowing some very specific information about $\mathbf{u}$, it is impossible to search a continuum for a deviation - even if we can narrow our search down to a very small region, any behavior of our function is possible on that region. It is impossible to determine whether or not a point is in the core without knowing some additional properties of $\mathbf{u}$.

Our solution: we will not attempt to decide the problem of core membership exactly. Instead, we will attack an ϵ-approximate version of the problem. We define:

Definition 2.7. An action vector $\mathbf{d}$ is an ϵ-deviation from an action vector $\mathbf{a}$ for a coalition C if $\mathbf{u}_C(\mathbf{d}) \gneq \mathbf{u}_C(\mathbf{a}) + \epsilon$ (where $\epsilon = \langle \epsilon, \ldots, \epsilon \rangle$) and $\mathbf{d}_{N-C} = \mathbf{0}$.

Definition 2.8. An action vector $\mathbf{a}$ is in the ϵ-core if no coalition has an ϵ-deviation from C.

We shall loosen the restrictions on our algorithm: given a point $\mathbf{a}$, our algorithm must *either* find a deviation from $\mathbf{a}$, *or* report that there are no ϵ-deviations from $\mathbf{a}$ (so $\mathbf{a}$ is in the ϵ-core). This means that, in the narrow class of games that carry deviations but not ϵ-deviations from $\mathbf{a}$, our algorithm is allowed to either report a deviation or report $\mathbf{a}$ is in the ϵ-core, and either option is considered correct.

We mitigate this inaccuracy by ensuring that the runtime of our algorithm is cheap (polynomial) in the value of ϵ, so very precise results are still computationally feasible.

2.2.3 Stating the Problem

We are now ready to write the following problem:

Problem. (CORE-MEMBERSHIP)

Input:

1. An integer n
2. A number $\epsilon > 0$
3. An integer m
4. An action vector $\mathbf{a}$
5. A utility function $\mathbf{u}$ that satisfies all required assumptions of a utility function in a negotiation game

Output:

1. Either report that $\mathbf{a}$ is not in the core of the game, or
2. report that $\mathbf{a}$ is in the ϵ-core of the game.

2.2.4 CORE-MEMBERSHIP is NP-Hard

Unfortunately, the problem as stated is not tractable.

Theorem 2.9. *CORE-MEMBERSHIP is NP-Hard.*

Proof. The proof is by reduction to SAT. Consider an instance of SAT with variables $x_1, \ldots, x_k$. Set up a 1-player negotiation game with $m = 1$, $n = \epsilon = 1, a = 2^k + 1$. Define the utility function u as follows:

1. Set $u(0) = 0$.
2. For input $z \geq a$, set $u(z) = 0$.

3. For input $z < a$, note that the integer $z-1$ can always be expressed in at most k bits. Consider the k-digit representation of $z-1$, and set each x_j in the SAT instance according to the j^{th} bit of this representation (for example, $k=1$ corresponds to the bit representation 0^k, which corresponds to the setting $x_1 = \cdots = x_k = \text{FALSE}$). Check to see if this variable setting satisfies the boolean expression. If it does, then return $u(z) = 2$. If not, then check to see if the variable setting corresponding to z or $z-2$ corresponds to a satisfying variable setting. It either one does, then return $u(z) = 1$. If not, then return $u(z) = 0$.

If we run CORE-MEMBERSHIP on this negotiation game, it will either (1) output a value z such that z, $z-1$, or $z-2$ corresponds to a variable setting that satisfies the boolean formula, or (2) it will report that no such z exists (and therefore, the formula is unsatisfiable). □

We conclude that the core is a poor solution concept for negotiation games at this level of generality, because it is computationally infeasible to test core membership for a game of reasonable size. The natural next question is: what additional assumptions can we place on negotiation games in order to salvage the core? An ideal assumption would be simple, economically justifiable, and push the computational problem down into P.

2.2.5 The Concavity Assumption

We will find a perfect assumption in *concavity.*

Definition 2.10. (Concavity) A negotiation game $(N, \mathbf{u})$ is *concave* if $\mathbf{u}$ is concave; that is, $\lambda\mathbf{u}(\mathbf{v}) + (1-\lambda)\mathbf{u}(\mathbf{v}') \leq \mathbf{u}(\lambda\mathbf{v} + (1-\lambda)\mathbf{v}')$ for any action vectors $\mathbf{v}, \mathbf{v}'$ and any $\lambda \in [0, 1]$.

Accordingly, we will add the following assumption to our model:

Assumption. In any negotiation game $(N, \mathbf{u})$, $\mathbf{u}$ is concave.

It is very common for economists to assume concave utility functions. Concavity is the mathematical expression of the economic principle of *diminishing marginal returns. Diminishing marginal returns* is the idea that the value of 1 additional unit of a resource falls as more of that resource is acquired, because the first few units of the resource have already been put to their best available use. Here are some examples:

1. Money has diminishing marginal returns. The first part of a person's paycheck is spent on food, clothing, and shelter, which provide a large boost to quality of life. Once these have been acquired, the next part of a person's paycheck will be spent on luxury items, which provide a comparatively smaller boost in utility.
2. Food has diminishing marginal returns. The first slice of pizza is eaten when the subject is hungriest and wants pizza the most. The subject is then less hungry for the second slice, and even less so for the third, and so on until the hundredth slice will likely be thrown out.
3. Housework has diminishing marginal returns. If an individual plans to spend five minutes of time sweeping, he will spend that time cleaning the dirtiest part of the house. Once that area is clean, the next five minutes of sweeping must be spent on the second dirtiest part of the house, which gets rid of less dirt than the first five minutes did.

We assume that the work performed by each player has diminishing marginal returns to each other player. This straightforwardly implies our assumption of total group concavity.

Hopefully, the simplicity and economic justification of the concavity assumption are clear. We will now spend the rest of this paper showing that this assumption is sufficient to push the problem down into P.

Chapter 3

Economic Analysis of Negotiation Games

Before we begin our algorithm, we will develop a considerable amount of economic theory behind negotiation games.

3.1 Sperner's Lemma

We will first quote and prove *Sperner's Lemma* [9], an important and well-known result in combinatorics.

3.1.1 Sperner's Lemma

We begin with a fairly unsurprising result:

Proposition 3.1. *Let S be a line segment. Suppose we partition S into smaller line segments, and we color each endpoint of the new line segments one of two colors. If the endpoints of S are differently colored, then there are an odd number of line segments in the partition with differently-colored endpoints.*

Proof. Travel the line segment from left to right. Each time you switch colors, it is because you have encountered a two-colored line segment in the partition of S. If you switch colors an odd number of times, then the endpoints of S are different colors; if you switch colors an even number of times, then the endpoints of S are the same color. □

We next generalize this result to n dimensions.

Definition 3.2. A *simplex* is the set of affine linear combinations of $n+1$ points in $\mathbb{R}^n$. For example, a one-dimensional simplex is a line segment, a two-dimensional simplex is a triangle (including its interior), a three-dimensional simplex is a tetrahedron, etc.

Definition 3.3. A *simplicization* of a convex polyhedron is a partition of that polyhedron into many smaller simplexes.

Definition 3.4. Let S be an n-dimensional simplex. Suppose we simplicize S and we color each node of this simplicization one of $n+1$ colors. We say that S is *Sperner colored* if there does not exist a vertex v of S and a node n on the opposite face of S that share a color.

Note that a Sperner coloring implies that each vertex v of S is differently colored, since each vertex $v' \neq v$ of S lies on a face opposite v.

We are now ready to quote Sperner's Lemma:

Lemma 3.5. *(Sperner's Lemma) Let S be an n-dimensional simplex. Suppose we simplicize S, and we color each vertex of the new simplexes one of $n+1$ colors $0, 1, \ldots, n$. If S is Sperner colored, then an odd number of the simplexes into which S has been partitioned have every color represented on their vertices.*

Proof. The proof is by induction. The one-dimensional base case is shown in Proposition 3.1. The inductive step is as follows:

1. Find the vertex of S containing color 0, and consider the opposite face of S. This face must be an $n-1$-dimensional Sperner-colored simplex. By our inductive assumption, we may assume that there are an odd number of $n-1$-dimensional simplexes on this face with vertices colored $1, \ldots, n$. Additionally, by our coloring rules, there can be no other $n-1$-dimensional simplexes with these colors on any other face of S. Let M_0 be the set of these simplexes; then $|M_0|$ is odd.

2. Pick any element $m \in M_0$. Draw an arrow into S through m. We are now inside one a simplex T of the partition.

3. If the simplex T is fully-colored, then stop and repeat step (3) on the next element of M_0. Otherwise, draw an arrow from T through the face of T with colors $1, \ldots, n$ (aside from the one from which we entered T). We are now in a new simplex U.

4. Repeat the process with U. Eventually, the path we create will either dead-end in some simplex in S (corresponding to an $n+1$-colored simpelx), or it will exit S through another element $m' \in M_0$.

5. If the path exits through m', then delete m and m' from M_0 and start again. Otherwise, keep m in M_0. Suppose we delete j pairs of elements from $M_{1,2}$ in total.

6. Suppose there exists an additional $n+1$-colored simplex in S not reached by any element of M_0. Draw an arrow from this triangle through the face labeled $1, \ldots, n$, then continue drawing these arrows until we reach yet another $n+1$-colored simplex in S. This argument shows that additional three-colored triangles come in pairs. Suppose there are k additional pairs of three-colored triangles.

7. The total number of $n+1$-colored simplexes in S is equal to $|M_0| - 2j + 2k$. Since $j, k \in \mathbb{Z}$, this implies that the number of $n+1$-colored simplexes in S is equal to $|M_0| \pmod 2$.

□

Here is a graphical representation of the inductive step in two dimensions:

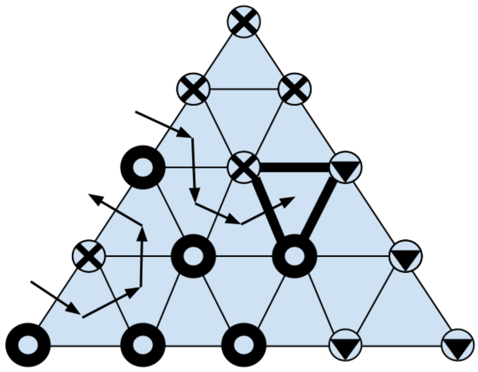

FIGURE 3.1: Proof of Sperner's Lemma. The three colors are represented by O's, X's, and $\triangledown$'s. This picture contains arrows corresponding to the set $M_{\triangledown}$.

3.1.2 Continuous Analog of Sperner

Sperner's Lemma is a discrete combinatorial result, it has a continuous analog as well.

Corollary 3.6. *Let S be an n-dimensional simplex. Suppose we color each point in S one of $n+1$ colors. If this is a Sperner coloring of S, then there exists a point $\mathbf{s} \in S$ for which every neighbourhood of $\mathbf{s}$ contains a point of each color.*

Proof. Let the index set $\{0,\ldots,n\}$ represent the colors of the simplex. Suppose for the sake of contradiction that the claim is *not* true. Let
$\epsilon = \inf\{\epsilon > 0 \,|\, \text{There exists an } \epsilon\text{-disk } D(\mathbf{s},\epsilon) \text{ containing a point of each color.}\}$. Either $\epsilon = 0$ or $\epsilon > 0$.

1. If $\epsilon > 0$, then let R be a simplicization of S such that each sub-simplex has side lengths no greater than ϵ. Then none of these sub-simplices can be distinctly colored. This contradicts Sperner's Lemma, so it must be the case that $\epsilon = 0$.

2. If $\epsilon = 0$, this implies that for any $\delta > 0$, the set
$H_\delta = \{\mathbf{s} \in S \,|\, D(\mathbf{s},\delta) \text{ contains a point of each color}\}$ is nonempty. Since this set is compact, we know from Cantor's Intersection Theorem[1] that $\cap_{\delta>0} H_\delta$ is nonempty. Any element of this set satisfies our premise.

□

3.2 Classifying Action Vectors by their Pareto Improvements

3.2.1 Directional Pareto Improvements

Our treatment of Sperner's Lemma was one big lead-up to the following result:

Proposition 3.7. *For every non-Pareto efficient action vector* $\mathbf{a}$*, one of the following two claims holds:*

1. *There exists a direction* $\mathbf{d} > \mathbf{0}$ *such that* $\mathbf{J}(\mathbf{a})\mathbf{d} \geq \mathbf{0}$.

2. *There exists a direction* $\mathbf{d} < \mathbf{0}$ *such that* $\mathbf{J}(\mathbf{a})\mathbf{d} \leq \mathbf{0}$.

Proof. Suppose there exists a player i for whom $(\mathbf{J}(\mathbf{a})\mathbf{e}^i)_i \geq 0$ (i.e. player i weakly gains utility from unilaterally increasing his action). A marginal increase in player i's action will grant each other player a utility gain. Then, each other player can increase their action by a small amount (but not enough to offset their utility gain), leaving player i with a utility gain as well. This produces a direction $\mathbf{d} > \mathbf{0}$ with $\mathbf{J}(\mathbf{a})\mathbf{d} > \mathbf{0}$.

Alternately, suppose that $(\mathbf{J}(\mathbf{a})\mathbf{e}^i) < 0$ for all players i. Let S be the $n-1$-dimensional simplex with vertices at $\mathbf{e}^1,\ldots,\mathbf{e}^n$. Color each point $\mathbf{s} \in S$ by assigning a color to each player, choosing a player i for whom $(\mathbf{J}(\mathbf{a})\mathbf{s})_i < 0$, and color $\mathbf{s}$

[1] The intersection of a descending sequence of compact sets is nonempty.

according to player i. Note that every vertex $\mathbf{e}^i$ can be colored according to player i, while the opposing side (which corresponds to other players increasing their action while i keeps his constant) will certainly not be.

Perhaps this coloring scheme is not well-defined; perhaps there exists $\mathbf{s} \in S$ to which no color would be assigned. That implies that $\mathbf{J}(\mathbf{a})\mathbf{s} \geq \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$ (since the boundary of S will certainly be colored).

Suppose this coloring scheme is well-defined; then we know that it is Sperner. By our continuous analog of Sperner's Lemma, there exists some point $\mathbf{s} \in S$ such that every neighbourhood of $\mathbf{s}$ contains a point of each color. This point must satisfy $\mathbf{J}(\mathbf{a})\mathbf{s} \leq \mathbf{0}$. □

We then base the following definitions on our claim:

Definition 3.8.

An action vector $\mathbf{a}$ is *below the Pareto frontier* if there exists a direction $\mathbf{d} > \mathbf{0}$ such that $\mathbf{J}(\mathbf{a})\mathbf{d} \geq \mathbf{0}$.

An action vector $\mathbf{a}$ is *above the Pareto frontier* if there exists a direction $\mathbf{d} < \mathbf{0}$ such that $\mathbf{J}(\mathbf{a})\mathbf{d} \leq \mathbf{0}$.

3.2.2 Uniqueness of Classification

The terminology *above* and *below* the Pareto frontier carries a lot of geometric intuition. We will next attempt to justify this intuition.

Proposition 3.9. *An action vector* $\mathbf{a}$ *on the Pareto frontier is both above and below the Pareto frontier.*

Proof. Suppose $\mathbf{a}$ is Pareto efficient. Then $\mathbf{J}(\mathbf{a})$ must be singular; otherwise, $\mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle$ corresponds to a direction that creates a Pareto improvement from $\mathbf{a}$. Let $\mathbf{x}$ be a nonzero point in the nullspace of $\mathbf{J}(\mathbf{a})$. Let C be the coalition of players i for whom $x_i < 0$. Consider the direction $\max(\mathbf{x}, \mathbf{0})$. All players in C must gain utility from movement in this direction (since $N - C$ increases its action), and all players in $N - C$ must gain utility from movement in this direction (since C increases its action from $\mathbf{x}$). Therefore, this signifies a Pareto improvement from $\mathbf{a}$.

This contradiction is resolved only if $C = N$ or $C = \emptyset$. If $C = N$ then $\mathbf{x} < \mathbf{0}$. If $C = \emptyset$ then $\mathbf{x} \gneq \mathbf{0}$. Additionally, it cannot be the case that $x_i = 0$ for any player i, since this player would gain utility in the direction $\mathbf{x}$. Therefore, the alternative is that $\mathbf{x} > \mathbf{0}$.

This means that the directions $\mathbf{x}$ and $-\mathbf{x}$ imply that $\mathbf{a}$ is both above and below the Pareto frontier. □

Proposition 3.10. *If* $\mathbf{a}$ *is both above and below the Pareto frontier, then it is on the Pareto frontier.*

Proof. Suppose otherwise, and let $\mathbf{p}$ be a Pareto improvement on $\mathbf{a}$.

1. Suppose $\mathbf{p} \lneq \mathbf{a}$. Then $\mathbf{J}(\mathbf{a})(\mathbf{p}-\mathbf{a}) \gneq \mathbf{0}$, so $\mathbf{J}(\mathbf{a})(\mathbf{a}-\mathbf{p}) \lneq \mathbf{0}$. Let $\mathbf{d} > \mathbf{0}$ be the direction that implies that $\mathbf{a}$ is above the Pareto frontier. Define $\lambda_i = \frac{-d_i}{(a_i - p_i)}$ (or $\lambda_i = \infty$ if $a_i = p_i$). Scale the vector $\mathbf{a}-\mathbf{p}$ by $\min_{i \in N} \lambda_i$; then we reach a vector $\mathbf{v}$ for which $v_i = -d_i$ for some player i, while $v_{N-\{i\}} \lneq -d_{N-\{i\}}$. Therefore, player i must prefer the direction $\mathbf{v}$ to $-\mathbf{d}$. But he doesn't: $-(\mathbf{J}(\mathbf{a})\mathbf{d})_i > 0$, but $(\mathbf{J}(\mathbf{a})\lambda_i(\mathbf{a}-\mathbf{p}))_i \leq 0$. This is a contradiction.
2. Suppose that $p_i > a_i$ for some player i. Let $\mathbf{p}' = \max(\mathbf{p}, \mathbf{0})$; then $\mathbf{p}'$ must also be a Pareto improvement on $\mathbf{a}$, and $\mathbf{p}' \gneq \mathbf{0}$. Let $\mathbf{d} < \mathbf{0}$ be the direction that implies that $\mathbf{a}$ is below the Pareto frontier, and the proof continues exactly as in the case above.

□

This proves that an action vector is both above and below the Pareto frontier if and only if it is on the Pareto frontier, so our terminology is sensible.

3.2.3 Determining the Classification

Our final task is to create a strategy to quickly classify a given action vector $\mathbf{a}$ as *above the Pareto frontier* or *below the Pareto frontier.*

Proposition 3.11. *If* $\mathbf{J}(\mathbf{a})$ *is nonsingular and* $\mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle \lneq \mathbf{0}$*, then* $\mathbf{a}$ *is above the Pareto frontier.*

Proof. Let $\mathbf{d} = \mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle$. Pick a player i for whom $d_i = 0$. Decrease player i's action slightly from $\mathbf{d}$ - more than 0, but not enough for any other player to lose more than half their current rate of utility gain (1). Then pick another player j for whom $d_j = 0$, and repeat the process until we have a new direction $\mathbf{d}' < 0$ with $\mathbf{J}(\mathbf{a})\mathbf{d}' > 0$. □

Proposition 3.12. *If* $\mathbf{J}(\mathbf{a})$ *is nonsingular and* $(\mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle)_i > 0$ *for some player* i*, then* $\mathbf{a}$ *is below the Pareto frontier.*

Proof. Let $\mathbf{d} = \mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle$, and let $\mathbf{d}' = \max(\mathbf{d}, \mathbf{0})$. Each player j for whom $d_j > 0$ must prefer $\mathbf{d}'$ to $\mathbf{d}$, since $\mathbf{d}'$ reflects an increase in the action of $N - \{j\}$ from $\mathbf{d}$. Each player k for whom $d_k \leq 0$ must gain utility at the margin from movement in the direction $\mathbf{d}'$, since $d'_k = 0$ but $\mathbf{d}' \geq \mathbf{0}$. So $\mathbf{d}' \geq \mathbf{0}$ with $\mathbf{J}(\mathbf{a})\mathbf{d}' > \mathbf{0}$. We can then increase the action of each player k for whom $d'_k = 0$ by a small amount, resulting in a new direction $\mathbf{d}'' > \mathbf{0}$ with $\mathbf{J}(\mathbf{a})\mathbf{d}'' > \mathbf{0}$. □

Proposition 3.13. *If $\mathbf{J}(\mathbf{a})$ is singular, then $\mathbf{a}$ is below the Pareto frontier.*[2]

Proof. Let $\mathbf{x} \neq \mathbf{0}$ be a vector in the nullspace of $\mathbf{J}(\mathbf{a})$. If $\mathbf{x} \gneq \mathbf{0}$, then it must be the case that $\mathbf{x} > \mathbf{0}$ (because $x_i = 0$ would imply $(\mathbf{J}(\mathbf{a})\mathbf{x})_i > 0$), which implies that $\mathbf{a}$ is both above and below the Pareto frontier. If $\mathbf{x} \lneq \mathbf{0}$, then it must be the case that $\mathbf{x} < \mathbf{0}$, which again implies that $\mathbf{a}$ is both above and below the Pareto frontier. Finally, suppose that $x_i < 0$ and $x_j > 0$ for two players i, j. Let $\mathbf{x}' = \max(\mathbf{0}, \mathbf{x})$. We know that player j must gain utility from movement in the direction $\mathbf{x}'$, since it reflects an increase of $N - \{j\}$'s action from $\mathbf{x}$. We know that player i must gain utility from movement in the direction $\mathbf{x}'$, since $\mathbf{x}' \gneq \mathbf{0}$ with $x'_i = 0$. Therefore, $\mathbf{x}' \geq \mathbf{0}$ and $\mathbf{J}(\mathbf{a})\mathbf{x}' > \mathbf{0}$. We can therefore once again perturb $\mathbf{x}'$ slightly to $\mathbf{x}''$ for which $\mathbf{x}'' > \mathbf{0}$ and $\mathbf{J}(\mathbf{a})\mathbf{x}'' > \mathbf{0}$. □

We will conclude this section by summarizing our classification of action vectors.

Lemma 3.14. *If $\mathbf{a}$ is nonsingular and $\mathbf{J}^{-1}(\mathbf{a})\langle 1, \ldots, 1\rangle \lneq \mathbf{0}$, then $\mathbf{a}$ is above the Pareto frontier. If either of these conditions fail to hold, then $\mathbf{a}$ is below the Pareto frontier.*

3.3 Main Lemma: A Topological Characterization of the Core

We next develop a major structural result about core membership. We re-characterize the core from an economic condition to a topological condition that is much easier to detect.

3.3.1 A Topological Characterization of the Core

Our new characterization of the core is as follows:

Lemma 3.15. *Suppose the following two conditions hold:*

[2]Note that this includes the possibility that $\mathbf{a}$ is on the Pareto frontier.

1. $\mathbf{a}$ *is above the Pareto frontier*

2. $\mathbf{a}$ *and* $\mathbf{0}$ *are path-connected in the set* $\{\mathbf{y} \in \mathbb{R}^n_{\geq \mathbf{0}} \,|\, \mathbf{u}(\mathbf{y}) \leq \mathbf{u}(\mathbf{a}) + \epsilon\}$

Then $\mathbf{a}$ *is in the* ϵ*-core. Additionally, if either of these conditions fails to hold, then* $\mathbf{a}$ *is not in the core.*

We will first prove that our two conditions imply that $\mathbf{a}$ is in the ϵ-core.

Proposition 3.16. *Let* $\mathbf{d}$ *be an* ϵ*-deviation for* C *from* $\mathbf{a}$ *for some* $\epsilon > 0$*. Assume that* $\mathbf{d} \leq \mathbf{a}$*. If there exists a path* P *from* $\mathbf{a}$ *to another action vector* $\mathbf{v}$ *with* $\mathbf{u}(\mathbf{p}) \leq \mathbf{u}(\mathbf{a}) + \epsilon$ *for all* $\mathbf{p} \in P$*, then* $\mathbf{d} < \mathbf{v}$*.*

Proof. Traverse P from $\mathbf{a}$ until we hit the first point $\mathbf{p} \in P$ such that $p_i = d_i$ for some player i. Then $\mathbf{d} \lneq \mathbf{p}$ with $d_i = p_i$, so by our assumption of positive externalities, we know that $u_i(\mathbf{d}) < u_i(\mathbf{p})$. But $u_i(\mathbf{p}) \leq u_i(\mathbf{a}) + \epsilon$ and $u_i(\mathbf{d}) \geq u_i(\mathbf{a}) + \epsilon$, so this is impossible.

Since this is a contradiction, there must not be any such point $\mathbf{p} \in P$. This implies that $\mathbf{d} < \mathbf{v}$. □

Proposition 3.17. *Let* $\mathbf{a}$ *be a Pareto efficient point. Then every deviation* $\mathbf{d}$ *from* $\mathbf{a}$ *satisfies* $\mathbf{d} \leq \mathbf{a}$*.*

Proof. Since $\mathbf{a}$ is above the Pareto frontier, there exists a direction $\mathbf{x} > 0$ such that $\mathbf{J}(\mathbf{a})\mathbf{x} < \mathbf{0}$. Travel from $\mathbf{a}$ in the direction $\mathbf{x}$ until you reach a point $\mathbf{d}' \geq \mathbf{d}$. Since $\mathbf{d} \leq \mathbf{d}'$ and there exists a path P from $\mathbf{d}'$ to $\mathbf{a}$ with $\mathbf{u}(\mathbf{p}) \leq \mathbf{u}(\mathbf{a})$ for each $\mathbf{p} \in P$, this implies that $\mathbf{d} \leq \mathbf{a}$. □

These results show that our two conditions imply membership of $\mathbf{a}$ in the ϵ-core. We will now show that a *lack* of either of these two results imply that $\mathbf{a}$ is not in the core.

If $\mathbf{a}$ is not Pareto efficient, it is trivially true that $\mathbf{a}$ is not in the core (a Pareto improvement is a deviation for the coalition N). All that remains is to show that if $\mathbf{a}$ and $\mathbf{0}$ are not path-connected in the set $\{\mathbf{y} \in \mathbb{R}^n_{\geq \mathbf{0}} \,|\, \mathbf{u}(\mathbf{y}) < \mathbf{u}(\mathbf{a}) + \epsilon\}$, then $\mathbf{a}$ is not in the core. In fact, we will show a slightly stronger result.

Proposition 3.18. *If the set* $V = \{\mathbf{y} \in \mathbb{R}^n_{\geq \mathbf{0}} \,|\, \mathbf{u}(\mathbf{y}) \leq \mathbf{u}(\mathbf{a}) + \epsilon\}$ *has more than one connected component, then* $\mathbf{a}$ *is not in the core.*

Proof. Let T be a connected component of V that does not contain $\mathbf{0}$. Since T is closed, we can choose a point $\mathbf{t} \in T$ on the lower envelope of T (i.e. there is no $t' \in T$

with $t' \lneq t$). Let C be the set of players whose action is 0 at t. Then every player i in the coalition $N - C$ must satisfy $u_i(\mathbf{t}) = u_i(\mathbf{a}) + \epsilon$; otherwise, we could reduce their action by a small amount and remain in T. That implies that $\mathbf{t}$ is a deviation for $N - C$ from $\mathbf{a}$. $\square$

This completes the proof of our main lemma. We will create the following terminology based on the lemma:

Definition 3.19. An action vector $\mathbf{v}$ is ϵ-*accessible* from $\mathbf{a}$ if it is path-connected to some point $\mathbf{a}' \geq \mathbf{a}$ in the set $\{\mathbf{v} \,|\, \mathbf{u}(\mathbf{v}) \leq \mathbf{u}(\mathbf{a}) + \epsilon\}$.

Definition 3.20. An action vector $\mathbf{v}$ is ϵ-*comparable* to $\mathbf{a}$ if $\mathbf{u}(\mathbf{v}) \leq \mathbf{u}(\mathbf{a}) + \epsilon$ or if $\mathbf{u}(\mathbf{v}) \geq \mathbf{u}(\mathbf{a})$.

When ϵ is clear from context, we will simply use "accessible" and "comparable."

3.3.2 Relationship to Point Classification

Our main lemma has a very important corollary:

Corollary 3.21. *If* $\mathbf{v}$ *is above the Pareto frontier and comparable to* $\mathbf{a}$*, then either* $\mathbf{v}$ *is a Pareto improvement on* $\mathbf{a}$ *or* $\mathbf{v}$ *is accessible.*

Proof. Assume that $\mathbf{v}$ is not a Pareto improvement on $\mathbf{a}$; then $\mathbf{u}(\mathbf{v}) \leq \mathbf{u}(\mathbf{a})$. Since $\mathbf{v}$ is above the Pareto frontier, we can choose the direction $\mathbf{d} < 0$ for which $\mathbf{J}(\mathbf{v})\mathbf{d} > 0$. Travel from $\mathbf{v}$ in the direction $-\mathbf{d}$ until you reach a point $\mathbf{a}' > \mathbf{a}$. By concavity, every point $\mathbf{p}$ on this path satisfies $\mathbf{u}(\mathbf{p}) \leq \mathbf{u}(\mathbf{a}) + \epsilon$. $\square$

It also makes accessible points beneath the Pareto frontier very powerful.

Corollary 3.22. *Suppose* $\mathbf{v}$ *is accessible and beneath the Pareto frontier. Let* $\mathbf{d} > 0$ *with* $\mathbf{J}(\mathbf{v})\mathbf{d} > 0$*. Then the player* i *who minimizes* $\frac{v_i}{d_i}$ *does not participate in any deviation from* $\mathbf{a}$*.*

Proof. Draw a linear path from $\mathbf{v}$ in the direction $-\mathbf{d}$. By concavity, every point $\mathbf{p}$ in this path satisfies $\mathbf{u}(\mathbf{p}) < \mathbf{u}(\mathbf{a}) + \epsilon$. This path must terminate when one player's action reaches 0; this is the player that minimizes $\frac{v_i}{d_i}$. By our main lemma, we then conclude that this player plays action 0 at any deviation from $\mathbf{a}$, so we may assume that he does not participate in any deviation from $\mathbf{a}$. $\square$

Chapter 4

The Algorithm

We are now finally ready to begin the discussion of our main algorithm.

4.1 Main Algorithm

4.1.1 Overview of the Main Algorithm

Our main algorithm is very straightforward and does not depend on any of the economic theory we have built (rather, the main algorithm will rely on a subroutine that requires our economic theory for its implementation). We will start with the grand coalition N, then run a subroutine named *Delete-Or-Deviate* that either (1) finds a deviation for N from $\mathbf{a}$, or (2) proves that some player $i \in N$ does not participate in any ϵ-deviation from $\mathbf{a}$, then deletes i from N. We then re-run our subroutine on the reduced coalition, and continue until we find a deviation for some coalition from $\mathbf{a}$ (in which case we report that $\mathbf{a}$ is not in the core), or we delete every player from the coalition (in which case we report that $\mathbf{a}$ is in the ϵ-core).

4.1.2 Main Algorithm Code

We implement with the following code:

---

if $\mathbf{a}$ is not above the Pareto frontier **then**

  Output "$\mathbf{a}$ is not in the core."

end if

$C \leftarrow \{1, \ldots, n\}$

while $C \neq \emptyset$ **do**

Delete-Or-Deviate($C, \mathbf{a}, \mathbf{u}, \epsilon, m$)

end while

Output "**a** is in the ϵ-core."

It is a matter of simple matrix math to check whether or not **a** is above the Pareto frontier (see Lemma 3.14). Once we do so, we have narrowed our search down to the box $\{\mathbf{v} \mid \mathbf{0} \leq \mathbf{v} \leq \mathbf{a}\}$; this is an important preprocessing step for our *Delete-Or-Deviate* subroutine.

We will now discuss the far more complicated *Delete-Or-Deviate* subroutine.

4.2 Subroutine: Delete-Or-Deviate

4.2.1 Overview of Delete-Or-Deviate

In our main algorithm, we established that **a** is above the Pareto frontier, so only the box $B = \{\mathbf{v} \mid \mathbf{0} \leq \mathbf{v} \leq \mathbf{a}\}$ must be considered. We make a variable **t** to hold the top corner of the box (initially **a**), and another variable **b** to hold the bottom corner of the box (initially **0**).

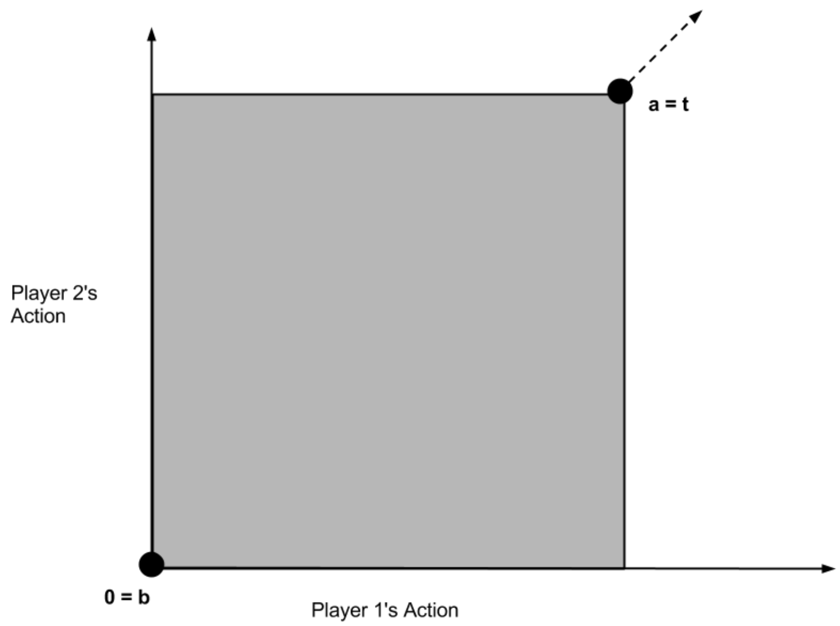

FIGURE 4.1: The initial configuration of Delete-Or-Deviate

The dashed line extending from $\mathbf{a}$ is due to the fact that $\mathbf{a}$ is above the Pareto frontier, and it is what allows us to restrict the action space to only those action vectors in the box. Our algorithm will proceed by repeatedly shrinking the box until $\mathbf{b}$ and $\mathbf{t}$ are very close together, but as an invariant, $\mathbf{t}$ will always remain above the Pareto frontier and $\mathbf{b}$ will always remain below the Pareto frontier.

The box-shrinking process is the crux of the algorithm, and it proceeds as follows. We first pick our favorite player $i \in C$ and declare that he will drive our algorithm. We fix his action at the level $\frac{b_i+t_i}{2}$.

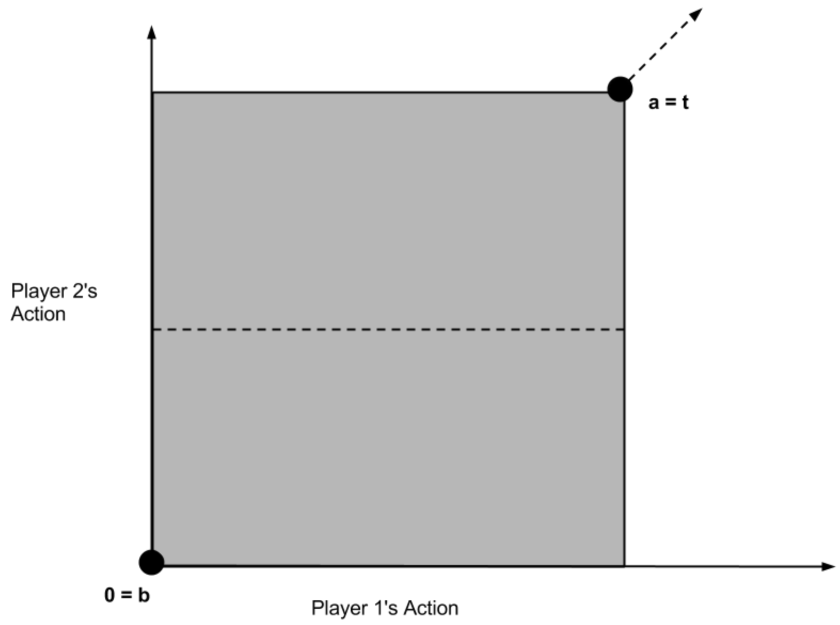

FIGURE 4.2: Player 2's action is temporarily fixed at the dashed line.

We then attempt to find the point $\mathbf{v}$ that maximizes the other players' actions, subject to the constraint $v_i = \frac{b_i+t_i}{2}$ (i.e. staying on the dashed line), and the additional constraint $\mathbf{u}_{C\setminus\{i\}}(\mathbf{v}) \geq \mathbf{u}_{C\setminus\{i\}}(\mathbf{a})$. Maybe we succeed:

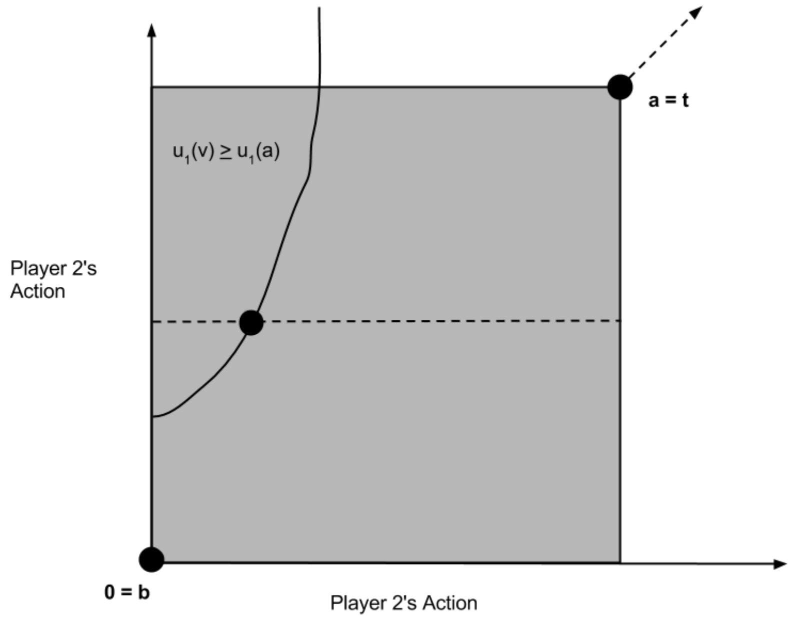

FIGURE 4.3: The dashed line intersects the region $\{\mathbf{v} \mid u_1(\mathbf{v}) \geq u_1(\mathbf{a})\}$.

Or maybe there does not exist a point that satisfies the given constraints:

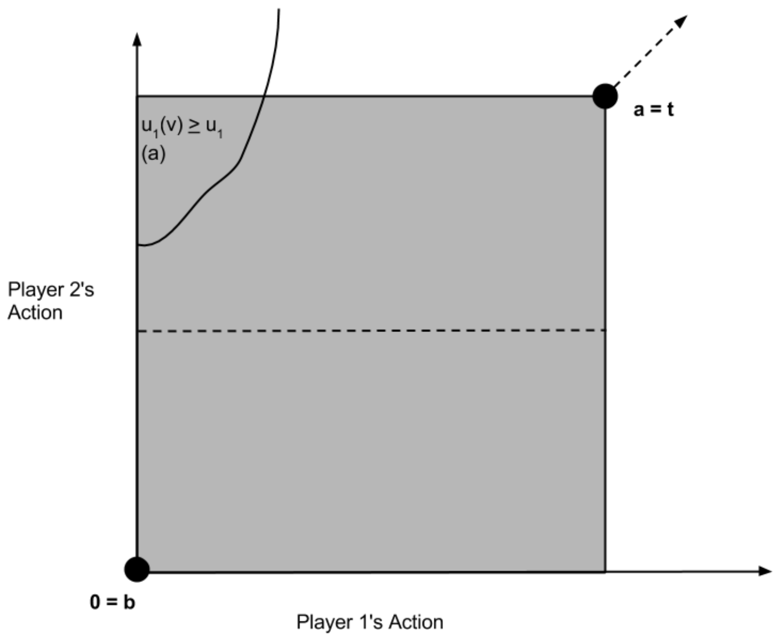

FIGURE 4.4: The dashed line does not intersect the region $\{\mathbf{v} \mid u_1(\mathbf{v}) \geq u_1(\mathbf{a})\}$.

If it's the latter case (no intersection), then we bring the bottom half of the box up to the dashed line.

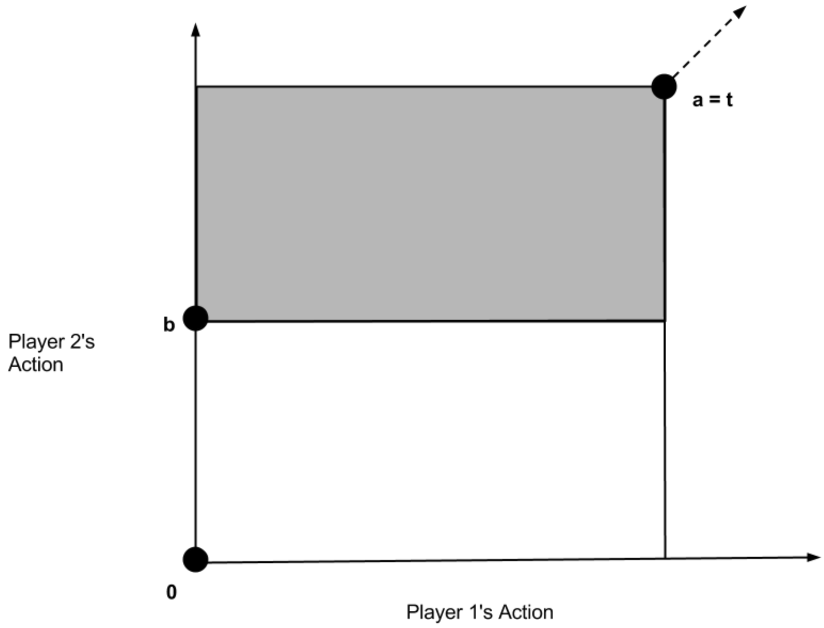

FIGURE 4.5: We have moved the lower bound of the box up to $v_i = \frac{b_i + t_i}{2}$.

Then we repeat the process. If instead our search succeeded (the dashed line does intersect the region $\{\mathbf{v} \mid u_1(\mathbf{v}) \geq u_1(\mathbf{a})\}$), then we check our point $\mathbf{v}$ and decide whether or not it's a deviation from $\mathbf{a}$; if it is, then we report this fact and exit. Otherwise, we check to see whether $\mathbf{v}$ is above or below the Pareto frontier. If it's above, then we move the top corner of the box down to $\mathbf{v}$.

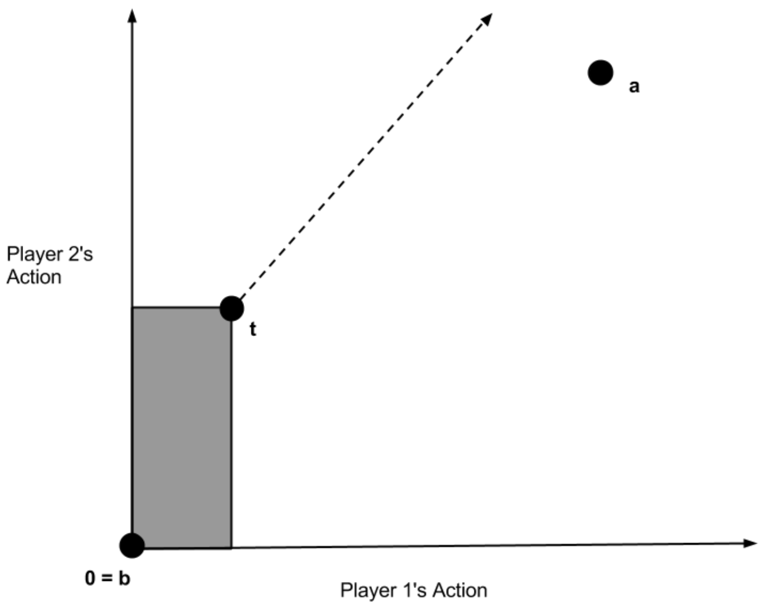

FIGURE 4.6: We have moved the top corner of the box down to **v**.

If **v** is below the Pareto frontier, then we move the bottom corner of the box up to **v**.

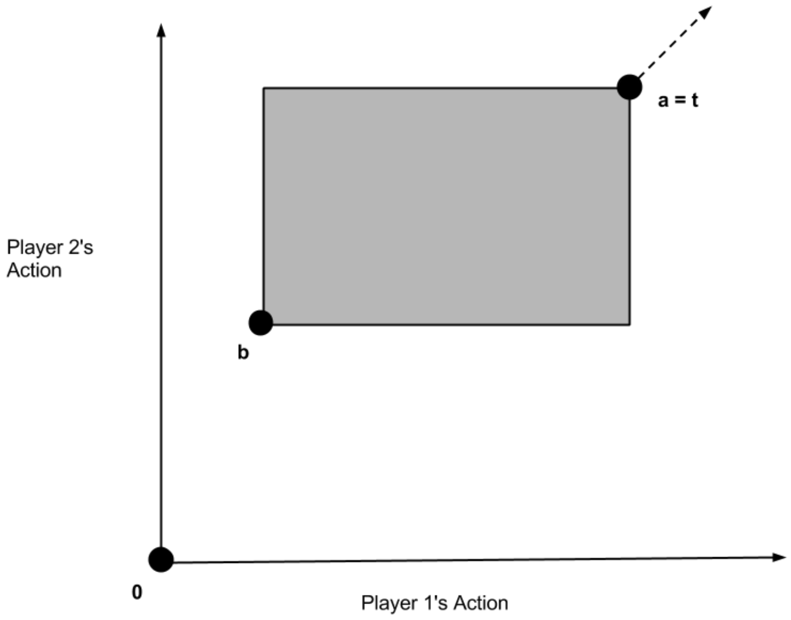

FIGURE 4.7: We have moved the bottom corner of the box up to $\mathbf{v}$.

We repeat this process until $t_i - b_i \leq \frac{\epsilon}{2nm}$.

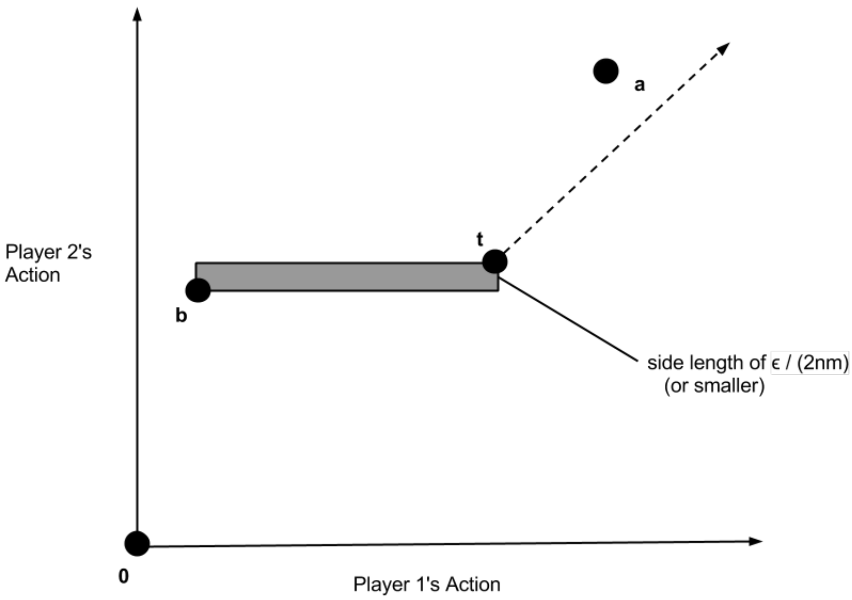

FIGURE 4.8: The box has shrunk so far that $t_i - b_i \leq \frac{\epsilon}{2nm}$.

We then freeze player 2's action at its current level, and we repeat the algorithm with a new player.

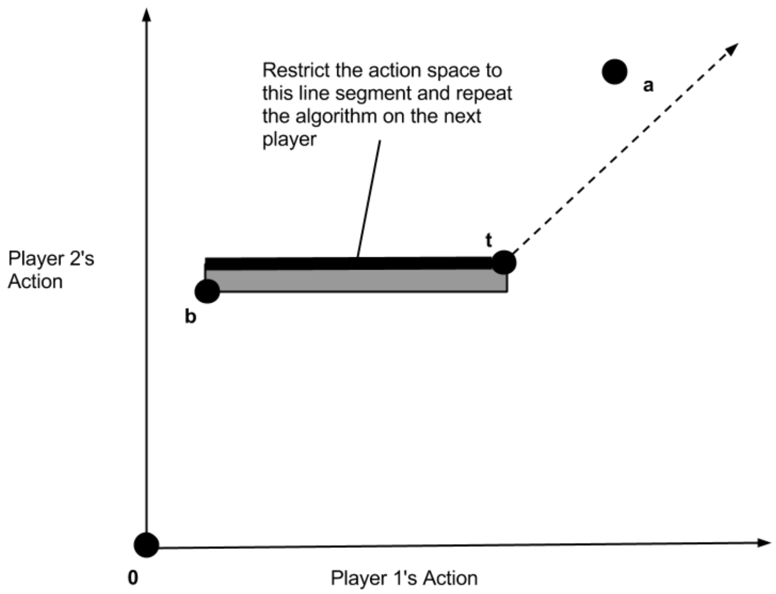

FIGURE 4.9: The algorithm recurses, with the bolded line segment serving as the initial "box" (we have dropped down 1 dimension).

After recursing through every possible player, our box now looks like this:

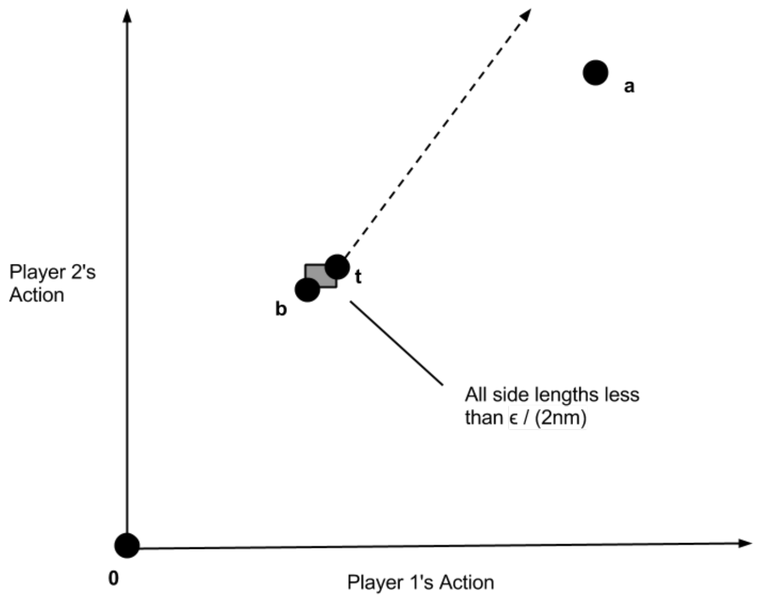

FIGURE 4.10: We have shrunk our box so far that every side length is less than $\frac{\epsilon}{2nm}$

Now we can simply walk from $\mathbf{t}$ straight to $\mathbf{b}$. This distance is no more than $\frac{\epsilon}{2m}$, so utilities will change along this line segment by no more than $\frac{\epsilon}{2}$. Since $\mathbf{t}$ satisfies $\mathbf{u}(\mathbf{t}) \leq \mathbf{u}(\mathbf{a}) + \frac{\epsilon}{2}$ (the reason for this error of $\frac{\epsilon}{2}$ is due to an error of distance $\frac{\epsilon}{2mn}$ in the search for $\mathbf{v}$; the reason behind this will be discussed later in our *convex optimization* subroutine), that means that $\mathbf{b}$ satisfies $\mathbf{u}(\mathbf{b}) \leq \mathbf{u}(\mathbf{a}) + \epsilon$.

We can therefore draw the following path:

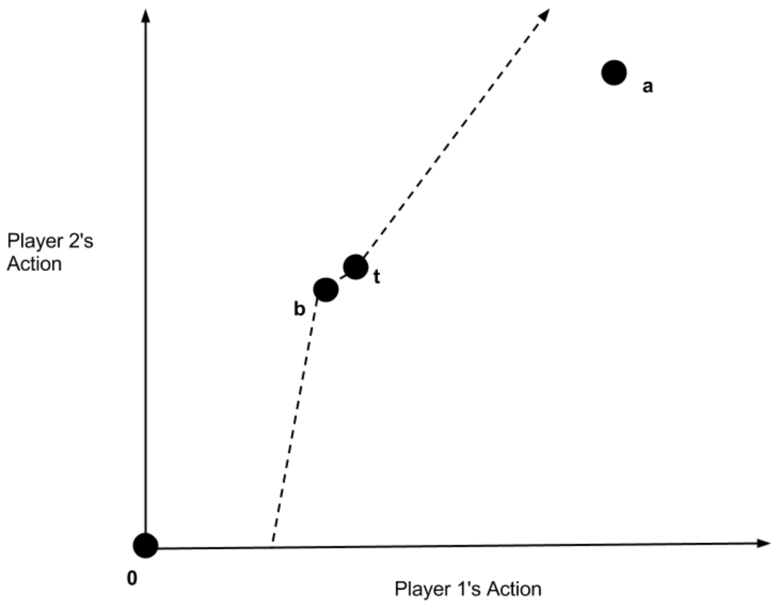

FIGURE 4.11: A path P of action vectors such that $\mathbf{u}(\mathbf{p}) \leq \mathbf{u}(\mathbf{a}) + \epsilon$ for all $\mathbf{p} \in P$.

This path implies that player 2 plays action 0 in any deviation from $\mathbf{a}$. Therefore, we can delete player 2 from N.

4.2.2 Delete-Or-Deviate Code

We can now complete our description of Delete-Or-Deviate by providing its implementation.

$\mathbf{b} \leftarrow \mathbf{0}$

$\mathbf{t} \leftarrow \mathbf{a}$

Pick any player $i \in C$

while $t_i - b_i > \frac{\epsilon}{2m}$ **do**

Let $\mathbf{v}$ be the result of the convex program: "Choose $\mathbf{v}$ to maximize $\sum_{j \neq i \in C} v_i$ subject to the constraints $v_i = \frac{b_i + t_i}{2}$ and $\mathbf{u}_{C \setminus \{i\}}(\mathbf{v}) \geq \mathbf{u}_{C \setminus \{i\}}(\mathbf{a})$ and $\mathbf{v}_F = \mathbf{a}_F$ and $\mathbf{v}_{N \setminus (C \cup F)} = \mathbf{0}$ within error $\frac{\epsilon}{2m}$."

if The program is not feasible **then**

$b_i \leftarrow v_i$

else if $\mathbf{u}_C(\mathbf{v}) \gneq \mathbf{u}_C(\mathbf{a})$ **then**

Output "**a** is not in the core."

Halt the program

else if $\mathbf{v}$ is above the Pareto frontier **then**

$\mathbf{t} \leftarrow \mathbf{v}$

else if $\mathbf{v}$ is below the Pareto frontier **then**

$\mathbf{b} \leftarrow \mathbf{v}$

end if

end while

if The convex program "Choose $\mathbf{v}$ subject to the constraints $v_i = t_i$ and $\mathbf{u}_{C\setminus\{i\}}(\mathbf{v}) \geq \mathbf{u}_{C\setminus\{i\}}(\mathbf{a})$ and $\mathbf{v}_F = \mathbf{a}_F$ and $\mathbf{v}_{N\setminus(C\cup F)} = \mathbf{0}$." is infeasible **and** $C = \{i\}$ **then**

Delete i from C at top level and return to top level

end if

$t_i \leftarrow b_i$

Call Delete-Or-Deviate($C \setminus \{i\}, F \cup \{i\}, \mathbf{t}, \mathbf{u}, \epsilon$)

4.3 Subroutine: Convex Optimization

Our implementation of the process described above relies on a secondary subroutine called *convex optimization.* Convex optimization is a body of study that solves problems of the form:

Problem. Input:

1. A convex function $f : \mathbb{R}^n \to \mathbb{R}$
2. A set of k convex functions $g_1, \ldots, g_k$, each $\mathbb{R}^n \to \mathbb{R}$.
3. A matrix $\mathbf{A} \in \mathbb{R}^{j\times n}$
4. A vector $\mathbf{c} \in \mathbb{R}^j$
5. A scalar $\epsilon > 0$

Let $\mathbf{x}^{\min}$ be the point that minimizes f within the *feasible set* $F = \{\mathbf{x} \,|\, g_1(\mathbf{x}) \leq 0, \ldots, g_n(\mathbf{x}) \leq 0, \mathbf{A}\mathbf{x} = \mathbf{c}\}$.

Output:

1. A point $\mathbf{x}$ for which $f(\mathbf{x}) - f(\mathbf{x}^{\min}) \leq \epsilon$

Convex optimization is a very deep field of theory and we will not attempt to do it justice here. Instead, we will simply mention that there exist well known algorithms that solve convex optimization problems in time polynomial in the values of n and $\frac{1}{\epsilon}$ (this assumes that j, k are polynomial in n) for continuous f. The reader is referred to Boyd & Vandenberghe [2] or Bertsekas [1] for a more detailed discussion of convex programming and the commonly used algorithms used to implement it.

Bibliography

[1] D. Bertsekas. *Convex Optimization Theory.* Cambridge university press, 2009.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge university press, 2004.

[3] X. Deng and C. Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19:257–266, 1994.

[4] T. Driessen. *Cooperative Games, Solutions and Applications.* Kluwer Academic Publishers, 1988.

[5] M. Elliott and B. Golub. A network centrality approach to coalitional stability. *Preprint, available at http://www.mit.edu/ bgolub/papers/jmp.pdf*, 2012.

[6] U. Faigle, S. Fekete, W. Hochstattler, and W. Kern. On the complexity of testing membership in the core of min-cost spanning tree games. *International Journal of Game Theory*, 26.3:361–366, 1997.

[7] G. Gigerenzer and R. Selten. *Bounded Rationality: The Adaptative Toolbox.* MIT Press, Cambridge, MA, 2001.

[8] Y. Kannai. The core and balancedness. *Handbook of Game Theory with Economic Applications*, I:355395, 1992.

[9] E. Sperner. Fifty years of further development of a combinatorial lemma. *Numberical solution of highly nonlinear problems*, pages Part A, p.183–197, Part B, p.199–214, 1980.

[10] S. C. Sung and D. Dimitrov. On core membership testing for hedonic coalition formation games. *Operations Research Letters*, 35(2):155–158, 2007.