The Cost of Stability in Coalitional Games

Yoram Bachrach[1], Edith Elkind[2,3], Reshef Meir[4], Dmitrii Pasechnik[3],
Michael Zuckerman[4], Jörg Rothe[5], and Jeffrey S. Rosenschein[4]

[1] Microsoft Research, Cambridge, United Kingdom
[2] University of Southampton, United Kingdom
[3] Nanyang Technological University, Singapore
[4] School of Engineering and Computer Science, Hebrew University, Jerusalem, Israel
[5] Institut für Informatik, Heinrich-Heine-Universität Düsseldorf, Germany

Abstract. A key question in cooperative game theory is that of coalitional stability, usually captured by the notion of the *core*—the set of outcomes such that no subgroup of players has an incentive to deviate. However, some coalitional games have empty cores, and any outcome in such a game is unstable.

In this paper, we investigate the possibility of stabilizing a coalitional game by using external payments. We consider a scenario where an external party, which is interested in having the players work together, offers a supplemental payment to the grand coalition (or, more generally, a particular coalition structure). This payment is conditional on players not deviating from their coalition(s). The sum of this payment plus the actual gains of the coalition(s) may then be divided among the agents so as to promote stability. We define the *cost of stability (CoS)* as the minimal external payment that stabilizes the game.

We provide general bounds on the cost of stability in several classes of games, and explore its algorithmic properties. To develop a better intuition for the concepts we introduce, we provide a detailed algorithmic study of the cost of stability in weighted voting games, a simple but expressive class of games which can model decision-making in political bodies, and cooperation in multiagent settings. Finally, we extend our model and results to games with coalition structures.

1 Introduction

In recent years, algorithmic game theory, an emerging field that combines computer science, game theory and social choice, has received much attention from the multiagent community [19, 8, 22, 20]. Indeed, multiagent systems research focuses on designing intelligent agents, i.e., entities that can coordinate, cooperate and negotiate without requiring human intervention. In many application domains, such agents are *self-interested*, i.e., they are built to maximize the rewards obtained by their creators. Therefore, these agents can be modeled naturally using game-theoretic tools. Moreover, as agents often have to function in rapidly changing environments, computational considerations are of great concern to their designers as well.

In many settings, such as online auctions and other types of markets, agents act individually. In this case, the standard notions of noncooperative game theory, such as *Nash equilibrium* or *dominant-strategy equilibrium*, provide a prediction of the outcome of the interaction. However, another frequently occurring type of scenario is that agents need to form teams to achieve their individual goals. In such domains, the focus turns from the interaction between single agents to the capabilities of subsets, or *coalitions*, of agents. Thus, a more appropriate modeling toolkit for this setting is that of *cooperative*, or *coalitional*, game theory [4], which studies what coalitions are most likely to arise, and how their members distribute the gains from cooperation. When agents are self-interested, the latter question is obviously of great importance. Indeed, the *total* utility generated by the coalition is of little interest to individual agents; rather, each agent aims to maximize her own utility. Thus, a *stable* coalition can be formed only if the gains from cooperation can be distributed in a way that satisfies all agents.

The most prominent solution concept that aims to formalize the idea of stability in coalitional games is the *core*. Informally, an *outcome* of a coalitional game is a *payoff vector* which for each agent lists her share of the profit of the *grand coalition*, i.e., the coalition that includes all agents. An outcome is said to be in the core if it distributes gains so that no subset of agents has an incentive to abandon the grand coalition and form a coalition of their own. It can be argued that the concept of the core captures the intuitive notion of stability in cooperative settings. However, it has an important drawback: the core of a game may be empty. In games with empty cores, any outcome is unstable, and therefore there is always a group of agents that is tempted to abandon the existing plan. This observation has triggered the invention of less demanding solution concepts, such as ε-core and the least core, as well as an interest in noncooperative approaches to identifying stable outcomes in coalitional games [5, 17].

In this paper, we approach this issue from a different perspective. Specifically, we examine the possibility of stabilizing the outcome of a game using external payments. Under this model, an external party (the *center*), which can be seen as a central authority interested in stable functioning of the system, attempts to incentivize a coalition of agents to cooperate in a stable manner. This party does this by offering the members of a coalition a supplemental payment if they cooperate. This external payment is given to the coalition as a whole, and is provided only if this coalition is formed.

Clearly, when the supplemental payment is large enough, the resulting outcome is stable: the profit that the deviators can make on their own is dwarfed by the subsidy they could receive by sticking to the prescribed solution. However, normally the external party would want to minimize its expenditure. Thus, in this paper we define and study the *cost of stability*, which is the minimal supplemental payment that is required to ensure stability in a coalitional game. We start by considering this concept in the context where the central authority aims to ensure that *all* agents cooperate, i.e., it offers a supplemental payment in order to stabilize the grand coalition. We then extend our analysis to the setting where the goal of the center is the stability of a *coalition structure*, i.e., a partition of all agents into disjoint coalitions. In this setting, the center does not expect the agents to work as a single team, but nevertheless wants each individual team to be immune to deviations. Finally, we consider the scenario where the center is concerned with the stability of a particular coalition within a coalition structure. This model is appropriate when the central authority wants a particular group of agents to work together, but is indifferent to other agents switching coalitions.

We first provide bounds on the cost of stability in general coalitional games. We then show that for some interesting special cases, such as super-additive games, these bounds can be improved considerably. We also propose a general algorithmic technique for computing the cost of stability. Then, to develop a better understanding of the concepts proposed in the paper, we apply them in the context of *weighted voting games* (WVGs), a simple but powerful class of games that have been used to model cooperation in settings as diverse as, on the one hand, decision-making in political bodies such as the United Nations Security Council and the International Monetary Fund and, on the other hand, resource allocation in multiagent systems. For such games, we are able to obtain a complete characterization of the cost of stability from an algorithmic perspective.

The paper is organized as follows. In Section 2, we provide the necessary background on coalitional games. In Section 3, we formally define the cost of stability for the setting where the desired outcome is the grand coalition, prove bounds on the cost of stability, and outline a general technique for computing it. We then focus on the computational aspects of the cost of stability in the context of our selected domain, i.e., weighted voting games. In Section 4.1, we demonstrate that computing the cost of stability in such games is coNP-hard if the weights are given in binary. On the other hand, for unary weights, we provide an efficient algorithm for this problem. We also investigate whether the cost of stability can be efficiently approximated. In Section 4.2, we answer this question positively by describing a fully polynomial-time approximation scheme (FPTAS) for our problem. We complement this result by showing that, by distributing the payments in a very natural manner, we get within a factor of 2 of the optimal adjusted gains, i.e., the sum of the value of the grand coalition and the external payments. While this method of allocating payoffs does not necessarily minimize the center's expenditure, the fact that it is both easy to implement and has a bounded worst-case performance may make it an attractive proposition in certain settings. In Section 5, we extend our discussion to the setting where the center aims to stabilize an arbitrary coalition structure, or a particular coalition within it, rather than the grand coalition. We end the paper with a discussion of related work and some conclusions.

We omit some of the proofs due to space constraints; the full version of the paper (with all proofs included) is available online [2]. A preliminary version of this paper was published in AAMAS'09 [3].

2 Preliminaries

Throughout this paper, given a vector $\boldsymbol{x} = (x_1, \dots, x_n)$ and a set $C \subseteq \{1, \dots, n\}$ we write $x(C)$ to denote $\sum_{i \in C} x_i$.

Definition 1. *A* (transferable utility) coalitional game $G = (I, v)$ *is given by a set of* agents *(synonymously,* players*)* $I = \{1, \dots, n\}$ *and a* characteristic function $v : 2^I \to \mathbb{R}^+ \cup \{0\}$ *that for any subset (coalition) of agents lists the total utility these agents achieve by working together. We assume* $v(\emptyset) = 0$.

A coalitional game $G = (I, v)$ is called *increasing* if for all coalitions $C' \subseteq C$ we have $v(C') \leq v(C)$, and *super-additive* if for all disjoint coalitions $C, C' \subseteq I$ we have $v(C) + v(C') \leq v(C \cup C')$. Note that since $v(C) \geq 0$ for any $C \subseteq I$, all super-additive games are increasing. A coalitional game $G = (I, v)$ is called *simple* if it is increasing and $v(C) \in \{0, 1\}$ for all $C \subseteq I$. In a simple game, we say that a coalition $C \subseteq I$ *wins* if $v(C) = 1$, and *loses* if $v(C) = 0$. Finally, a coalitional game is called *anonymous* if $v(C) = v(C')$ for any $C, C' \subseteq I$ such that $|C| = |C'|$. A particular class of simple games considered in this paper is that of *weighted voting games* (WVGs).

Definition 2. *A* weighted voting game *is a simple coalitional game given by a set of agents $I = \{1, \dots, n\}$, a vector $\mathbf{w} = (w_1, \dots, w_n)$ of nonnegative weights, where w_i is agent i's weight, and a threshold q. The* weight of a coalition *$C \subseteq I$ is $w(C) = \sum_{i \in C} w_i$. A coalition C* wins the game *(i.e., $v(C) = 1$) if $w(C) \geq q$, and* loses the game *(i.e., $v(C) = 0$) if $w(C) < q$.*

We denote the WVG with the weights $\mathbf{w} = (w_1, \dots, w_n)$ and the threshold q as $[\mathbf{w}; q]$ or $[w_1, \dots, w_n; q]$. Also, we set $w_{\max} = \max_{i \in I} w_i$. It is easy to see that WVGs are simple games; however, they are not necessarily super-additive. Throughout this paper, we assume that $w(I) \geq q$, i.e., the grand coalition wins.

The characteristic function of a coalitional game defines only the *total* gains a coalition achieves, but does not offer a way of distributing them among the agents. Such a division is called an imputation (or, sometimes, a payoff vector).

Definition 3. *Given a coalitional game $G = (I, v)$, a vector $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}^n$ is called an* imputation *for G if it satisfies $p_i \geq v(\{i\})$ for each i, $1 \leq i \leq n$, and $\sum_{i=1}^{n} p_i = v(I)$. We call p_i the* payoff of agent *i; the* total payoff of a coalition *$C \subseteq I$ is given by $p(C)$. We write $\mathcal{I}(G)$ to denote the set of all imputations for G.*

For an imputation to be stable, it should be the case that no subset of players has an incentive to deviate. Formally, we say that a coalition C *blocks* an imputation $\mathbf{p} = (p_1, \dots, p_n)$ if $p(C) < v(C)$. The *core* of a coalitional game G is defined as the set of imputations not blocked by any coalition, i.e., $\text{core}(G) = \{\mathbf{p} \in \mathcal{I}(G) \mid p(C) \geq v(C) \text{ for each } C \subseteq I\}$. An imputation in the core guarantees the stability of the grand coalition. However, the core can be empty.

In WVGs, and, more generally, in simple games, one can characterize the core using the notion of veto agents, i.e., agents that are indispensable for forming a winning coalition. Formally, given a simple coalitional game $G = (I, v)$, an agent $i \in I$ is said to be a *veto agent* if for all coalitions $C \subseteq I \setminus \{i\}$ we have $v(C) = 0$. The following is a folklore result regarding nonemptiness of the core.

Theorem 1. *Let $G = (I, v)$ be a simple coalitional game. If there are no veto agents in G, then the core of G is empty. Otherwise, let $I' = \{i_1, \dots, i_m\}$ be the set of veto agents in G. Then the core of G is the set of imputations that distribute all the gains among the veto agents only, i.e.,* $\text{core}(G) = \{\mathbf{p} \in \mathcal{I}(G) \mid p(I') = 1\}$.

So far, we have tacitly assumed that the only possible outcome of a coalitional game is the formation of the grand coalition. However, often it makes more sense for the agents to form several disjoint coalitions, each of which can focus on its own task. For example, WVGs can be used to model the setting where each agent has a certain amount of resources (modeled by her weight), and there are a number of identical tasks each of which requires a certain amount of these resources (modeled by the threshold) to be completed. In this setting, the formation of the grand coalition means that only one task will be completed, even if there are enough resources for several tasks.

The situation when agents can split into teams to work on several tasks simultaneously can be modeled using the notion of a coalition structure, i.e., a partition of the set of agents into disjoint coalitions. Formally, we say that $CS = (C^1, \dots, C^m)$ is a *coalition structure* over a set of agents I if $\bigcup_{i=1}^m C^i = I$ and $C^i \cap C^j = \emptyset$ for all $i \neq j$; we write $CS \in \mathcal{CS}(I)$. Also, we overload notation by writing $v(CS)$ to denote $\sum_{C^j \in CS} v(C^j)$. If coalition structures are allowed, an outcome of a game is not just an imputation, but a pair $(CS, \mathbf{p})$, where $\mathbf{p}$ is an imputation for the coalition structure CS, i.e., $\mathbf{p}$ distributes the gains of every coalition in CS among its members. Formally, we say that $\mathbf{p} = (p_1, \dots, p_n)$ is an *imputation for a coalition structure* $CS = (C^1, \dots, C^m)$ in a game $G = (I, v)$ if $p_i \geq 0$ for all i, $1 \leq i \leq n$, and $p(C^j) = v(C^j)$ for all j, $1 \leq j \leq m$; we write $\mathbf{p} \in \mathcal{I}(CS, G)$. We can also generalize the notion of the core introduced earlier in this section to games with coalition structures. Namely, given a game $G = (I, v)$, we say that an outcome $(CS, \mathbf{p})$ is in the *CS-core of* G if CS is a coalition structure over I, $\mathbf{p} \in \mathcal{I}(CS, G)$ and $p(C) \geq v(C)$ for all $C \subseteq I$; we write $(CS, \mathbf{p}) \in$ *CS-core*(G). Note that if $\mathbf{p}$ is in the core of G then $(I, \mathbf{p})$ is in the CS-core of G; however, the converse is not necessarily true.

3 The Cost of Stability

In many games, forming the grand coalition maximizes social welfare; this happens, for example, in super-additive games. However, the core of such games may still be empty. In this case, it would be impossible to distribute the gains of the grand coalition in a stable way, so it may fall apart despite being socially optimal. Thus, an external party, such as a benevolent central authority, may want to incentivize the agents to cooperate, e.g., by offering the agents a supplemental payment Δ if they stay in the grand coalition. This situation can be modeled as an *adjusted coalitional game* derived from the original coalitional game G.

Definition 4. *Given a coalitional game $G = (I, v)$ and $\Delta \geq 0$, the* adjusted coalitional game *$G(\Delta) = (I, v')$ is given by $v'(C) = v(C)$ for $C \neq I$, and $v'(I) = v(I) + \Delta$.*

We call $v'(I) = v(I) + \Delta$ the *adjusted gains* of the grand coalition. We say that a vector $\mathbf{p} \in \mathbb{R}^n$ is a *super-imputation* for a game $G = (I, v)$ if $p_i \geq 0$ for all $i \in I$ and $p(I) \geq v(I)$. Furthermore, we say that a super-imputation $\mathbf{p}$ is *stable* if $p(C) \geq v(C)$ for all $C \subseteq I$. A super-imputation $\mathbf{p}$ with $p(I) = v(I) + \Delta$ distributes the adjusted gains, i.e., it is an imputation for $G(\Delta)$; it is stable if and only if it is in the core of $G(\Delta)$. We say that a supplemental payment Δ *stabilizes* the grand coalition in a game G if the adjusted game $G(\Delta)$ has a nonempty core. Clearly, if Δ is large enough (e.g., $\Delta = n \max_{C \subseteq I} v(C)$), the game $G(\Delta)$ will have a nonempty core. However, usually the central authority wants to spend as little money as possible. Hence, we define the cost of stability as the *smallest* external payment that stabilizes the grand coalition.

Definition 5. *Given a coalitional game $G = (I, v)$, its* cost of stability *$CoS(G)$ is defined as $CoS(G) = \inf\{\Delta \mid \Delta \geq 0 \text{ and } \mathrm{core}(G(\Delta)) \neq \emptyset\}$.*

We have argued that the set $\{\Delta \mid \Delta \geq 0 \text{ and } \mathrm{core}(G(\Delta)) \neq \emptyset\}$ is nonempty. Therefore, $G(\Delta)$ is well-defined. Now, we prove that this set contains its greatest lower bound

$CoS(G)$, i.e., that the game $G(CoS(G))$ has a nonempty core. While this can be shown using a continuity argument, we will now give a different proof, which will also be useful for exploring the cost of stability from an algorithmic perspective. Fix a coalitional game $G = (I, v)$ and consider the following linear program $\mathcal{LP}^*$:

$$\begin{aligned} &\min \Delta \quad \text{subject to:} \\ &\Delta \geq 0, && (1) \\ &p_i \geq 0 \quad \text{for each } i = 1, \dots, n, && (2) \\ &\sum_{i \in I} p_i = v(I) + \Delta, && (3) \\ &\sum_{i \in C} p_i \geq v(C) \quad \text{for all } C \subseteq I. && (4) \end{aligned}$$

It is not hard to see that the optimal value of this linear program is exactly $CoS(G)$. Moreover, any optimal solution of $\mathcal{LP}^*$ corresponds to an imputation in the core of $G(CoS(G))$ and therefore the game $G(CoS(G))$ has a nonempty core.

As an example, consider a uniform weighted voting game, i.e., a WVG $G = [\mathbf{w}; q]$ with $w_1 = \dots = w_n = w$. We can derive an explicit formula for $CoS(G)$.

Theorem 2. *For a WVG $G = [w, w, \dots, w; q]$, we have $CoS(G) = \frac{n}{\lceil q/w \rceil} - 1$.*

For example, if $w(n-1) < q \leq wn$, then $CoS(G) = 0$, i.e., G has a nonempty core. On the other hand, if $w = 1$, $n = 3k$ and $q = 2k$ for some integer $k > 0$, i.e., $q = \frac{2}{3}n$, we have $CoS(G) = \frac{3}{2} - 1 = \frac{1}{2}$.

3.1 Bounds on $CoS(G)$ in General Coalitional Games

Consider an arbitrary coalitional game $G = (I, v)$. Clearly, $CoS(G) = 0$ if and only if G has a nonempty core. Further, we have argued that $CoS(G)$ is upper-bounded by $n \max_{C \subseteq I} v(C)$, i.e., $CoS(G)$ is finite for any fixed coalitional game. Moreover, the bound of $n \max_{C \subseteq I} v(C)$ is (almost) tight. To see this, consider a (simple) game G' given by $v'(\emptyset) = 0$ and $v'(C) = 1$ for all $C \neq \emptyset$. Clearly, we have $CoS(G') = n - 1$: any super-imputation that pays some agent less than 1 will not be stable, whereas setting $p_i = 1$ for all $i \in I$ ensures stability. Thus, the cost of stability can be quite large relative to the value of the grand coalition.

On the other hand, we can provide a lower bound on $CoS(G)$ in terms of the values of coalition structures over I. Indeed, for an arbitrary coalition structure $CS \in \mathcal{CS}(I)$, we have $CoS(G) \geq v(CS) - v(I)$. To see this, note that if the total payment to the grand coalition is less than $(v(CS) - v(I)) + v(I)$, then for some coalition $C \in CS$ it will be the case that $p(C) < v(C)$. It would be tempting to conjecture that $CoS(G) = \max_{CS \in \mathcal{CS}(I)}(v(CS) - v(I))$. However, a counterexample is provided by Theorem 2 with $w = 1$, $q = \frac{2}{3}n$: indeed, in this case we have $CoS(G) = \frac{1}{2}$, yet $\max_{CS \in \mathcal{CS}(I)}(v(CS) - v(I)) = 0$. We can summarize these observations as follows.

Theorem 3. *For any coalitional game $G = (I, v)$, we have*

$$\max_{CS \in \mathcal{CS}(I)}(v(CS) - v(I)) \leq CoS(G) \leq n \max_{C \subseteq I} v(C).$$

For super-additive games, we can strengthen the upper bound considerably. Note that in such games the grand coalition maximizes social welfare, so its stability is particularly desirable. Yet, as the second part of Theorem 4 implies, ensuring stability may turn out to be quite costly even in this restricted setting.

Theorem 4. *For any super-additive game $G = (I, v)$, $|I| = n$, we have $CoS(G) \leq (\sqrt{n} - 1)v(I)$, and this bound is asymptotically tight.*

For anonymous super-additive games, further improvements are possible.

Theorem 5. *For any anonymous super-additive game $G = (I, v)$, we have $CoS(G) \leq 2v(I)$, and this bound is asymptotically tight.*

A somewhat similar stability-related concept is the *least core*, which is the set of all imputations $\mathbf{p}$ that minimize the maximal *deficit* $v(C) - p(C)$. In particular, the *value* of the least core $\varepsilon(G)$, defined as $\varepsilon(G) = \inf_{\mathbf{p}\in\mathcal{I}(G)}\{\max\{v(C) - p(C) \mid C \subseteq I\}\}$, is strictly positive if and only if the cost of stability is strictly positive. The following proposition provides a more precise description of the relationship between the value of the least core and the cost of stability.

Proposition 1. *For any coalitional game $G = (I, v)$ with $|I| = n$ such that $\varepsilon(G) \geq 0$, we have $CoS(G) \leq n\varepsilon(G)$, and this bound is asymptotically tight.*

3.2 Algorithmic Properties of $CoS(G)$

The linear program $\mathcal{LP}^*$ provides a way of computing $CoS(G)$ for any coalitional game G. However, this linear program contains exponentially many constraints (one for each subset of I). Thus, solving it directly would be too time-consuming for most games. Note that for general coalitional games, this is, in a sense, inevitable: in general, a coalitional game is described by its characteristic function, i.e., a list of 2^n numbers. Thus, to discuss the algorithmic properties of $CoS(G)$, we need to restrict our attention to games with compactly representable characteristic functions.

A standard approach to this issue is to consider games that can be described by polynomial-size circuits. Formally, we say that a class $\mathcal{G}$ of games has a *compact circuit representation* if there exists a polynomial p such that for every $G \in \mathcal{G}$, $G = (I, v)$, $|I| = n$, there exists a circuit $\mathcal{C}$ of size $p(n)$ with n binary inputs that on input $(b_1, \dots, b_n)$ outputs $v(C)$, where $C = \{i \in I \mid b_i = 1\}$.

Unfortunately, it turns out that having a compact circuit representation does not guarantee efficient computability of $CoS(G)$. Indeed, it is easy to see that WVGs with integer weights have such a representation. However, in the next section we will show that computing $CoS(G)$ for such games is computationally intractable (Theorem 7). We can, however, provide a *sufficient* condition for $CoS(G)$ to be efficiently computable. To do so, we will first formally state the relevant computational problems.

SUPER-IMPUTATION-STABILITY: Given a coalitional game G (compactly represented by a circuit), a supplemental payment $\varDelta$ and an imputation $\mathbf{p} = (p_1, \dots, p_n)$ in the adjusted game $G(\varDelta)$, decide whether $\mathbf{p} \in \mathrm{core}(G(\varDelta))$.

COS: Given a coalitional game G (compactly represented by a circuit) and a parameter $\varDelta$, decide whether $CoS(G) \leq \varDelta$, i.e., whether $\mathrm{core}(G(\varDelta)) \neq \emptyset$.

Consider first SUPER-IMPUTATION-STABILITY. Fix a game $G = (I, v)$. For any super-imputation $\mathbf{p}$ for G, let $d(G, \mathbf{p}) = \max_{C \subseteq I}(v(C) - p(C))$ be the maximum deficit of a coalition under $\mathbf{p}$. Clearly, $\mathbf{p}$ is stable if and only if $d(G, \mathbf{p}) \leq 0$. Observe also that for any $\Delta > 0$ it is easy to decide whether $\mathbf{p}$ is an imputation for $G(\Delta)$. Thus, a polynomial-time algorithm for computing $d(G, \mathbf{p})$ can be converted into a polynomial-time algorithm for SUPER-IMPUTATION-STABILITY. Further, we can decide COS via solving $\mathcal{LP}^*$ by the ellipsoid method. The ellipsoid method runs in polynomial time given a polynomial-time *separation oracle*, i.e., a procedure that takes as input a candidate feasible solution, checks if it indeed is feasible, and if this is not the case, returns a violated constraint. Now, given a vector $\mathbf{p}$ and a parameter Δ, we can easily check if they satisfy constraints (1)–(3), i.e., if $\mathbf{p}$ is an imputation for $G(\Delta)$. To verify constraint (4), we need to check if $\mathbf{p}$ is in the core of $G(\Delta)$. As argued above, this can be done by checking whether $d(G, \mathbf{p}) \leq 0$. We summarize these results as follows.

Theorem 6. *Consider a class of coalitional games $\mathcal{G}$ with a compact circuit representation. If there is an algorithm that for any $G \in \mathcal{G}$, $G = (I, v)$, $|I| = n$, and for any super-imputation $\mathbf{p}$ for G computes $d(G, \mathbf{p})$ in time $\mathrm{poly}(n, |\mathbf{p}|)$, where $|\mathbf{p}|$ is the number of bits in the binary representation of $\mathbf{p}$, then for any $G \in \mathcal{G}$ the problems* SUPER-IMPUTATION-STABILITY *and* COS *are polynomial-time solvable.*

We mention in passing that for games with poly-time computable characteristic functions both problems are in coNP. For SUPER-IMPUTATION-STABILITY, the membership is trivial; for COS, it follows from the fact that the game $G(\Delta)$ has a poly-time computable characteristic function as long as G does, and hence we can apply the results of [14] (see the proof of Theorem 7 for details).

4 Cost of Stability in WVGs Without Coalition Structures

In this section, we focus on computing the cost of stabilizing the grand coalition in WVGs. We start by considering the complexity of exact algorithms for this problem.

4.1 Exact Algorithms

In what follows, unless specified otherwise, we assume that all weights and the threshold are integers given in binary, whereas all other numeric parameters, such as the supplemental payment Δ and the entries of the payoff vector $\mathbf{p}$, are rationals given in binary. Standard results on linear threshold functions [16] imply that WVGs with integer weights have a compact circuit representation. Thus, we can define the computational problems SUPER-IMPUTATION-STABILITY-WVG and COS-WVG by specializing the problems SUPER-IMPUTATION-STABILITY and COS to WVGs. Both of the resulting problems turn out to be computationally hard.

Theorem 7. *The problems* SUPER-IMPUTATION-STABILITY-WVG *and* COS-WVG *are* coNP*-complete.*

The reductions in the proof of Theorem 7 are from PARTITION. Consequently, our hardness results depend in an essential way on the weights being given in binary. Thus, it is natural to ask what happens if the agents' weights are polynomially bounded (or given in unary). It turns out that in this case the results of Section 3.2 imply that SUPER-IMPUTATION-STABILITY-WVG and COS-WVG are in P, since for WVGs with small weights one can compute $d(G, \mathbf{p})$ in polynomial time.

Theorem 8. SUPER-IMPUTATION-STABILITY-WVG *and* COS-WVG *are in* P *when the agents' weights are polynomially bounded (or given in unary).*

4.2 Approximating the Cost of Stability in Weighted Voting Games

For large weights, the algorithms outlined at the end of the previous section may not be practical. Thus, the center may want to trade off its payment and computation time, i.e., provide a slightly higher supplemental payment for which the corresponding stable super-imputation can be computed efficiently. It turns out that this is indeed possible, i.e., $CoS(G)$ can be efficiently approximated to an arbitrary degree of precision.

Theorem 9. *There exists an algorithm $\mathcal{A}(G, \varepsilon)$ that, given a WVG $G = [\mathbf{w}; q]$ in which the weights of all players are nonnegative integers given in binary and a parameter $\varepsilon > 0$, outputs a value Δ that satisfies $CoS(G) \leq \Delta \leq (1 + \varepsilon)\, CoS(G)$ and runs in time* $\mathrm{poly}(n, \log w_{\max}, 1/\varepsilon)$. *That is, there exists a fully polynomial-time approximation scheme (FPTAS) for $CoS(G)$.*

Moreover, one can get a 2-approximation to the adjusted gains simply by paying each agent in proportion to her weight, and this bound can be shown to be tight.

Theorem 10. *For any WVG $G = [\mathbf{w}; q]$ with $CoS(G) = \Delta$, the super-imputation $\mathbf{p}^*$ given by $p^*_i = \min\{1, \frac{w_i}{q}\}$ is stable and satisfies $p^*(I) \leq 2p(I)$ for any super-imputation* $\mathbf{p} \in \mathrm{core}(G(\Delta))$.

5 Cost of Stability in Games with Coalition Structures

If a coalitional game is not super-additive, the formation of the grand coalition is not necessarily the most desirable outcome: for example, it may be the case that by splitting into several teams the agents can accomplish more tasks than by working together. In such settings, the central authority may want to stabilize a coalition structure, i.e., a partition of agents into teams. We now generalize the cost of stability to such settings.

5.1 Stabilizing a Fixed Coalition Structure

We first consider the setting where the central authority wants to stabilize a particular coalition structure.

Given a coalitional game $G = (I, v)$, a coalition structure $CS = (C^1, \dots, C^m)$ over I and a vector $\boldsymbol{\Delta} = (\Delta^1, \dots, \Delta^m)$, let $G(\boldsymbol{\Delta})$ be the game with the set of agents I and the characteristic function v' given by $v'(C^i) = v(C^i) + \Delta^i$ for $i = 1, \dots, m$ and

$v'(C) = v(C)$ for any $C \notin \{C^1, \dots, C^m\}$. We say that the game $G(\boldsymbol{\Delta})$ is *stable with respect to* CS if there exists an imputation $\mathbf{p} \in \mathcal{I}(CS, G(\boldsymbol{\Delta}))$ such that $(CS, \mathbf{p})$ is in the CS-core of $G(\boldsymbol{\Delta})$. Also, we say that an external payment Δ *stabilizes* a coalition structure CS with respect to a game G if there exist $\Delta^1 \geq 0, \dots, \Delta^m \geq 0$ such that $\Delta = \Delta^1 + \dots + \Delta^m$ and the game $G(\boldsymbol{\Delta})$ is stable with respect to CS. We are now ready to define the cost of stability of a coalition structure CS in G.

Definition 6. *Given a coalitional game* $G = (I, v)$ *and a coalition structure* $CS = (C^1, \dots, C^m)$ *over* I*, the* cost of stability $CoS(CS, G)$ *of the coalition structure* CS *in* G *is the smallest external payment needed to stabilize* CS*, i.e.,*

$$CoS(CS, G) = \inf\{\sum_{i=1}^{m} \Delta^i \mid \Delta^i \geq 0 \text{ for } i = 1, \dots, m \quad \text{and}$$
$$\exists \mathbf{p} \in \mathcal{I}(CS, G(\boldsymbol{\Delta})) \quad \text{s.t.} \quad (CS, \mathbf{p}) \in \textit{CS-core}(G(\boldsymbol{\Delta}))\}.$$

Fix a game $G = (I, v)$ and set $v_{\max} = \max_{C \subseteq I} v(C)$. It is easy to see that for any coalition structure $CS = (C^1, \dots, C^m)$ the game $G(\boldsymbol{\Delta})$, where $\Delta^i = |C^i| v_{\max}$, is stable with respect to CS, and therefore $CoS(CS, G)$ is well-defined and satisfies $CoS(CS, G) \leq n v_{\max}$. Moreover, as in the case of games without coalition structures, the value $CoS(CS, G)$ can be obtained as an optimal solution to a linear program. Indeed, we can simply take the linear program $\mathcal{LP}^*$ and replace the constraint $\sum_{i \in I} p_i = v(I) + \Delta$ with the constraint $\sum_{i \in I} p_i = v(CS) + \Delta$. It is not hard to see that the resulting linear program, which we will denote by $\mathcal{LP}^*_{CS}$, computes $CoS(CS, G)$: in particular, the constraints $\Delta^i \geq 0$ for $i = 1, \dots, m$ are implicitly captured by the constraints $\sum_{i \in C^i} p_i \geq v(C^i)$ in line (4) of $\mathcal{LP}^*_{CS}$.

We now turn to the question of computing the cost of stability of a given coalition structure in WVGs. To this end, we will modify the decision problems stated in Section 4.1 as follows.

SUPER-IMPUTATION-STABILITY-WVG-CS: Given a WVG $G = [\mathbf{w}; q]$ with the set of agents I, a coalition structure $CS = (C^1, \dots, C^m)$ over I, a vector $\boldsymbol{\Delta} = (\Delta^1, \dots, \Delta^m)$ and an imputation $\mathbf{p} \in \mathcal{I}(CS, G(\boldsymbol{\Delta}))$, decide if $(CS, \mathbf{p})$ is in the CS-core of $G(\boldsymbol{\Delta})$.

COS-WVG-CS: Given a WVG $G = [\mathbf{w}; q]$ with the set of agents I, a coalition structure CS over I and a parameter Δ, decide whether $CoS(CS, G) \leq \Delta$.

The results of Section 4.1 immediately imply that both of these problems are computationally hard even for $m = 1$. Moreover, using the results of [9], we can show that SUPER-IMPUTATION-STABILITY-WVG-CS remains coNP-complete even if $\boldsymbol{\Delta}$ is fixed to be $(0, \dots, 0)$. On the other hand, when weights are integers given in unary, both COS-WVG-CS and SUPER-IMPUTATION-STABILITY-WVG-CS are polynomial-time solvable. Indeed, to solve SUPER-IMPUTATION-STABILITY-WVG-CS, one needs to check if there is a coalition C with $w(C) \geq q$, $p(C) < 1$. This can be done using the dynamic programming algorithm from the proof of Theorem 8. Moreover, to solve COS-WVG-CS, we can simply run the ellipsoid algorithm on the linear program $\mathcal{LP}^*_{CS}$ described earlier in this section, using the algorithm for SUPER-IMPUTATION-STABILITY-WVG-CS as a separation oracle. Thus, we obtain the following result.

Theorem 11. *When all players' weights are integers given in unary, the problems* CoS-WVG-CS *and* Super-Imputation-Stability-WVG-CS *are in* P.

Finally, we adapt the approximation algorithm presented in Section 4.2 to this setting.

Theorem 12. *There exists an FPTAS for $CoS(CS, G)$ in WVGs.*

5.2 Finding the Cheapest Coalition Structure to Stabilize

So far, we have focused on the setting where the external party wants to stabilize a particular coalition structure. However, it can also be the case that the central authority simply wants to achieve stability, and does not care which coalition structure arises, as long as it can be made stable using as little money as possible. We will now introduce the notion of *cost of stability for games with coalition structures* to capture this type of setting. Recall that $\mathcal{CS}(I)$ denotes the set of all coalition structures over I.

Definition 7. *Given a coalitional game $G = (I, v)$, let the* cost of stability for G with coalition structures, *denoted by $CoS_{CS}(G)$, be* $\min\{CoS(CS, G) \mid CS \in \mathcal{CS}(I)\}$.

Clearly, one can compute $CoS_{CS}(G)$ by enumerating all coalition structures over I and picking the one with the smallest value of $CoS(CS, G)$. Alternatively, note that the linear program $\mathcal{LP}^*_{CS}$ depends only on the value of the coalition structure CS. Hence, stabilizing all coalition structures with the same total value has the same cost. Moreover, this implies that the cheapest coalition structure to stabilize is the one that maximizes social welfare. Hence, if we could compute the value of the coalition structure CS^* that maximizes social welfare, we could find $CoS_{CS}(G)$ by solving $\mathcal{LP}^*_{CS^*}$.

For WVGs, paper [9] (see Theorem 2 there) shows that if weights are given in binary, it is NP-hard to decide whether a given game has a nonempty CS-core. As this question is equivalent to asking whether $CoS_{CS}(G) = 0$, the latter problem is NP-hard, too. One might hope that computing $CoS_{CS}(G)$ is easy if the weights of all players are given in unary. However, this does not seem to be the case. Indeed, our algorithms for computing the cost of stability in other settings relied on solving the corresponding linear program. To implement this approach in our scenario, we would need to compute the value of the coalition structure that maximizes social welfare. However, a straightforward reduction from 3-Partition, a classic problem that is known to be NP-hard even for unary weights, shows that the latter problem is NP-hard even if weights are given in unary. While this does not immediately imply that computing $CoS_{CS}(G)$ is hard for small weights, it means that finding the cheapest-to-stabilize outcome is NP-hard even if weights are given in unary.

5.3 Stabilizing a Particular Coalition

We now consider the case where the central authority wants a particular group of agents to work together, but does not care about the stability of the overall game. Thus, it wants to identify a coalition structure containing a particular coalition C and the minimal subsidy to the players that ensures that no set of players that includes members of C wants to deviate. We omit the formal definition of the corresponding cost of stability concept, as well as its algorithmic analysis due to space constraints. However, we would like to mention several subtle points that arise in this context. First, one might think that the optimal way to stabilize a coalition is to offer payments to members of this coalition only. However, this turns out not to be true (see [2]). Second, stabilizing a given coalition may be strictly cheaper than stabilizing *any* of the coalition structures that contain it (see [2]). Thus choosing a good definition of the cost of stability of an individual coalition is a nontrivial issue.

6 Related Work

The complexity of various solution concepts in coalitional games is a well-studied topic [6, 13, 7, 23]. In particular, [10] analyzes some important computational aspects of stability in WVGs, proving a number of results on the complexity of the least core and the nucleolus. The complexity of the CS-core in WVGs is studied in [9]. Paper [15] is similar to ours in spirit. It considers the setting where an external party intervenes in order to achieve a certain outcome using monetary payments. However, [15] deals with the very different domain of *non*cooperative games. There are also similarities between our work and the recent research on bribery in elections [11], where an external party pays voters to change their preferences in order to make a given candidate win. A companion paper [18] studies the cost of stability in network flow games.

7 Conclusion

We have examined the possibility of stabilizing a coalitional game by offering the agents additional payments in order to discourage them from deviating, and defined the cost of stability as the minimal total payment that allows a stable division of the gains. We focused on the computational aspects of this concept for weighted voting games. In the setting where the outcome to be stabilized is the grand coalition, we provided a complete picture of the computational complexity of the related decision problems. We then extended our results to settings where agents can form a coalition structure.

There are several lines of possible future research. First, while the focus of this paper was on weighted voting games, the notion of the cost of stability is defined for any coalitional game. Therefore, a natural research direction is to study the cost of stability in other classes of games. Second, we would like to develop a better understanding of the relationship between the cost of stability of a game, and its least core and nucleolus. Finally, it would be interesting to extend the notion of the cost of stability to games with nontransferable utility and partition function games.

Acknowledgments: We thank the reviewers for many helpful comments. This work was supported in part by DFG grants RO 1202/{11-1, 12-1}, the ESF's EUROCORES program LogICCC, the Alexander von Humboldt Foundation's TransCoop program, EPSRC grant GR/T10664/01, ESRC grant ES/F035845/1, NRF Research Fellowship, Singapore Ministry of Education AcRF Tier 1 Grant, and by ISF grant #898/05.

References

1. R. J. Aumann and S. Hart, editors. *Handbook of Game Theory, with Economic Applications, Vol. 2.* North-Holland, 1994.
2. Y. Bachrach, E. Elkind, R. Meir, D. Pasechnik, M. Zuckerman, J. Rothe, and J. S. Rosenschein. The cost of stability and its application to weighted voting games. Tech. Rep. arXiv: 0907.4385 [cs.GT], ACM Comp. Research Repository, July 2009. Preliminary version: [3].
3. Y. Bachrach, R. Meir, M. Zuckerman, J. Rothe, and J. S. Rosenschein. The cost of stability in weighted voting games (extended abstract). In *Proc. of AAMAS-09*, pp. 1289–1290, 2009.
4. R. Branzei, D. Dimitrov, and S. Tijs. *Models in Cooperative Game Theory*. Springer, 2008.
5. K. Chatterjee, B. Dutta, and K. Sengupta. A noncooperative theory of coalitional bargaining. *Review of Economic Studies*, 60:463–477, 1993.
6. V. Conitzer and T. Sandholm. Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains. In *Proc. of AAAI-04*, pp. 219–225, 2004.
7. V. Conitzer and T. Sandholm. Complexity of constructing solutions in the core based on synergies among coalitions. *Artificial Intelligence*, 170(6-7):607–619, 2006.
8. E. Ephrati and J. S. Rosenschein. The Clarke Tax as a consensus mechanism among automated agents. In *Proc. of IJCAI-91*, pp. 173–178, 1991.
9. E. Elkind, G. Chalkiadakis, and N. R. Jennings. Coalition structures in weighted voting games. In *Proc. of ECAI-08*, pp. 393–397, 2008
10. E. Elkind, L. A. Goldberg, P. W. Goldberg, and M. Wooldridge. Computational complexity of weighted threshold games. In *Proc. of AAAI-07*, pp. 718–723, 2007.
11. P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. The complexity of bribery in elections. In *Proc. of AAAI-06*, pp. 641–646, 2006. Full version to appear in *JAIR*.
12. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.
13. S. Ieong and Y. Shoham. Marginal contribution nets: A compact representation scheme for coalitional games. In *Proc. of EC-05*, pp. 193–292, 2005.
14. E. Malizia, L. Palopoli, and F. Scarcello. Infeasibility certificates and the complexity of the core in coalitional games. In *Proc. of IJCAI-07*, pp. 1402-1407, 2007.
15. D. Monderer and M. Tennenholtz. K-implementation. *Journal of Artificial Intelligence Research*, 21:37–62, 2004.
16. S. Muroga. Threshold Logic and its Applications. John Wiley & Sons, 1971.
17. A. Okada. A noncooperative coalitional bargaining game with random proposers. *Games and Economic Behavior*, 16:97–108, 1996.
18. E. Resnick, Y. Bachrach, R. Meir, and J. S. Rosenschein. The cost of stability in network flow games. To appear in *Proc. of MFCS-09*, 2009.
19. J. S. Rosenschein and M. R. Genesereth. Deals among rational agents. In *Proc. of IJCAI-85*, pp. 91–99, 1985.
20. T. Sandholm and V. Lesser. Issues in automated negotiation and electronic commerce: Extending the contract net framework. In *Proc. of ICMAS-95*, pp. 328–335, 1995.
21. A. Taylor and W. Zwicker. *Simple Games: Desirability Relations, Trading, Pseudoweightings*. Princeton University Press, 1999.
22. M. P. Wellman. The economic approach to artificial intelligence. *ACM Computing Surveys*, 27:360–362, 1995.
23. M. Yokoo, V. Conitzer, T. Sandholm, N. Ohta, and A. Iwasaki. Coalitional games in open anonymous environments. In *Proc. of AAAI-05*, pp. 509-514, 2005.