Technical Report

Theorem 1: Deciding whether a given number of ballot substitutions guarantees a fair outcome of a general instance of the 2 attribute case is weakly NP-Complete.

Proof: Given a solution that specifies the ballot substitutions in an instance of the 2 attribute case it is easy to check whether the solution satisfies the fairness conditions. To prove the hardness we reduce the weakly NP-Hard Partition problem to our problem.

The input for the Partition problem [1] consists of $2k$ positive integers $c_1, \ldots c_{2k}$ and the problem is to determine whether there exists a set $S \subset \{1, \ldots, 2k\}$, where $|S| = k$ such that $\sum_{i \in S} c_i = \sum_{i \notin S} c_i$. Assume without loss of generality that the sequence $c_1, \ldots c_{2k}$ is sorted in non-increasing order.

Given an instance of the Partition problem we define a corresponding instance of the margin problem with 2 attributes as follows. Define the attributes $A[i, j]$, for $i \in \{1, 2\}$ and $j \in [1..2k + 2]$. Assume that the number of candidates that need to be elected is $4k$. A fair outcome has to satisfy the following fairness constraints.

- For every $j \in [1..2k]$, there is exactly one candidate whose first attribute is $A[1, j]$
- For every $j \in [1..2k]$, there is exactly one candidate whose second attribute is $A[2, j]$
- There are k candidates whose first attribute is $A[1, 2k+1]$
- There are k candidates whose first attribute is $A[1, 2k+2]$
- There are k candidates whose second attribute is $A[2, 2k + 1]$
- There are k candidates whose second attribute is $A[2, 2k + 2]$

There are $8k$ candidates with the following attributes and number of votes.

- For $i \in [1..2k]$ there is a candidate with attributes $A[1, i],\ A[2, 2k + 1]$ who received $c_1 + c_i$ votes
- For $i \in [1..2k]$ there is a candidate with attributes $A[1, 2k + 1],\ A[2, i]$ who received $c_1 + 1 - c_i$ votes
- For $i \in [1..2k]$ there is a candidate with attributes $A[1, i],\ A[2, i]$ who received $c_1 + 1$ votes
- There are $2k$ candidates with attributes $A[1, 2k + 2],\ A[2, 2k + 2]$, each of whom received c_1 votes

We claim that the answer to the instance of the Partition problem is "yes" if and only if the corresponding instance of the margin problem can be solved with margin $k + \frac{1}{2}\sum_{i=1}^{2k} c_i$.

Suppose that the answer to the instance of the Partition problem is "yes". Then there exists a set S such that $\sum_{i \in S} c_i = \sum_{i \notin S} c_i$. The solution for the corresponding instance of the margin problem is defined as follows. For $i \in S$, c_i ballots of the candidate with attributes $A[1, i],\ A[2, 2k + 1]$ are removed, leaving this candidate with c_1 votes. For $i \notin S$, c_i ballots of the candidate with attributes $A[1, 2k + 1],\ A[2, i]$ are added, resulting in this candidate having $c_1 + 1$ votes. For $i \notin S$, a single ballot of the candidate with attributes $A[1, i],\ A[2, i]$ is removed, leaving this candidate with c_1 votes. Lastly, a single ballot is added to k candidates with attributes $A[1, 2k+2],\ A[2, 2k+2]$ resulting in each of these candidates having $c_1 + 1$ votes.

The following candidates have at least $c_1 + 1$ votes after the manipulations.

- For $i \notin S$, the candidate with attributes $A[1, i],\ A[2, 2k + 1]$
- For $i \notin S$, the candidate with attributes $A[1, 2k + 1],\ A[2, i]$
- For $i \in S$, the candidate with attributes $A[1, i],\ A[2, i]$
- k candidates with attributes $A[1, 2k + 2],\ A[2, 2k + 2]$

Since $|S| = k$ the number of these candidates is exactly $4k$ and they satisfy the fairness properties. The total number of ballot additions is $k + \sum_{i \notin S} c_i$. The total number of ballot removals is $k + \sum_{i \in S} c_i$. Both quantities are identical and thus the number of ballot substitutions is $k + \frac{1}{2}\sum_{i=1}^{2k} c_i$.

We now prove the opposite direction. Suppose that the corresponding instance of the margin problem can be solved with margin $k + \frac{1}{2}\sum_{i=1}^{2k} c_i$. We show that in this case the answer to the instance of the Partition problem is "yes". Consider the outcome after making the $k + \frac{1}{2}\sum_{i=1}^{2k} c_i$ ballot substitutions. Let q be the threshold, namely the number of votes such that each of the $4k$ elected candidates received at least q votes and there is at least one elected candidate who received q votes.

We prove that the threshold q must be $c + 1$. For this we prove that if $q \geq c + 1$ then $q = c + 1$. The proof that if $q \leq c + 1$ then also $q = c + 1$ is symmetric and thus omitted. From now on suppose that $q \geq c_1 + 1$.

A fair outcome has to contain exactly k candidates whose first attribute is $A[1, 2k + 1]$. The second attribute of such a candidate is $A[2, i]$, for $i \in [1..2k]$. We say that the *index* of this candidate is i. Since all the candidates whose first attribute is $A[1, 2k + 1]$ received less than q votes, a subset of size k of such candidates needs to be selected and their number of votes needs to be increased so that it is (at least) q. Let T be the index set of these elements; that is, for every $i \in T$, the candidate with attributes $A[1, 2k + 1],\ A[2, i]$ is elected.

A fair outcome has to contain exactly k candidates whose second attribute is $A[2, 2k + 1]$. The first attribute of such a candidate is $A[1, i]$, for $i \in [1..2k]$. Let T' be the index set of these candidates. We claim that $T \setminus T' = \emptyset$. Otherwise, there is an index $i \in T \setminus T'$. Since $i \notin T'$ the candidate with attributes $A[1, i],\ A[2, 2k+1]$ is not elected, but since a candidate whose first attribute is $A[1, i]$ must be elected, the only other candidate whose first attribute is $A[1, i]$ has to be elected. However, the second attribute of this candidate is $A[2, i]$, which implies that the outcome contains two elected candidates whose second attribute is $A[2,i]$, one is the candidate whose attributes are $A[1,2k{+}1],\ A[2,i]$ (since $i \in T$) and the other is the candidate whose attributes are $A[1,i],\ A[2,i]$. This violates the fairness of the outcome. It follows that $T \backslash T' = \emptyset$. Since $|T| = |T'| = k$ also $T' \setminus T = \emptyset$ and thus $T = T'$.

Let $\bar{T} = [1..2k] \setminus T$. Since $|T| = |\bar{T}| = k$ we can define an arbitrary one to one and onto mapping $\mathcal{M} : T \to \bar{T}$. Consider $i \in T$. The number of votes that need to be added to the candidate with attributes $A[1,2k+1],\ A[2,i]$ is (at least) $q-(c_1+1-c_i)$. Now consider $j = \mathcal{M}(i)$. If $q > c_1 + c_j$ then $q-(c_1+1-c_i) \geq c_1+c_j+1-(c_1+1-c_i) = c_i+c_j$. Otherwise, $q \leq c_1 + c_j$, and in this case since the candidate with attributes $A[1,j],\ A[2,2k+1]$ is not elected, at least $c_1 + c_j - (q-1)$ votes need to be removed from this candidate, implying that the total number of votes additions and removals in this case is at least $q - (c_1 + 1 - c_i) + c_1 + c_j - (q-1) = c_i + c_j$. Thus, in both cases the total number of votes manipulations is at least $c_i + c_j$. Summing over all $i \in T$ we get the number of vote additions and removals so far is at least $\sum_{i=1}^{2k} c_i$.

A fair outcome has to contain exactly k candidates whose first attribute is $A[1,2k+2]$, and exactly k candidates whose second attribute is $A[2,2k+2]$. Thus $q - c_1$ votes need to be added to k of the candidates with attributes $A[1,2k+2],\ A[2,2k+2]$. A fair outcome has to contain exactly one candidate whose first attribute is $A[1,i]$, and exactly one candidates whose second attribute is $A[2,i]$, for $i \in [1..2k]$. Recall that there are $2k$ candidates with attributes $A[1,i],\ A[2,i]$ who received $c_1 + 1$ votes, for $i \in [1..2k]$. If $q > c_i + 1$ then $q - c_1 - 1$ votes need to be added to the candidates with attributes $A[1,i],\ A[2,i]$, for $i \notin T$. If $q = c_i + 1$ then at least one vote needs to be removed from the candidates with attributes $A[1,i],\ A[2,i]$, for $i \in T$.

Suppose that $q > c + 1$. In this case we get that the total number of vote additions and removals is at least $k(q-c_1) + k(q-c_1-1) + \sum_{i=1}^{2k} c_i \geq 3k + \sum_{i=1}^{2k} c_i$. But in this case the number of ballot substitutions is at least $1.5k + \frac{1}{2}\sum_{i=1}^{2k} c_i$, which is too much. If $q = c_1 + 1$ the number vote additions and removals is $2k + \sum_{i=1}^{2k} c_i$. Also in this case c_i votes need to be removed from each of the k candidates with attributes $A[1,j],\ A[2,2k+1]$, for $i \notin T$. It follows that the number of vote additions is $k + \sum_{i\in T} c_i$ and the number of vote subtractions is $k + \sum_{i \notin T} c_i$. Since the margin is $k + \frac{1}{2}\sum_{i=1}^{2k} c_i$. We must have $\frac{1}{2}\sum_{i=1}^{2k} c_i \geq \sum_{i\in T} c_i$ and $\frac{1}{2}\sum_{i=1}^{2k} c_i \geq \sum_{i\notin T} c_i$. Thus, the answer to the Partition instance is "yes" since $\sum_{i\in T} c_i = \sum_{i \notin T} c_i = \frac{1}{2}\sum_{i=1}^{2k} c_i$. ■

Theorem 2: Deciding the feasibility of a general instance of the 3 attribute case (and thus any $d \geq 3$ attributes as well) is NP-Complete.

Proof: Given a solution that specifies the ballot substitutions in an instance of the 3 attribute case it is easy to check whether the solution satisfies the fairness conditions. To prove the hardness we reduce the 3-Dimensional Matching problem (3DM) to our problem, The input for the 3DM problem [1] consists of node sets X, Y, and Z, where $|X| = |Y| = |Z| = k$, and a set of hyper-edges E, where each edge has 3 coordinates (x,y,z), and $x \in X,\ y \in Y,\ z \in Z$. The problem is to determine whether there is a subset of hyper-edges $S \subseteq E$ such that $|S| = k$ and S touches every node, that is, $\forall x \in X\ \exists e \in S$ such that the first coordinate of e is x, and a similar condition applies $\forall y \in Y$ and $\forall z \in Z$.

Given an instance of the 3DM we define a corresponding instance of the margin problem with 3 attributes as follows. Define the attributes $A[i,j]$, for $i \in \{1,2,3\}$ and $j \in [1..k]$, where each $x \in X$ corresponds to a distinct first attribute value $A[1,i(x)]$, each $y \in Y$ corresponds to a distinct second attribute value $A[2,i(y)]$, and each $z \in Z$ corresponds to a distinct third attribute value $A[3,i(z)]$. For each $(x,y,z) \in E$ define a candidate with attributes $A[1,i(x)], A[2,i(y)], A[3,i(z)]$. The outcome has to contain k candidates with exactly one candidate with attribute value $A[i,j]$, for $i \in \{1,2,3\}$ and $j \in [1..k]$. It is easy to see that this problem has a solution for margin m if and only if the 3DM instance has a solution. (It is enough to set the margin to be m since with m ballot substitutions we can generate any election outcome since it allows us to change all ballots.) ■

References

[1] "Computers and intractability," Michael R Garey and David S Johnson. volume 174,freeman San Francisco, 1974.