Choosing Collectively Optimal Sets of Alternatives Based on the Condorcet Criterion

Edith Elkind
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore
eelkind@ntu.edu.sg

Jérôme Lang Abdallah Saffidine
LAMSADE
Univ. Paris-Dauphine, France
lang@lamsade.dauphine.fr
abdallah.saffidine@dauphine.fr

Abstract

In elections, an alternative is said to be a Condorcet winner if it is preferred to any other alternative by a majority of voters. While this is a very attractive solution concept, many elections do not have a Condorcet winner. In this paper, we propose a set-valued relaxation of this concept, which we call a *Condorcet winning set*: such sets consist of alternatives that *collectively* dominate any other alternative. We also consider a more general version of this concept, where instead of domination by a majority of voters we require domination by a given fraction θ of voters; we refer to this concept as θ-*winning set*. We explore social choice-theoretic and algorithmic aspects of these solution concepts, both theoretically and empirically.

1 Introduction

When a group of agents is trying to decide on a joint plan, it is often the case that no single alternative is consensual enough to be chosen as the collective decision. For instance, this is the case if consensual alternatives are identified with Condorcet winners, since some profiles do not have Condorcet winners. Similarly, in approval voting, we might view an alternative as consensual if it is approved by all voters: again, in many cases there will be no consensual alternative.

For instance, consider a common practical problem, namely, the choice of a time slot for a departamental seminar. Each faculty member approves of some of the time slots, and disapproves of others, based on their teaching schedule, so it is natural to make this choice using approval voting. Now, ideally, research seminars in a department should always be held on the same day of the week and at the same time. However, this requirement (R) will have the unfortunate consequence that some members of the department will always miss the research seminar, just because they teach every week at that time. In this case, rather than choosing the time slot acceptable to the highest number of voters, we might instead relax requirement (R) and allow *two* slots to be chosen, i.e., ask whether there is a set of two slots $\{s, s'\}$ such that every voter approves either s or s'. If such a pair exists, then by alternating between s and s', we can ensure that everyone can attend at least some seminars. More generally, we are interested in a set of slots of minimal cardinality such that every voter approves at least one slot in this set.

In this paper, we extend this approach to the more traditional model of voting, where voters submit rankings of candidates, and to a weaker notion of collective acceptability. That is, we consider the problem of finding a *set* of alternatives such that collectively the voters are happy enough with *at least one* alternative in the set. Sets of alternatives are thus considered *disjunctively*: a set of alternatives Y is deemed to satisfy property π if, in the profile obtained by (a) replacing the top alternative in Y in each vote by a new alternative *merge*(Y) (b) removing all alternatives in Y from each vote, the property π is satisfied by the new alternative *merge*(Y).

Applied to approval voting, and taking π to be "being approved by every voter", this approach reduces to finding a subset of alternatives Y such that every voter approves at least one alternative in Y, which is the well-known *hitting set* problem. If π relates to the Borda count or, more generally, some scoring function, then we recover some approaches to proportional representation (see Section 6). If we take π to be the Condorcet criterion, we get the following: Y is a *Condorcet winning set* if for any candidate z in $X \setminus Y$, a majority of agents prefer some candidate in Y to z; in particular, a Condorcet winner is a Condorcet winning set of size 1. This concept can be generalized by varying the quota: a set S is a θ-*winning set*, $\theta \in [0, 1)$, for an n-voter profile if for any alternative x not in S, more than θn voters prefer *some* alternative in S to x; Condorcet winning sets correspond to $\theta = \frac{1}{2}$.

In this paper we give a detailed study of θ-winning sets, and, in particular, Condorcet winning sets. In Section 2 we provide the main definitions. In Section 3 we relate our solution concept to standard tournament solution concepts, and briefly discuss its social choice-theoretic properties. In Section 4 we study the corresponding optimization problem, with a particular focus on its computational aspects. We complement our theoretical results with empirical analysis in Section 5. Section 6 provides an overview of related work.

2 Definitions and Basic Properties

Throughout the paper, we consider elections with a set of candidates (alternatives) $X = \{x_1, \ldots, x_m\}$ and n voters. Each voter i is associated with a linear order $\succ_i$ over X; the vector $\langle \succ_1, \ldots, \succ_n \rangle$ of all voters' preferences is called a *preference profile* and is usually denoted by P.

We say that an alternative $x \in X$ *beats* another alternative $y \in X$ in a pairwise election if the majority of voters prefer x to y; if exactly half of the voters prefer x to y, then x and y are said to be *tied*. A candidate is said to be a *Condorcet winner* if he wins all pairwise elections that he participates in. Clearly, each election has at most one Condorcet winner, but many elections have no Condorcet winners.

A *voting correspondence* is a mapping that given a preference profile P over X outputs a non-empty subset $X' \subset X$; the elements of X' are called the *election winners*.

We can now define the concepts that we intend to study.

Definition 1. *Consider an n-voter election over X with a preference profile $P = \langle \succ_1, \dots, \succ_n \rangle$. A set $Y \subseteq X$ is said to θ-*cover *an alternative $z \in X \setminus Y$ for $\theta \in [0, 1)$ if*

$$\#\{i \in N \mid \exists y \in Y \text{such that } y \succ_i z\} > \theta n.$$

For $\theta = 1$, we say that a set $Y \subseteq X$ 1-covers *an alternative $z \in X \setminus Y$ if $\#\{i \in N \mid \exists y \in Y \text{such that } y \succ_i z\} = n$. Given a $\theta \in [0, 1]$, we say that Y is a θ-*winning set *if Y θ-covers each alternative in $z \in X \setminus Y$.*

A $\frac{1}{2}$-winning set is also called a Condorcet winning set*.*

Definition 1 can be rephrased as follows. Given a preference profile $P = \langle \succ_1, \dots, \succ_n \rangle$ over X and a set $Y \subseteq X$, consider a preference profile $P_{merge}(Y) = \langle \succ'_1, \dots, \succ'_n \rangle$ over the set of candidates $(X \setminus Y) \cup \{merge(Y)\}$, where for each $i = 1, \dots, n$ the preference order $\succ'_i$ is obtained from $\succ_i$ by replacing the top candidate from Y with *merge*(Y) and removing all other candidates in Y. Then Y is a θ-winning set for $\theta < 1$ if and only if for each $x \in X \setminus Y$ more than θn voters prefer *merge*(Y) to x. For $\theta = 1$, this reformulation is particularly convenient: Y is a 1-winning set if and only if *merge*(Y) is ranked first by all voters.

Given a $\theta \in [0, 1]$ and a $k \in \mathbb{N}$, we will denote by $D(P, \theta, k)$ the set of all θ-winning sets of size at most k in a profile P; also, we set $D(P, \theta) = \cup_k D(P, \theta, k)$. Observe that the set family $D(P, \theta)$ is upwards-closed: if $X \in D(P, \theta)$ and $X \subseteq Y$, then $Y \in D(P, \theta)$.

Typically, we are interested in winning sets that are as small as possible and dominate all alternatives not in the set as strongly as possible, i.e., we aim to minimize k and to maximize θ. Thus, if we want to turn the mapping $D(P, \cdot, \cdot)$ into a voting correspondence, it is natural to do so by fixing one of its arguments and optimizing over the other one. That is, we can fix θ, let $k(P, \theta)$ be the smallest value of k such that $D(P, \theta, k) \neq \emptyset$, and output either a member of $D(P, \theta, k(P, \theta))$ (chosen according to a pre-specified tie-breaking rule) or the union of all sets in $D(P, \theta, k(P, \theta))$; the latter option is particularly appealing if $k(P, \theta) = 1$.

Optimizing over θ requires more care: while we can set $\theta(P, k) = \sup\{\theta \mid D(P, \theta, k) \neq \emptyset\}$, since the definition of a θ-winning set involves a strict inequality, we have $D(P, \theta(P, k), k) = \emptyset$. Consequently, the voting correspondence should output a set in $D(P, \theta(P, k) - \varepsilon, k)$ (or, just as in the previous case, the union of all sets in $D(P, \theta(P, k) - \varepsilon, k)$), where ε is a sufficiently small parameter; for instance, we can set $\varepsilon = \frac{1}{n}$, where n is the number of voters in P.

The following examples illustrate the notions we have just introduced.

$\theta = \frac{1}{2}, k = 1$: If P has a Condorcet winner c, then $D(P, \frac{1}{2}, 1) = \{\{c\}\}$, otherwise, $D(P, \frac{1}{2}, 1) = \emptyset$.

$\theta = 1$: Let *top*(P) be the set of all candidates ranked first by some voter. Then $D(P, 1) = \{S \subseteq X \mid top(P) \subseteq S\}$.

Interestingly, the voting correspondence that fixes $k = 1$ and optimizes over θ, i.e., maps P to $\cup D(P, \theta(P, 1) - \frac{1}{|P|}, 1)$ is well-known in the social choice literature under a different name: namely, it is simply the Maximin correspondence. We recall that under Maximin, a candidate's score is the number of votes he gets in his worst pairwise election, i.e., x's score equals $\min_{y \in X \setminus \{x\}} \#\{i \mid x \succ_i y\}$; the Maximin winners are the candidates with maximal score.

Proposition 1. *Given an n-voter preference profile P, the set $D_1(P) = \cup D(P, \theta(P, 1) - \frac{1}{n}, 1)$ is exactly the set of Maximin winners in P.*

In the rest of the paper, we will focus on the case $\theta = \frac{1}{2}$, i.e., the Condorcet winning sets. We will say that the *Condorcet dimension* $\dim_C(P)$ of a given profile P is the size of the smallest Condorcet winning set for P.

3 Social Choice-Theoretic Properties

In this section, we will discuss some social-choice theoretic properties of Condorcet winning sets.

3.1 Condorcet Winning Sets and Tournament Functions

Given a profile P, the *majority graph* $G(P)$ is the directed graph (also called a *weak tournament*) with the vertex set X that contains a directed edge from x to y if x beats y in a pairwise election. If the number of voters is odd, then for each pair (x, y) either (x, y) or (y, x) is present in the graph, i.e., $G(P)$ is a tournament. A *weighted majority graph* labels each edge (x, y) with the number of voters that prefer x to y. A *(weighted) tournament function* is a mapping F defined on preference profiles that satisfies $F(P) = F(Q)$ whenever P and Q have the same (weighted) majority graph (see, *e.g.*, [Laslier, 1997]). It turns out that the mapping that outputs all Condorcet winning sets for a given profile is *not* a (weighted) tournament function.

Proposition 2. *The mapping $P \mapsto D(P, \frac{1}{2})$ is not a (weighted) tournament function.*

Proof. Consider the following 4-candidate, 3-voter profiles $P = \langle abcd, cdab, dabc \rangle$ and $Q = \langle acdb, bcda, dabc \rangle$. The weighted majority graphs for P and Q coincide: a beats b and c, b beats c, c beats d, d beats a and b. However, $\{a, b\}$ is a Condorcet winning set for Q, but not for P. □

There is, however, a certain similarity between Condorcet winning sets and a classic tournament function, namely, the dominating sets. Recall that a set of vertices S of a directed graph $G = (N, A)$ is called a *dominating set* if for every vertex $x \in N \setminus S$ there exists a vertex $y \in S$ such that the directed edge (y, x) is in A. Now, clearly, if S is a dominating set for $G(P)$, then S is a Condorcet winning set for P. However, the proof of Proposition 2 shows that the converse is not true. Another example, which illustrates that a minimal Condorcet winning set for P can be very different from any dominating set of $G(P)$, is provided by a preference profile P in which all pairwise elections end in a tie, i.e., $G(P)$ consists of $|X|$ isolated vertices (for instance, a profile that contains a single copy of each of the possible $|X|!$ orderings of the candidates has this property). Clearly, the only dominating set for $G(P)$ is X. In contrast, it is not hard to see that P has a Condorcet winning set of size 2.

Proposition 3. *If P is an n-voter preference profile over X in which all pairwise elections are tied, then* $\dim_C(P) = 2$.

Proposition 3 shows that if we are looking for a small Condorcet winning set, simply outputting the smallest dominating set results in a very poor approximation factor. However, this observation crucially relies on allowing ties among the candidates. In Section 4, we will show that if there are no ties (this happens, in particular, if the number of voters is odd), then the dominating set-based approach works quite well.

3.2 Properties of Condorcet Winning Sets

Condorcet winning sets have several desirable social choice properties. For instance, it is easy to see that they satisfy *monotonicity* (we omit the proof due to lack of space).

Proposition 4. *If an alternative x belongs to some minimal Condorcet winning set, and some voter moves it up in her preference ranking without changing the relative order of the remaining alternatives, then x is contained in some minimal Condorcet winning set in the resulting profile.*

Further, Condorcet winning sets satisfy a weak form of *consistency*, a.k.a. *reinforcement* (again we omit the proof).

Proposition 5. *Let P_1 and P_2 be two profiles over a set of candidates X. If Y is a Condorcet winning set for both P_1 and P_2, then it is also a Condorcet winning set for $P_1 \cup P_2$.*

However, *minimal* Condorcet winning sets do not satisfy reinforcement: Let $P_1 = \langle abc, abc, bca, bca, cab, cab\rangle$ and $P_2 = \langle acb, acb, bac, bac, cba, cba, cba\rangle$. Then $\{a, b\}$ is a minimal Condorcet winning set for P_1 and P_2, but not for $P_1 \cup P_2$, which has c as its Condorcet winner. Also, minimal Condorcet winning sets do not satisfy *efficiency*: there exist profiles for which some minimal Condorcet winning set contains a Pareto-dominated alternative. For instance, for $P = \langle abcd, cdab, dabc\rangle$, the set $\{b, d\}$ is a minimal Condorcet winning set, although b is dominated by a.

4 Computing Condorcet Winning Sets

In this section, we focus on the problem of finding small Condorcet winning sets. Since we limit ourselves to the case of $\theta = \frac{1}{2}$, we use the term "cover" instead of "$\frac{1}{2}$-cover" throughout this section. Also, we assume that the number of voters is odd and hence no pairwise election ends in a tie.

The first issue we consider, which turns out to be surprisingly difficult, is constructing a profile with high Condorcet dimension. To start, observe that the Condorcet dimension of a given profile P is 1 if and only if P has a Condorcet winner. Thus, it is easy to construct a profile whose dimension exceeds 1: consider, for instance, a *Condorcet cycle* of size m, $m \geq 3$, i.e., an m-voter, m-candidate preference profile, where the i-th voter places the j-th candidate in position $i + j - 1 \mod m$ (e.g., for $m = 3$ the Condorcet cycle over the candidate set $X = \{1, 2, 3\}$ is given by $P = \langle 123, 231, 312\rangle$). P has no Condorcet winner, so $\dim_C(P) \geq 2$. In fact, it is not hard to see that $\dim_C(P) = 2$: indeed, candidates 1 and $\lceil m/2\rceil$ form a dominating set in $G(P)$ (and hence a Condorcet winning set).

To exhibit a profile of Condorcet dimension 3. we borrow some terminology from linear algebra. Let $A = (a_{ij})$ be a p-by-q matrix and let $B = (b_{k\ell})$ be a p'-by-q' matrix. Recall that the *Kronecker product* of A and B is a pp'-by-qq' matrix $A \otimes B$ of the form

$$\begin{pmatrix} a_{11}B & \dots & a_{1q}B \\ \dots & \dots & \dots \\ a_{p1}B & \dots & a_{pq}B \end{pmatrix}.$$

Now, any n-voter, m-candidate preference profile P can be associated with an m-by-n matrix $M(P)$: the i-th entry in the j-th column of $M(P)$ is the name of the candidate that is ranked in the i-th position by the j-th voter. Thus, given a preference profile P over a set of candidates X and a preference profile Q over a set of candidates Y, we can define $P \otimes Q$ as a preference profile over the set of candidates $X \times Y$ (i.e., pairs of the form (x_i, y_j), where $x_i \in X$, $y_j \in Y$) that corresponds to the matrix $M(P) \otimes M(Q)$, where we identify the product $x_i y_j$ with the pair (x_i, y_j).

For instance, if $P = \langle 321, 231\rangle$ is a preference profile over $X = \{1, 2, 3\}$ and $Q = \langle 12, 21\rangle$ is a preference profile over $Y = \{1, 2\}$, then the first voter in $P \otimes Q$ has the preference ordering $(3, 1)(3, 2)(2, 1)(2, 2)(1, 1)(1, 2)$ over the candidate set $\{(i, j) \mid i = 1, 2, 3, j = 1, 2\}$.

We are now ready to present our example. Let P be the Condorcet cycle over $X = \{0, 1, 2\}$ and let Q be the Condorcet cycle over $Y = \{0, 1, 2, 3, 4\}$; the preference profile $P \otimes Q$ over $Z = X \times Y$ is given in the table below, where we identify the element (i, j) with $5i + j + 1$.

v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	3	4	5	1	7	8	9	10	6	12	13	14	15	11
3	4	5	1	2	8	9	10	6	7	13	14	15	11	12
4	5	1	2	3	9	10	6	7	8	14	15	11	12	13
5	1	2	3	4	10	6	7	8	9	15	11	12	13	14
6	7	8	9	10	11	12	13	14	15	1	2	3	4	5
7	8	9	10	6	12	13	14	15	11	2	3	4	5	1
8	9	10	6	7	13	14	15	11	12	3	4	5	1	2
9	10	6	7	8	14	15	11	12	13	4	5	1	2	3
10	6	7	8	9	15	11	12	13	14	5	1	2	3	4
11	12	13	14	15	1	2	3	4	5	6	7	8	9	10
12	13	14	15	11	2	3	4	5	1	7	8	9	10	6
13	14	15	11	12	3	4	5	1	2	8	9	10	6	7
14	15	11	12	13	4	5	1	2	3	9	10	6	7	8
15	11	12	13	14	5	1	2	3	4	10	6	7	8	9

Proposition 6. $\dim_C(P \otimes Q) = 3$.

Proof. It is immediate that $P \otimes Q$ does not have a Condorcet winner. Now, suppose that S is a Condorcet winning set of size 2 for $P \otimes Q$, and set $Z_i = \{(i, j)\}$ for $i = 0, 1, 2$. Assume first that $S \subseteq Z_i$ for some $i = 0, 1, 2$; by symmetry we can assume that $S \subseteq Z_0$. Then the candidates in Z_2 are not covered, a contradiction. Thus, we have $|S \cap Z_i| = 1$, $|S \cap Z_j| = 1$ for some $i \neq j$; again, by symmetry we can assume that $|S \cap Z_0| = 1$, $|S \cap Z_1| = 1$, i.e., $S \cap Z_1 = \{(0, k)\}$

$S \cap Z_1 = \{(1, \ell)\}$ for some $k, \ell = 0, \ldots, 4$. Now, consider the candidate $(0, k')$, where $k' = k - 1 \mod 5$. This candidate is ranked below $(0, k)$ in 3 votes and below $(1, \ell)$ in 5 votes; however, there is 1 vote where $(0, k')$ is ranked below both $(0, k)$ and $(1, \ell)$, so altogether $(0, k')$ is ranked below $(0, k)$ or $(1, \ell)$ in 7 votes, i.e., $(0, k')$ is not covered.

Finally, it is easy to see that $P \otimes Q$ has a Condorcet winning set of size 3: for instance, take any set S such that $S \cap Z_i \neq \emptyset$ for $i = 0, 1, 2$ or, e.g., the set $\{(0, 0), (0, 2), (1, 0)\}$. ☐

We do not know whether there exists a profile of dimension 3 with less than 15 candidates; finding a profile with the smallest number of candidates that admits a profile of dimension 3 or more is an interesting research question.

One might think that by taking a Kronecker product of s sufficiently long Condorcet cycles, we always get a profile of Condorcet dimension $s + 1$. However, it turns out that this is not the case: a product of a Condorcet cycle and *any* profile has Condorcet dimension that does not exceed 3.

Proposition 7. *Let P be a Condorcet cycle over a set of candidates $X = \{x_1, \ldots, x_m\}$, $m \geq 2$, and let Q be some other profile over a set of candidates $Y = \{y_1, \ldots, y_k\}$. Then $\dim_C(P \otimes Q) \leq 3$.*

Proof. The set $\{(x_1, y_1), (x_i, y_1), (x_m, y_1)\}$, where $i = \lfloor \frac{m}{2} \rfloor + 1$, is a Condorcet winning set for $P \otimes Q$. Indeed, consider any $z = (x_k, y_\ell) \in X \times Y$. If $1 < k \leq i$, then z is covered by (x_1, y_1), if $i < k \leq m$, then z is covered by (x_i, y_1), and if $k = 1$, then z is covered by (x_m, y_1). ☐

In fact, we do not have any examples of preference profiles of Condorcet dimension 4 or higher: while we do believe that profiles of arbitrarily high dimension exist, we were not able to construct them. Finding such profiles is perhaps the most important open question suggested by this work.

One could try to solve this problem using the probabilistic method [Alon and Spencer, 1992], i.e., generate a preference profile randomly and argue that it has high Condorcet dimension with non-zero probability. However, the following proposition shows that a direct application of the probabilistic method fails in our case.

Proposition 8. *Let P be an n-voter preference profile over a set of candidates X, $|X| = m$, obtained by drawing each of the n votes uniformly at random from all permutations of X, and let $S = \{a, b\}$ be an arbitrary 2-element subset of X. With probability at least $1 - me^{-n/24}$ the set S is a Condorcet winning set for P.*

Proof. Consider an arbitrary candidate $c \in X \setminus \{a, b\}$. Any given vote is equally likely to contain any of the six possible permutations of a, b, and c. Therefore, with probability $\frac{2}{3}$ in any given vote either a or b is ranked above c. Hence, the expected number of votes where a or b beats c is $\frac{2n}{3}$. By Chernoff bound (see, e.g., [Alon and Spencer, 1992]), the probability that c is ranked above a and b in at least $\frac{n}{2}$ votes is at most $e^{-n/24}$. Thus, by the union bound, the probability that $\{a, b\}$ is not a Condorcet winning set is at most $me^{-n/24}$. ☐

Thus, for a random profile, every pair of alternatives is likely to be a Condorcet winning set; *a fortiori*, with high probability every alternative is contained in some minimal Condorcet winning set; this includes Pareto-dominated alternatives and Condorcet losers[1]. Therefore, the voting correspondence that returns all alternatives contained in a minimal Condorcet winning set is not particularly attractive; we will return to this issue at the end of Section 5.

The probability distribution used in the proof of Proposition 8 is known as the *impartial culture*, and is standard in the probabilistic analysis of elections. Thus, Proposition 8 shows that when the number of voters is large, under the impartial culture assumption a random preference profile has Condorcet dimension 2 or lower with very high probability; this is also confirmed by our empirical analysis (see Section 5). In fact, applying the union bound once again, we see that with probability at least $1 - m^3 e^{-n/24}$ *any* set of size 2 is a Condorcet winning set. Note also that our argument can be generalized to θ-winning sets, giving us the following corollary.

Corollary 1. *Let P be a random n-voter preference profile over a set X, $|X| = m$, generated under the impartial culture assumption, and let S be a k-element subset of X. Then S is a $(\frac{k}{k+1} - \varepsilon)$-winning set for P with probability at least $1 - m \cdot \exp(-\frac{k\varepsilon^2}{(k+1)n})$.*

To upper-bound the Condorcet dimension of a given profile, we will exploit the connection between Condorcet winning sets and dominating sets in tournaments that was established in the previous section. [Megiddo and Vishkin, 1988] show that any tournament on m vertices has a dominating set of size $\lceil \log_2 m \rceil$: the proof proceeds by selecting the vertex with the highest outdegree, which by the pigeonhole principle dominates at least half of the other vertices, deleting this vertex and all vertices dominated by it, and recursively applying the same procedure to the remaining graph. This implies the following result.

Proposition 9. *For any profile P with $n = 2r + 1$ voters and m candidates, we have $\dim_C(P) \leq \lceil \log_2 m \rceil$.*

From the algorithmic perspective, it is natural to ask if we can efficiently compute the Condorcet dimension of a given profile; we will refer to this problem as CONDORCET DIMENSION. Now, if, contrary to our hypothesis, there exists a constant K such that $\dim_C(P) \leq K$ for every profile P, CONDORCET DIMENSION can be solved in polynomial time by direct enumeration. However, even if this is not the case, CONDORCET DIMENSION is nevertheless unlikely to be too hard for profiles with an odd number of voters. Indeed, by Proposition 9, if no pairwise election is tied, we can find a Condorcet winning set by enumerating all subsets of candidates of size $\log m$. Thus, CONDORCET DIMENSION can be solved in time $\mathrm{poly}(n, m) m^{\log m}$, i.e., it is in the class QP of *quasi-polynomial* problems, exactly like the minimum dominating set problem [Megiddo and Vishkin, 1988]. It is strongly believed that QP is strictly contained in NP. Therefore, for elections with an odd number of candidates CONDORCET DIMENSION is unlikely to be NP-complete[2].

[1] We thank an anonymous reviewer for this remark.

[2] We thank Bruno Escoffier for pointing this out.

Observe also that Proposition 8 shows that our problem admits an algorithm whose running time is polynomial *in expectation* under the impartial culture assumption when n is sufficiently large relative to m. Indeed, given a profile P, we pick an arbitrary set of candidates of size 2, and check if it is a Condorcet winning set for P. If this is not the case, we perform the same check for all 2^m subsets of candidates. The probability that we execute the second stage is at most $me^{-n/24}$, so the expected running time of this algorithm is polynomial in n and m as long as $n \geq 24m$. In fact, if the number of voters is odd, we can restrict ourselves to sets of size at most $\log m$ during the second stage of the algorithm.

We can use the same technique to find θ-winning sets for $\theta \neq \frac{1}{2}$. However, as θ increases, we may have to search larger sets during the first step of the algorithm; indeed, Corollary 1 suggests that one should search all sets of size $\lceil \frac{\theta}{1-\theta} \rceil$.

5 Empirical Analysis

In this section, we provide an empirical analysis of θ-winning sets under the impartial culture assumption.

In our first experiment, we generate r preference profiles with m voters and n candidates, for various values of r, m, and n, and compute the Condorcet dimension of each profile. In *all* of our experiments, the Condorcet dimension was either 1 or 2. While this is consistent with Proposition 8, it is remarkable that not even a single profile of a higher dimension has been discovered. The results of our experiments are summarized in Table 1. Note that the probability that the Condorcet dimension of a given profile exceeds 1 is simply the probability of the Condorcet paradox, which has been extensively studied in the social choice literature (see, e.g., [Jones *et al.*, 1995]). Our results are consistent with those results, lending credence to our implementation. Observe also the drastic difference between the columns that correspond to odd and even values of n, which is especially pronounced for larger values of m.

	$\dim_C$	$n=10$	$n=11$	$n=20$	$n=21$	$n=100$	$n=101$
$m=15$,	1	14.5	43.4	19.0	41.3	28.0	39.6
$r=10^6$	2	85.5	56.6	81.0	58.7	72.0	60.4
$m=50$,	1	4.3	19.7	6.0	17.6	10.0	16.0
$r=10^6$	2	95.7	80.3	94.0	82.4	90.0	84.0
$m=100$,	1	2.2	11.9	3.1	10.2	5.3	9.1
$r=10^5$	2	97.8	88.1	96.9	89.8	94.7	90.9

Table 1: Proportion of profiles with Condorcet dimension ≤ 2

Figure 1 maps the probability that a fixed set of candidates of size k, $k = 1, 2, 3, 4$, is a Condorcet winning set for a 30-candidate election, as a function of the number of voters. We see that for $k > 1$ this probability quickly reaches 1.

Figure 2 plots the empirical distribution of $\theta(k)$ for $m = 30$, $n = 100$ and $k = 1, 2, 3, 4$. Our results agree with Corollary 1: the values of $\theta(k)$ are clustered around $\frac{k}{k+1}$.

Figure 3 shows the distribution of the number of $\frac{2}{3}$-winning sets of size 2 for 20 candidates and a varying number of voters. We see that out of $(19 \cdot 20)/2$ pairs of candidates, with a high probability, only a few of them are $\frac{2}{3}$-winning sets, with a peak at 1. Therefore, using $\theta = \frac{2}{3}$, we are likely to get an output consisting of a unique winning set of size 2. For other values of k, the situation is similar; we simply need to replace $\frac{2}{3}$ by $\frac{k}{k+1}$. Therefore, if we search for a Condorcet set of minimal size k, and among those, one that maximizes $\theta(P, k)$, then, *in most cases*, $k = 2$ and $\theta = \frac{2}{3}$ will give us a solution. This analysis suggests an appealing way of turning θ-winning sets into a voting correspondence (though this is not the primary focus of our work): if we list all candidates contained in some smallest $\frac{2}{3}$-winning set, with high probability, we output a set of size 2.

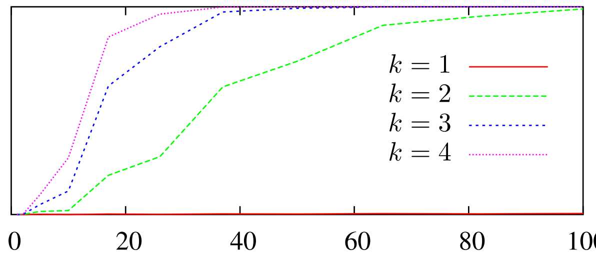

Figure 1: Probability that a fixed set of size k is a Condorcet winning set as a function of n

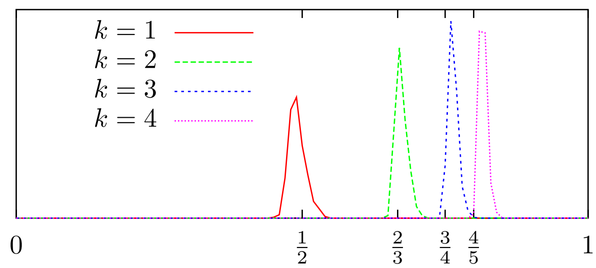

Figure 2: Empirical distribution of $\theta(k)$

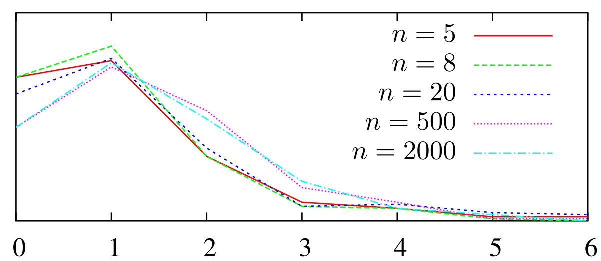

Figure 3: Empirical distribution of the number of $\frac{2}{3}$-winning sets of size 2

6 Related Work

There are two streams of research that are closely related to the problem we consider. The first stream deals with the problem of *proportional representation*, where each voter specifies a subset of candidates that would represent her, and the goal is to build a committee of a given cardinality that represents every voter. There are approaches to this problem that select a subset of candidates in a disjunctive manner. Specifically, [Chamberlin and Courant, 1983] propose a method that chooses the highest-ranking alternative from the given set in each vote, but uses the Borda score as a basis: a set Y receives $\max_{y\in Y} s_B(y,i)$ points from a voter i, where $s_B(y,i)$ is the number of candidates that i ranks below y, and the winning committee of size k is the k-element set of candidates with the highest score. Note that, unlike in our work, the committee size k is determined exogenously. This method is discussed in several subsequent papers, such as [Monroe, 1995]. [Procaccia *et al.*, 2008] show that computing a winning commitee of size k is NP-hard. A relaxation of this approach which allows for tradeoffs between committee size and quality of representation is proposed by [Lu and Boutilier, 2010], who also address the computation of optimal sets and show experimental results on real-world data sets.

A second stream of research has the same starting point as we do, namely, finding a generalization of the notion of a Condorcet winner to committees, or, more generally, sets of alternatives. There are two possible ways to do so: [Gerhrlein, 1985] assumes, like us, that individual preferences are given by rankings of the alternatives, and defines a *Condorcet committee* as a set of alternatives $Y \subseteq X$ such that every alternative in Y beats every alternative in $X \setminus Y$. This method has a *conjunctive* interpretation, as opposed to the disjunctive interpretation of Condorcet winning sets: Y is a Condorcet committee if for every alternative $x \in X\setminus Y$, and for *every* alternative $y \in Y$, a majority of voters prefers alternative y to x. Thus, a Condorcet committee is a Condorcet winning set, but the converse is not necessarily true.

As Condorcet committees of size k do not always exist, except for the trivial case $k = m$, [Ratliff, 2003] suggests to generalize the Dodgson and Kemeny voting rules to sets of alternatives. For instance, the extension of Dodgson computes, for a given subset Y, the minimal number of elementary swaps in the votes needed to make Y a Condorcet committee. The second approach, taken by [Fishburn, 1981] proceeds by defining a preference relation on *sets* of alternatives and looks for a subset that beats any other subset in a pairwise election. The two ways are bridged by [Kaymak and Sanver, 2003], who examine under which condition a Condorcet committee in the sense of Fishburn can be derived from preferences over single alternatives. [Ratliff, 2003] gives further discussion on committee selection. It is not immediately clear whether a Condorcet committee in the sense of [Fishburn, 1981] and [Kaymak and Sanver, 2003] is also a Condorcet winning set: the answer depends on the extension function used for lifting the preference relation from single alternatives to subsets; for standard extension functions (*e.g.*, ones that compare the sets lexicographically or according to their best element) this is not the case.

In a sense, we reconciliate both approaches discussed above: Condorcet winning sets treat the alternatives disjunctively, yet satisfy the Condorcet criterion for sets of alternatives of size 1.

7 Conclusions

We have defined a framework for defining collectively preferred sets of alternatives that is based on generalizing the Condorcet principle. Such a framework is useful whenever it makes sense to output several alternatives, such as committee elections, multiple recommendations to groups of users (as in [Lu and Boutilier, 2010]), choices of time slots for regular events, etc. A number of questions remain open for future work; perhaps the most pressing of them is whether there exist profiles of Condorcet dimension 4 or higher.

Acknowledgements

We thank Bruno Escoffier, Remzi Sanver, and the anonymous reviewers for their helpful comments. Jérôme Lang thanks the ANR project ComSoc (ANR-09-BLAN-0305). Edith Elkind was supported by NRF (Singapore) under grant RF2009-08 and NTU start-up grant.

References

[Alon and Spencer, 1992] N. Alon and J. Spencer. *Probabilistic method*. John Wiley, 1992.

[Chamberlin and Courant, 1983] J. R. Chamberlin and P. N. Courant. Representative deliberations and representative decisions: proportional representation and the Borda rule. *American Political Science Review*, 77(3):718–733, 1983.

[Fishburn, 1981] P. Fishburn. An analysis of simple voting systems for electing committees. *SIAM Journal on Applied Mathematics*, 41(3):499–502, 1981.

[Gerhrlein, 1985] W. Gerhrlein. The Condorcet criterion and committee selection. *Math. Social Sciences*, 10(3):199–209, 1985.

[Jones *et al.*, 1995] B. Jones, B. Radcliff, C. Taber, and R. Timpone. Condorcet winners and the paradox of voting: probability calculations for weak preference orders. *American Political Science Review*, 89(1):137–144, 1995.

[Kaymak and Sanver, 2003] B. Kaymak and R. Sanver. Sets of alternatives as Condorcet winners. *Social Choice and Welfare*, 20(3):477–494, 2003.

[Laslier, 1997] J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer-Verlag, 1997.

[Lu and Boutilier, 2010] T. Lu and C. Boutilier. Budgeted social choice: A framework for multiple recommendations in consensus decision making. In *Proc. of COMSOC'10*, 2010.

[Megiddo and Vishkin, 1988] N. Megiddo and U. Vishkin. On finding a minimum dominating set in a tournament. *Theoretical Computer Science*, 61:307–316, 1988.

[Monroe, 1995] B. L. Monroe. Fully proportional representation. *American Political Science Review*, 89:925–940, 1995.

[Procaccia *et al.*, 2008] A. Procaccia, J. Rosenschein, and A. Zohar. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3):353–362, 2008.

[Ratliff, 2003] T. Ratliff. Some startling inconsistencies when electing committees. *Social Choice and Welfare*, 21(3):433–454, 2003.