Bounded single-peaked width and proportional representation[1]

Denis Cornaz, Lucie Galand and Olivier Spanjaard

Abstract

This paper is devoted to the proportional representation (PR) problem when the preferences are clustered single-peaked. PR is a "multi-winner" election problem, that we study in Chamberlin and Courant's scheme [6]. We define clustered single-peakedness as a form of single-peakedness with respect to clusters of candidates, i.e. subsets of candidates that are consecutive (in arbitrary order) in the preferences of all voters. We show that the PR problem becomes polynomial when the size of the largest cluster of candidates (width) is bounded. Furthermore, we establish the polynomiality of determining the single-peaked width of a preference profile (minimum width for a partition of candidates into clusters compatible with clustered single-peakedness) when the preferences are narcissistic (i.e., every candidate is the most preferred one for some voter).

1 Introduction

Social choice theory deals with making collective choices on the basis of the individual preference relations of a set of agents (or voters) over a set of alternatives (or candidates). In this field, an active stream of research deals with "multi-winner" elections, where one aims at electing a subset of candidates rather than a single candidate. This occurs for instance when electing an assembly. In such situation, a combinatorial difficulty arises: while there are only m possible outputs of a single-winner election with m candidates, there are $\binom{m}{\kappa}$ possible assemblies of κ representatives. This difficulty is often overcome by organizing κ single-winner elections over κ subelectorates. With this way of partitioning the election, it may nevertheless happen that the elected assembly fails to represent minorities [4]: assume that the representatives of a party are in second position for the κ single-winner elections, then the party will have no representative in the assembly. Proportional representation aims at tackling this issue by performing a single multi-winner election ensuring that collectively the voters are satisfied enough by at least one elected candidate. This can be achieved for instance by using Chamberlin and Courant's scheme [6], where one elects a subset of κ candidates minimizing a *misrepresentation* score. The effective computation of such winning subsets of candidates has been studied by several authors.

Procaccia et al. have shown that the problem is NP-hard in the general case, but polynomial for a fixed κ [12]. Lu and Boutilier provided a polynomial approximation algorithm with performance guarantee (for maximizing a *representation* score), and show, on different experimental datasets, that it almost always returns an optimal solution [10]. Their setting is nevertheless different from proportional representation in political science: they aim at designing a system able to recommend a set of options to a group, based on the individual preferences of its members. Such a system could be used for instance by a conference organizer wishing to select a subset of sushis for the gala dinner, based on the individual preferences of the participants over the varieties of sushis. Clearly, this context authorizes suboptimality. Coming back to voting procedures, it is nevertheless important to note that the scores only provide an ordinal information: if an assembly A has a misrepresentation score 1 while an assembly B has a misrepresentation score $1 + \varepsilon$, one can only conclude that A is better than B, and not that B is close to be as good as A. Furthermore, in a political setting, it is simply not possible to elect an assembly without guaranteeing that it is the true winner. To our

---

[1] This paper appeared in the proceedings of ECAI 2012.

knowledge, the only general exact approaches proposed for proportional representation in Chamberlin and Courant's scheme are based on integer programming as the ones by Potthoff and Brams [11] and by Balinski [1] (in this latter reference, the formulation was actually proposed for the κ-median location problem, which is equivalent to the proportional representation problem [12]). The solution of these IP formulations might of course take exponential time in the worst case.

Very recently, Betzler et al. proposed an extensive investigation of parameterized complexity results for the problem [4]. Besides, they established that the problem becomes polynomial when the preferences are *single-peaked* [5]. Single-peakedness is the most popular domain restriction in social choice theory. In single-winner elections, it makes it possible to overcome Arrow's impossibility theorem (that states that no voting rule can simultaneously fulfill a set of basic axioms). In particular, there always exists a Condorcet winner (i.e., a candidate who is preferred to any other candidate by a majority of voters) if preferences are single-peaked. Such preferences are typically encountered in political science. Intuitively, preferences are single-peaked when 1) all voters agree on a left-right axis on the candidates reflecting their political convictions, and 2) the preferences of all voters decrease along the axis when moving away from their preferred candidate to the right or left. Nevertheless, this condition on preferences can be a bit restrictive when several candidates share similar opinions (e.g. they belong to the same party) since it is unlikely that the preferences of all voters are single-peaked on this subset of candidates.

We therefore study a new domain restriction, *clustered single-peakedness*, where single-peakedness holds on subsets of candidates (parties or more generally clusters), and not within clusters. The candidates belonging to the same cluster are ranked consecutively in the preferences of all voters, though *not necessarily in the same order*. Given a partition of the candidates into clusters such that the preferences are clustered single-peaked, the *width* of the partition is the size of the largest cluster minus one. Note that, for a given set of individual preference relations, several partitions into clusters can be compatible with clustered single-peakedness: we call *single-peaked width* the minimum width among all possible partitions of candidates into clusters. We show that the single-peaked width is computable in polynomial time if preferences are narcissistic, and that a bounded single-peaked width makes it possible to design a polynomial time solution algorithm for the proportional representation problem. Note that the same structures have been studied by Elkind et al. [7], under another terminology (in particular, clusters are called clone sets). Their main concern is not to study how clustered single-peakedness can be used to determine the winner of an election, but they show interesting connections with PQ-trees, and use them to design an algorithm to compute a partition of the candidates into (as many as possible) clusters. The links between their work and ours will be detailed in Section 4.

The paper is organized as follows. We first formally introduce the proportional representation problem and clustered single-peakedness (Sect. 2). Then we present a dynamic programming procedure for solving the proportional representation problem when the preferences are clustered single-peaked (Sect. 3). A key parameter for the efficiency of the procedure is the width of the partition into clusters. We therefore study the complexity of determining the single-peaked width of a set of individual preference relations (Sect. 4), and show the polynomiality of the problem for *narcissistic* preferences.

2 Preliminaries

2.1 Proportional representation

Let V be a set of n voters and C a set of m candidates. Let $\mathcal{P}$ be an $m \times n$ *preference profile matrix* over C, that is, each candidate appears exactly once in each column. So the set of columns of $\mathcal{P}$ is the set V and each column v is the preference relation of voter v. We denote by $r(v, c)$ the rank of candidate c in the preferences of voter v, and by $x \prec_v y$ the preference for y over x. A non-decreasing misrepresentation function $\mu : \{1, \ldots, m\} \to \mathbb{N}$ is defined such that

$\mu(r(v, c))$ is the misrepresentation value of c for v. The proportional representation problem aims at determining a subset $S \subseteq C$ of κ candidates such that the total misrepresentation score is minimized. In Chamberlin and Courant's scheme, the scoring function $s : 2^C \rightarrow \mathbb{N}$ is defined as follows:

$$s(S) = \sum_{v \in V} \min_{c \in S} \mu(r(v, c))$$

The proportional representation problem can then be simply written: $\min_{|S|=\kappa} s(S)$. The following example illustrates the value of using Chamberlin and Courant's scheme.

Example 1. *Consider a proportional representation problem with 6 voters $1, 2, 3, 4, 5, 6$ (indices of the columns) and 4 candidates a, b, c, d, and the following preference profile matrix:*

$$\mathcal{P} = \begin{pmatrix} a & c & a & c & d & c \\ b & b & b & a & c & d \\ c & a & c & b & b & a \\ d & d & d & d & a & b \end{pmatrix}$$

Assume that the misrepresentation function is $\mu(r) = r - 1$. If $\kappa = 2$, then the possible subsets and scores are (for simplicity ab stands for $\{a, b\}$):

ab	ac	ad	bc	bd	cd
6	1	4	3	6	4

The optimal solution is subset ac with score 1. With such a solution, only one voter is not represented by her preferred candidate (but by her second choice). Assume now this multi-winner election is divided into two single-winner elections, namely an election L between b and c for voters $1, 2, 3$, and an election R between a and d for voters $4, 5, 6$. The winner of election L (resp. R) is b (resp. d). Consequently, the winning solution is bd, which is the worst one according to the misrepresentation scores!

2.2 Clustered single-peakedness

Definition 1. *Let $\mathcal{C} = (C_1, \ldots, C_q)$ be an ordered partition of C into q non-empty subsets (called clusters). Preference profile matrix $\mathcal{P}$ is* clustered single-peaked *with respect to $\mathcal{C}$ if for all $v \in V$ there exists an index p in $\{1, \ldots, q\}$ such that:*

$$i < j < p \quad \Rightarrow \quad x \prec_v y \prec_v z$$
$$p < j < i \quad \Rightarrow \quad x \prec_v y \prec_v z$$

for all $x \in C_i$, $y \in C_j$ and $z \in C_p$.

For a voter v, we call C_p the *peak* of v, which means that any candidate in C_p is preferred to any candidate in $C \setminus C_p$. This definition coincides with usual single-peakedness when $|C_i| = 1$ for all i. The only candidate in C_p is then the most preferred one.

Example 2. *Coming back to Example 1, it can be easily seen that the preferences are not single-peaked w.r.t. axis (a, b, c, d), by considering Figure 1 where each curve represents a preference ranking of a voter, namely voters 1, 2, 6. For each curve and each candidate on the X-axis, the value on the Y-axis is the rank in the corresponding preference ranking (the better the rank the higher the point). Preferences are single-peaked w.r.t. an X-axis iff all curves have a single peak. This is not the case in the left graph since the curve of voter 6 (in bold) spikes down for b and then spikes up for a. More generally, it can be shown that the preferences in Example 1 are not single-peaked, whatever permutation of candidates on the X-axis is considered. However, the preferences are clustered single-peaked with respect to $(\{a, b\}, \{c\}, \{d\})$, denoted by (ab, c, d) for simplicity. Note that a and b are adjacent in all preference rankings, which is a necessary condition to be clustered (but not sufficient for clustered single-peakedness!). A preference profile is clustered single-peaked with respect to an ordered partition $(C_1, \ldots, C_q)$ iff it is single-peaked when considering each subset C_i as a single candidate. In the example, introducing cluster $\{a, b\}$ amounts to considering a and b as a "single candidate" ab. The preference profile matrix becomes then the one indicated on the right-hand side of Figure 1. In the graph on the right, one can observe that the preferences become then single-peaked, i.e. they are clustered single-peaked with respect to (ab, c, d).*
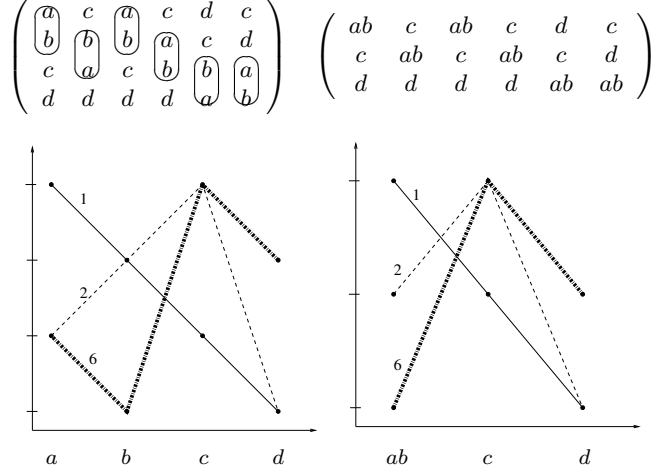
$$\begin{pmatrix} \boxed{a} & c & \boxed{a} & c & d & c \\ \boxed{b} & \boxed{b} & \boxed{b} & \boxed{a} & c & d \\ c & \boxed{a} & c & \boxed{b} & \boxed{b} & \boxed{a} \\ d & d & d & d & \boxed{a} & \boxed{b} \end{pmatrix} \quad \begin{pmatrix} ab & c & ab & c & d & c \\ c & ab & c & ab & c & d \\ d & d & d & d & ab & ab \end{pmatrix}$$

Figure 1: Clustered single-peakedness.

3 Dynamic Programming

We now present a dynamic programming algorithm that generalizes the one proposed by Betzler et al. for single-peaked preferences [4]. Let $P(i, C', k)$ denote the subproblem where all candidates in $C' \subseteq C$ are made mandatory and one selects $k - |C'|$ candidates in $C_1 \cup \ldots \cup C_i$. For the convenience of the reader, we briefly recall the recursion scheme of the procedure proposed by Betzler et al., with an alternative proof. Assume that the preferences are single-peaked with respect to axis $(x_1, \ldots, x_m)$ (i.e. clustered single-peaked with respect to $(C_1, \ldots, C_m)$, where $C_i = \{x_i\}$ $\forall i$). Let $z(i, k)$ denote the optimal score for problem $P(i - 1, \{x_i\}, k)$, where one selects x_i and $k - 1$ candidates among $\{x_1, \ldots, x_{i-1}\}$ (the $i - 1$ leftmost candidates on the axis). The authors use the following recursion:

$$z(i, k) = \min_{j \in [k-1 \cdots i-1]} \left\{ z(j, k - 1) - \sum_{v \in V} \max\{0, \mu(r(v, x_j)) - \mu(r(v, x_i))\} \right\}$$

The optimal score for a subset of κ candidates is then $\min_{i \in \{\kappa \cdots m\}} z(i, \kappa)$. The validity of the recursion can be established by showing that selecting a subset of k candidates, including x_j and x_i (mandatory candidates), in $\{x_1, \ldots, x_j, x_i\}$ (problem $P(j - 1, \{x_j, x_i\}, k)$) amounts to selecting $k - 1$ candidates, including x_j, in $\{x_1, \ldots, x_j\}$ (problem $P(j - 1, \{x_j\}, k - 1)$). Indeed, it reduces to computations on the same *minor* of the preference profile.

Definition 2. *Any preference profile matrix that depicts the individual preferences of a subset $V' \subseteq V$ of voters over a subset $C' \subseteq C$ of candidates is called a* minor *and denoted by $\mathcal{P}(V', C')$.*

The voters can be partitioned into two sets: the set $V_{[1,j-1]}$ of voters whose peak x_p is in $\{x_1, \ldots, x_{j-1}\}$, and the set $V_{[j,m]}$ of voters whose peak x_p is in $\{x_j, \ldots, x_m\}$. Both problems $P(j - 1, \{x_j, x_i\}, k)$ and $P(j - 1, \{x_j\}, k - 1)$ amount to computations in the same minor:

• *Problem $P(j-1, \{x_j, x_i\}, k)$*: all voters in $V_{[j,m]}$ can be deleted from the preference profile matrix since their preferred candidate among $\{x_1, \ldots, x_j, x_i\}$ is either x_i or x_j, that are mandatory, and therefore the preferences of these voters play no role in the determination of the optimal solution to $P(j - 1, \{x_j, x_i\}, k)$. Furthermore, all voters in $V_{[1,j-1]}$ prefer x_j to x_i since their peak is to the left of x_j, and therefore candidate x_i plays no role since x_j is mandatory. Consequently, the problem reduces to selecting $k - 1$ candidates, including x_j, according to minor $\mathcal{P}(V_{[1,j-1]}, \{x_1, \ldots, x_j\})$.

- *Problem $P(j-1, \{x_j\}, k-1)$*: for all voters in $V_{[j,m]}$, candidate x_j is necessarily the most preferred one in $\{x_1, \ldots, x_j\}$. Since candidate x_j is mandatory, all voters in $V_{[j,m]}$ can be deleted from the preference profile matrix. The problem reduces then to selecting $k-1$ candidates, including x_j, according to minor $\mathcal{P}(V_{[1,j-1]}, \{x_1, \ldots, x_j\})$.

The two problems $P(j-1, \{x_j, x_i\}, k)$ and $P(j-1, \{x_j\}, k-1)$ are thus equivalent, which establishes the validity of the recursion. We now show how this recursion scheme can be extended to handle clustered single-peaked preferences. Assume that the preferences are clustered single-peaked with respect to an ordered partition $(C_1, \ldots, C_q)$. Let $z(i, C_i', k)$ denote the optimal score when candidates in $C_i' \subseteq C_i$ are mandatory, candidates in $C_i \setminus C_i'$ are forbidden, and one selects $k - |C_i'|$ candidates in $C_1 \cup \ldots \cup C_{i-1}$. In our setting, the recursion can be written as follows:

$$z(i, C_i', k) = \min_{j \in [1 \cdots i-1]} \min_{C_j' \subseteq C_j \, C_j' \neq \emptyset} \left\{ z(j, C_j', k - |C_i'|) \right.$$
$$\left. - \sum_{v \in V} \max\{0, \min_{y \in C_j'} \mu(r(v,y)) - \min_{x \in C_i'} \mu(r(v,x))\} \right\} \quad (1)$$

where $z(i, C_i', k) = +\infty$ if $|C_i'| > k$ or $|C_1 \cup \ldots \cup C_{i-1}| < k - |C_i'|$.
The optimal score for a subset of κ candidates is then:

$$\min_{i \in [1 \cdots q]} \min_{C_i' \subseteq C_i \, C_i' \neq \emptyset} z(i, C_i', \kappa)$$

The proof of the recursion is similar to the one in the single-peaked case. It amounts to establishing the equivalence of problems $P(j-1, C_j' \cup C_i', k)$ and $P(j-1, C_j', k - |C_i'|)$, by considering a partition of V into the set $V_{[1,j-1]}$ of voters whose peak is in $\{C_1, \ldots, C_{j-1}\}$ and the set $V_{[j,q]}$ whose peak is in $\{C_j, \ldots, C_q\}$:

- *Problem $P(j-1, C_j' \cup C_i', k)$*: all voters in $V_{[j,q]}$ can be deleted from the preference profile matrix since their preferred candidate among $C_1 \cup \ldots \cup C_{j-1} \cup C_j' \cup C_i'$ is either in C_i' or in C_j'. All voters $V_{[1,j-1]}$ prefer a candidate in C_j' to a candidate in C_i' since their peak is to the left of C_j. Consequently, the problem reduces to selecting $k - |C_i'|$ candidates, including candidates in C_j', according to minor $\mathcal{P}(V_{[1,j-1]}, C_1 \cup \ldots \cup C_{j-1} \cup C_j')$.

- *Problem $P(j-1, C_j', k - |C_i'|)$*: for all voters in $V_{[j,q]}$, the most preferred candidate in $C_1 \cup \ldots \cup C_{j-1} \cup C_j'$ necessarily belongs to C_j'. The voters can therefore be deleted from the preference profile. The problem reduces then to selecting $k - |C_i'|$ candidates, including candidates in C_j', according to minor $\mathcal{P}(V_{[1,j-1]}, C_1 \cup \ldots \cup C_{j-1} \cup C_j')$.

Both problems are thus equivalent, which establishes the validity of the recursion. Algorithm 1 describes the ensuing dynamic programming procedure.

Algorithm 1: Dynamic programming

> **for** $i = 1, \ldots, q$ **do**
>   **for** $C_i' \subseteq C_i$ *with* $|C_i'| \leq \kappa$, $C_i' \neq \emptyset$ **do**
> $z(i, C_i', |C_i'|) = \sum_{v \in V} \min_{x \in C_i'} \mu(r(v,x))$
> **for** $i = 2, \ldots, q$ **do**
>   **for** $C_i' \subseteq C_i$ *with* $|C_i'| \leq \kappa$, $C_i' \neq \emptyset$ **do**
>   **for** $k = |C_i'| + 1, \ldots, \min\{\kappa, |C_i'| + \sum_{j=1}^{i-1} |C_j|\}$ **do**
>   compute $z(i, C_i', k)$ by Equation 1
> **return** $\min_{i \in [1 \cdots q]} \min_{C_i' \subseteq C_i, \, C_i' \neq \emptyset} z(i, C_i', \kappa)$

Example 3. *For simplicity, a set $\{a, b\}$ is denoted by ab in this example, and $\{a, b\} \cup \{c, d\}$ by abcd. Consider a proportional representation problem with 6 candidates a, b, c, d, e, f having clustered single-peaked preferences with respect to (ab, cd, e, f). Let us study how many triples of*

candidates are examined by the procedure when computing $z(4, f, 3)$. Given r subsets $S_1, \ldots, S_r$ of candidates, let us denote by $\mathrm{opt}\{S_1, \ldots, S_r\}$ a subset in $\arg \min_i s(S_i)$. The following computation is performed by the procedure:

$$z(4, f, 3) = s \left(\mathrm{opt} \left\{ \begin{array}{l} fab, f\mathrm{opt}\{ca, cb\}, f\mathrm{opt}\{da, db\}, \\ fcd, f\mathrm{opt}\{ea, eb, ec, ed\} \end{array} \right\} \right)$$

Therefore 5 subsets are examined (three of the four "opt" operations have been performed during the previous iterations) while there are 10 subsets of cardinality 3 including f.

For a small single-peaked width, the computational savings become of course more and more significant when the size of the instance increases. Actually, the following complexity analysis shows that the dynamic programming procedure is polynomial for a bounded singe peaked width. Equation 1 requires indeed a computational time within $O(nqt2^t)$ where $t = \max_i |C_i| - 1$. Furthermore, the number of computed terms $z(i, C_i', k)$ is upper bounded by $q2^t\kappa$. Therefore the running time of the procedure is within $O(nq^2t2^{2t}\kappa)$, which amounts to $O(nm^3)$ for a bounded single-peaked width t (we recall that $q \leq m$ and $\kappa \leq m$).

Theorem 1. *The proportional representation problem over bounded single-peaked width preferences is polynomial.*

The complexity analysis shows that $\max_i |C_i|$ is a key parameter for the efficiency of the algorithm. Note that there always exists an ordered partition for which the preferences are clustered single-peaked: in the worst case, it is sufficient to consider the partition (C). It is nevertheless interesting from an algorithmic viewpoint to have an ordered partition where each subset includes few candidates. Two cases can occur: either the partition is known in advance (for instance, when the candidates indicate their affiliation to a political party and the preferences of the voters are consistent with the displayed affiliations) or it is unknown. In both cases, it is desirable to be able to compute an ordered partition compatible with clustered single-peakedness and such that $\max_i |C_i|$ is minimized. In the next section, we show the polynomiality of this problem for *narcissistic preferences* [3, 13].

4 Single peaked width

We call *width* of an ordered partition $(C_1, \ldots, C_q)$ the value $\max_i |C_i| - 1$. Given a preference profile matrix, we call *single-peaked width* the minimum width among all ordered partitions compatible with clustered single-peakedness. This can be seen as a distance measuring near-single-peakedness (the single-peaked width is indeed equal to 0 for single-peaked preferences). Note that this should not be confused with other distance measures that have been proposed in the literature, such as the number of voters to remove to make a profile single-peaked [9].

Example 4. *Consider the preference profile matrix $\mathcal{P}$ represented in Figure 2, where the preferences are not single-peaked. It is easy to check that they are nevertheless clustered single-peaked with respect to ordered partition (ac, efg, bd, h) (see the left part of the figure, where the subsets of the partition are encircled), whose width is $|\{e, f, g\}| - 1 = 2$. However the preferences are also clustered single-peaked with respect to (ac, f, eg, b, d, h) (right part of the figure). The single-peaked width of this preference matrix is thus 1.*

Ballester and Haeringer [2] recently showed that single-peakedness can be lost just because of the existence of two voters and four candidates, or three voters and three candidates. Conversely, they showed that if a profile is not single-peaked there must exist a set of two voters (resp. three) whose preferences over four candidates (resp. three) are not single-peaked. More precisely, the authors characterize single-peakedness with the following two conditions:

$$\mathcal{P} = \begin{pmatrix} \boxed{f} & \boxed{b} & \boxed{c} \\ \boxed{g} & \boxed{d} & \boxed{a} \\ \boxed{e} & h & \boxed{f} \\ \boxed{c} & \boxed{g} & \boxed{e} \\ \boxed{a} & \boxed{e} & \boxed{g} \\ \boxed{b} & \boxed{f} & \boxed{b} \\ \boxed{d} & \boxed{a} & \boxed{d} \\ h & \boxed{c} & h \end{pmatrix} \qquad \mathcal{P} = \begin{pmatrix} f & b & \boxed{c} \\ \boxed{g} & d & \boxed{a} \\ \boxed{e} & h & f \\ \boxed{c} & \boxed{g} & \boxed{e} \\ \boxed{a} & \boxed{e} & \boxed{g} \\ b & f & b \\ d & \boxed{a} & d \\ h & \boxed{c} & h \end{pmatrix}$$

Figure 2: Single-peaked width.

• *Worst-restriction:* Given a triple $V' \subseteq V$ of voters and a triple $C' \subseteq C$ of candidates, let $L(V', C')$ be the set of all candidates ranked last in C' by at least one voter in V'. The worst-restriction condition holds if $|L(V', C')| < 3$ for all triples V' and C'.

• *α-restriction:* the α-restriction condition holds if there do not exist two voters v and v' and four candidates w, x, y, and z such that their preferences over w, x and z are opposite ($w \succ_v x \succ_v z$ and $z \succ_{v'} x \succ_{v'} w$) and the voters agree about the preference for y over x ($y \succ_v x$ and $y \succ_{v'} x$).

Interestingly, these conditions amount to forbidding five minors in the profile $\mathcal{P}$ (Lemma 1). In this formalism, we propose here a shorter proof of the characterization result of Ballester and Haeringer. Our proof is based on the polynomial algorithm proposed by Escoffier et al. [8] to determine if a profile is single-peaked with respect to some axis. This algorithm runs in time $O(mn)$ improving on the $O(mn^2)$ algorithm proposed by Bartholdi and Trick [3]. Before stating Lemma 1, let us present the algorithm of Escoffier et al. It works recursively and takes as arguments the left part $(x_1, \ldots, x_i)$ and the right part $(x_j, \ldots, x_m)$ of the axis under construction. A third argument is the subset C' of candidates which remains to be positioned on the axis. This algorithm returns an axis compatible with $\mathcal{P}$ or proves that the preferences are not single-peaked (by raising a contradiction between voters). The recursion is made possible by the fact that single-peakedness over $\mathcal{P}$ implies single-peakedness over any of its minors. It heavily uses the property that candidates ranked last in the preferences are necessarily at the extremities of the axis. At each step of the algorithm, one candidate x or two candidates x and y are ranked last in $\mathcal{P}(V, C')$ and will be positioned in x_{i+1} or x_{j-1} on the axis. There is a *contradiction* if a candidate has to be placed in two different positions (according to the preferences of two voters). These positions depend on the way x and y are positioned with respect to x_i and x_j in the preferences of all the voters. The whole procedure is detailed in Algorithm 2. The initial call is Make-axis$(C, (), ())$.

Before presenting Lemma 1 (on which our algorithm to compute single-peaked width strongly relies), we need to introduce the notion of *isomorphic* minors. A minor $\mathcal{P}'$ is isomorphic to $\mathcal{P}$ if there exists a bijection ϕ such that $\mathcal{P}$ and $\mathcal{P}'$ are identical up to column permutation if one renames every candidate x in $\mathcal{P}$ as $\phi(x)$. For instance, preference profile matrix $\mathcal{P}'$ below is isomorphic to $\mathcal{P}$ (take $\phi(a) = b$, $\phi(b) = c$, $\phi(c) = a$ and permute the columns).

$$\mathcal{P} = \begin{pmatrix} a & c \\ b & a \\ c & b \end{pmatrix}, \quad \mathcal{P}' = \begin{pmatrix} a & b \\ b & c \\ c & a \end{pmatrix}$$

Definition 3. *A minor is called* forbidden *if it is isomorphic to one of the following profiles:*

$$\mathcal{T}_1 = \begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix}, \mathcal{T}_2 = \begin{pmatrix} a & c & a \\ b & b & c \\ c & a & b \end{pmatrix},$$

$$\mathcal{F}_1 = \begin{pmatrix} a & c \\ d & d \\ b & b \\ c & a \end{pmatrix}, \mathcal{F}_2 = \begin{pmatrix} a & d \\ d & c \\ b & b \\ c & a \end{pmatrix}, \text{ or } \mathcal{F}_3 = \begin{pmatrix} d & d \\ a & c \\ b & b \\ c & a \end{pmatrix}.$$

Algorithm 2: Make-axis(C',$(x_1,\ldots,x_i)$,$(x_j,\ldots,x_m)$)
---
1 **if** $C' = \emptyset$ **then return** $(x_1,\ldots,x_i,x_j,\ldots,x_m)$
2 **if** $C' = \{x\}$ **then return** $(x_1,\ldots,x_i,x,x_j,\ldots,x_m)$
3 $L \leftarrow$ candidates ranked last in $\mathcal{P}(C',V)$ by at least one voter
4 **if** $L = \{x\}$ **then** $y \leftarrow$ a candidate in $C' \setminus \{x\}$ /* $x \prec_v y, \forall v$ */
5 **if** $|L| \geq 3$ **then return** *not single-peaked*
6 **for** $v = 1,\ldots,n$ **do**
7   **if** $L = \{x,y\}$ **then** let $x \prec_v y$ (w.l.o.g)
8   **if** $x_i \prec_v x \prec_v x_j \prec_v y$ **or** $x_i \prec_v x \prec_v y \prec_v x_j$ **then**
9   **if** *no contradiction* **then** $x_{i+1} \leftarrow x$; $x_{j-1} \leftarrow y$
10   **else return** *not single-peaked*
11   **if** $x_j \prec_v x \prec_v x_i \prec_v y$ **or** $x_j \prec_v x \prec_v y \prec_v x_i$ **then**
12   **if** *no contradiction* **then** $x_{i+1} \leftarrow y$; $x_{j-1} \leftarrow x$
13   **else return** *not single-peaked*
14 **if** $L = \{x\}$ **then**
15   **if** $x = x_{i+1}$ **then** Make-axis($C' \setminus \{x\}$,$(x_1,\ldots,x_i,x_{i+1})$,$(x_j,\ldots,x_m)$)
16   **else** Make-axis($C' \setminus \{x\}$,$(x_1,\ldots,x_i)$,$(x_{j-1},x_j,\ldots,x_m)$)
17 Make-axis($C' \setminus \{x,y\}$,$(x_1,\ldots,x_i,x_{i+1})$,$(x_{j-1},x_j,\ldots,x_m)$)

Lemma 1. *$\mathcal{P}$ is single-peaked iff it has no forbidden minor.*

Proof (sketch) *Necessity*: it suffices to check that none of the five forbidden minors is single-peaked, since the single-peakedness property is closed under taking minors.
Sufficiency: run Algorithm 2 and suppose that it returns *not single-peaked*. If it stops at Line 5, then $\mathcal{P}$ has a minor $\mathcal{T}_1$ or $\mathcal{T}_2$. Otherwise it stops at Line 10 or 13 and $\mathcal{P}$ has a minor $\mathcal{F}_1$, $\mathcal{F}_2$ or $\mathcal{F}_3$.   ∎

The rest of the section is devoted to the problem of determining an ordered partition of minimum width among the ones that are compatible with clustered single-peakedness. Note that Elkind et al. [7] studied a closely related problem, namely finding an ordered partition $(C_1,\ldots,C_q)$ maximizing q. Both problems are not equivalent, as shown by the following example.

Example 5. *Consider the preference profile matrix $\mathcal{P}$:*

$$\mathcal{P} = \begin{pmatrix} d & d \\ x & a \\ y & v \\ c & b \\ b & c \\ a & x \\ v & y \end{pmatrix}$$

Both partitions $(abcv,d,x,y)$ and $(v,a,d,bcxy)$ maximize q and are compatible with clustered single-peakedness, but (abv,d,cxy) is the only partition that minimizes the single-peaked width.

However, for *narcissistic preferences* [3, 13], one can show that the algorithm proposed by Elkind et al. for their problem returns an ordered partition of minimum width. Nevertheless, our approach proves that there is a unique (up to reversal) ordered partition maximizing q. Preferences are said to be *narcissistic* when each candidate is most preferred by some voter. In politics, as soon as the candidates are voting, this assumption seems reasonable. In the remainder, we prove the following result:

Theorem 2. *Finding the single-peaked width is polynomial if $\mathcal{P}$ is narcissistic.*

For each voter $v \in V$ and candidates $a, b \in C$ we denote $I_v(a,b) := \{c \in C : c = a$ or $c = b$ or $a \succ_v c \succ_v b$ or $b \succ_v c \succ_v a\}$ the set of candidates between a and b in the preferences of voter v.

By convention, $I_v(a,a) = \{a\}$. A subset I of C is called an *interval of $\mathcal{P}$* if for each $v \in V$, one can choose two candidates $a, b \in I$ such that $I = I_v(a,b)$. This definition coincides with the notion of *clone set* studied by Elkind et al. [7]. Notice that the set of intervals $\mathcal{I}$ of $\mathcal{P}$ is not closed under taking subsets. Nevertheless, it is closed under intersection [7]. Given $a, b \in C$, the minimal interval w.r.t. inclusion that contains a and b is thus uniquely defined: we denote it by $I(a,b)$. The following lemma will prove useful in order to design an algorithm able to compute a partition compatible with clustered single-peakedness. For simplicity, if $\mathcal{P}'$ is isomorphic to $\mathcal{P}$ for ϕ, we write $I(x,y)$ for $I(\phi(x), \phi(y))$.

Lemma 2. *The following properties hold:*
- *If $\mathcal{T}_1$ is a minor of $\mathcal{P}$, then $I(a,b) = I(a,c) = I(b,c)$;*
- *If $\mathcal{T}_2$ is a minor of $\mathcal{P}$, then $I(a,b)$ and $I(a,c)$ include $I(b,c)$;*
- *If $\mathcal{F}_1$ is a minor of $\mathcal{P}$, then $I(a,b)$, $I(a,c)$, $I(a,d)$, $I(b,c)$ and $I(c,d)$ include $I(b,d)$;*
- *If $\mathcal{F}_2$ is a minor of $\mathcal{P}$, then $I(a,b)$, $I(a,c)$, $I(a,d)$, $I(b,d)$ and $I(c,d)$ include $I(b,c)$;*
- *If $\mathcal{F}_3$ is a minor of $\mathcal{P}$, then*

  - *$I(a,c)$, $I(a,d)$, $I(b,d)$ and $I(c,d)$ include $I(a,b)$,*
  - *$I(a,c)$, $I(a,d)$, $I(b,d)$ and $I(c,d)$ include $I(b,c)$.*

Proof (sketch) Let v be the voter of the first column of $\mathcal{T}_1$. Since $b \in I_v(a,c)$, it follows that $b \in I(a,c)$. Thus $I(a,b)$ and $I(b,c) \subseteq I(a,c)$. The second column gives $I(b,c)$ and $I(a,c) \subseteq I(a,b)$, and the third column gives $I(a,c)$ and $I(a,b) \subseteq I(b,c)$. Finally $I(a,b) = I(a,c) = I(b,c)$. The proofs for the four other forbidden minors go along the same lines. ∎

We propose a greedy algorithm to compute the clusters of an ordered partition compatible with clustered single-peakedness. This algorithm proceeds by *contracting* candidates so that no forbidden minor remains in the preference profile matrix. *Contracting two candidates $a, b \in C$* consists in contracting $I(a,b)$. *Contracting an interval I* consists in collapsing all candidates in I into a single "cluster" candidate. This amounts to choosing a representative in I and removing from $\mathcal{P}$ all the other candidates in I. For instance, contracting b and d in $\mathcal{P}$ yields cluster $\{b,d,e\}$ (since $I(b,d) = \{b,d,e\}$) and profile $\mathcal{P}'$:

$$\mathcal{P} = \begin{pmatrix} a & c \\ d & d \\ e & b \\ b & e \\ c & a \end{pmatrix} \qquad \mathcal{P}' = \begin{pmatrix} a & c \\ b & b \\ c & a \end{pmatrix}$$

Notice that the preference profile matrix $\mathcal{P}'$ obtained by contracting an interval of $\mathcal{P}$ is a minor of $\mathcal{P}$. Note also that if $I, J \in \mathcal{I}$ are two intervals of $\mathcal{P}$, then the two minors obtained from $\mathcal{P}$ either by contracting I then J, or by contracting J then I coincide (even if I and J overlap). Besides, if $\mathcal{P}'$ is a minor of $\mathcal{P}$ and $\mathcal{F}'$ is a minor of $\mathcal{P}'$, then $\mathcal{F}'$ is also a minor of $\mathcal{P}$. The greedy procedure is detailed in Algorithm 3. The termination follows from the fact that contracting candidates cannot create new forbidden minors.

Example 6. *Consider the preference profile matrix $\mathcal{P}$ in Figure 2 and apply Algorithm 3, assume that it detects:*

$$\text{the minor} \quad \begin{pmatrix} g & g & c \\ c & a & a \\ a & c & g \end{pmatrix} \quad \text{and then the minor} \quad \begin{pmatrix} f & h \\ g & g \\ e & e \\ h & f \end{pmatrix}$$

Algorithm 3: Greedy algorithm
---
let $\mathcal{P}'$ be a minor of $\mathcal{P}$ isomorphic for ϕ to:

$\mathcal{T}_1$. Contract $\phi(a)$ and $\phi(b)$

$\mathcal{T}_2$. Contract $\phi(b)$ and $\phi(c)$

$\mathcal{F}_1$. Contract $\phi(b)$ and $\phi(d)$

$\mathcal{F}_2$. Contract $\phi(b)$ and $\phi(c)$

$\mathcal{F}_3$. Contract $\phi(a)$ and $\phi(b)$, or $\phi(b)$ and $\phi(c)$

apply these contractions (non-deterministically) until no forbidden minor remains.

The first minor is isomorphic to $\mathcal{T}_2$ for $\phi(a) = g$, $\phi(b) = a$ and $\phi(c) = c$. Therefore candidates a and c are contracted. The second minor is isomorphic to $\mathcal{F}_1$ for $\phi(a) = f$, $\phi(b) = e$, $\phi(c) = h$ and $\phi(d) = g$. Therefore candidates e and g are contracted. Taking candidate a (resp. e) as the representative of cluster $\{a, c\}$ (resp. $\{e, g\}$), the preference profile becomes:

$$\mathcal{P}' = \begin{pmatrix} f & b & a \\ e & d & f \\ a & h & e \\ b & e & b \\ d & f & d \\ h & a & h \end{pmatrix}$$

There is no more forbidden minor in the preference profile, and thus the greedy procedure stops. The clusters are $\{a, c\}$ and $\{e, g\}$.

This algorithm is polynomial since the forbidden minors can be enumerated in $O(m^3 n^3)$ for $\mathcal{T}_1$, $\mathcal{T}_2$, and $O(m^4 n^2)$ for $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$. The clusters identified by the algorithm belong to an ordered partition compatible with clustered single-peakedness. The ordered partition itself can be computed by applying Algorithm 2 on the final preference profile. Coming back to Example 6, Algorithm 2 returns axis (h, d, b, e, f, a) on $\mathcal{P}'$, which corresponds to the ordered partition (h, d, b, eg, f, ac) since e (resp. a) is the representative of $\{e, g\}$ (resp. $\{a, c\}$). This is an ordered partition of minimum width for this profile. However, in the general case, the width of the returned ordered partition is not guaranteed to be minimal. We now show how to refine the greedy procedure to get an ordered partition of minimum width when preferences are narcissistic. To this end, we introduce a notion of similarity between candidates that enables us to identify necessary and sufficient contractions for clustered single-peakedness.

Definition 4. *Two candidates a and b are said to be* similar *if they belong to the same cluster in all ordered partitions w.r.t. which $\mathcal{P}$ is clustered single-peaked.*

It results from Lemma 2 that the following properties hold:

- If $\mathcal{T}_1$ (T_2) is a minor of $\mathcal{P}$, then a and b (b and c) are similar;

- If $\mathcal{F}_1$ (F_2) is a minor of $\mathcal{P}$, then b and d (b and c) are similar;

- If $\mathcal{F}_3$ is a minor of $\mathcal{P}$, then
    - if $I(b, c) \subseteq I(a, b)$, then b and c are similar;
    - if $I(a, b) \subseteq I(b, c)$, then a and b are similar.

These properties imply that all contractions but one ($\mathcal{F}_3$) in the greedy algorithm cover candidates which belong to the same cluster in any ordered partition of minimum width. The only case of a forbidden minor that cannot be removed from $\mathcal{P}$ by contracting similar candidates is thus $\mathcal{F}_3$ when $I(a, b)$ and $I(b, c)$ intersect properly, i.e. $I(a, b) \nsubseteq I(b, c)$ and $I(b, c) \nsubseteq I(a, b)$. We call such forbidden minors *ambiguous*. If one finds an ambiguous minor $\mathcal{M}$, at least two candidates in $\mathcal{M}$

must be in the same cluster. Nevertheless the single-peaked width of an ordered partition depends on the choice of the candidates to contract. Furthermore this choice does not only depend on the maximum number of candidates involved in the possible interval contractions. For instance, consider the preference profile matrix $\mathcal{P}$ of Example 5 which has the following minor:

$$\begin{pmatrix} d & d \\ c & v \\ b & b \\ v & c \end{pmatrix}$$

The smallest contraction implied by the given minor would be to contract b and c (2 candidates in the interval). But (abv, d, cxy), where b and c are not in the same cluster, is the only minimum width ordered partition compatible with clustered single-peakedness.

For this reason, the greedy algorithm may fail to provide clusters belonging to an ordered partition of minimum width. However when preferences are narcissistic, no ambiguous minor can exist in the preference profile matrix. Assume indeed that there exists a minor $\mathcal{M}$ isomorphic to $\mathcal{F}_3$ for ϕ. Since $\mathcal{P}$ is narcissistic, candidate $\phi(b)$ is the most preferred one for some voter v, and consequently: $\phi(b) \succ_v \phi(a) \succ_v \phi(c)$ or $\phi(b) \succ_v \phi(c) \succ_v \phi(a)$. Therefore we have $I(a,b) \subseteq I(b,c)$ or $I(b,c) \subseteq I(a,b)$. The minor is thus unambiguous. To obtain an optimal greedy algorithm for narcissistic preferences, contraction related to $\mathcal{F}_3$ must then be modified as follows:

let v be a voter whose most preferred candidate is $\phi(b)$
if $\phi(b) \succ_v \phi(a) \succ_v \phi(c)$ then contract $\phi(a)$ and $\phi(b)$ else contract $\phi(b)$ and $\phi(c)$.

Furthermore, the greedy algorithm uses necessary and sufficient contractions to make the profile clustered single-peaked, and thus partition $(C_1, \ldots, C_q)$ of minimum width is clearly unique under maximizing the number q of clusters.

5   Conclusion

An interesting open question is whether there exists a general polynomial algorithm to compute the single-peaked width (not necessarily in the narcissistic case). Adapting the PQ-trees based algorithm of Elkind et al. [7] to our problem could work. Besides, the concept of minors could be a tool for finding a short validity proof.

6   Acknowledgements

References

[1] M. Balinski. On finding integer solutions to linear programs. In *Proc. IBM Scientific Computing Symposium on Combinatorial Problems*, pages 225–248, 1966.

[2] M. A. Ballester and G. Haeringer. A characterization of the single-peaked domain. *Social Choice and Welfare*, 36(2):305–322, 2011.

[3] J. Bartholdi III and M. A. Trick. Stable matching with preferences derived from a psychological model. *Operations Research Letters*, 5(4):165 – 169, 1986.

[4] N. Betzler, A. Slinko, and J. Uhlmann. On the computation of fully proportional representation. Available at SSRN: http://ssrn.com/abstract=1952497.

[5] D. Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

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

[7] E. Elkind, P. Faliszewski, and A. Slinko. Clone structures in voters' preferences. *ArXiv e-prints*, 1110.3939, 2011.

[8] B. Escoffier, J. Lang, and M. Öztürk. Single-peaked consistency and its complexity. In *Proceedings of 18th European Conference on Artificial Intelligence (ECAI 2008)*, pages 366–370. IOS Press, 2008.

[9] P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra. The complexity of manipulative attacks in nearly single-peaked electorates. In *TARK*, pages 228–237, 2011.

[10] T. Lu and C. Boutilier. Budgeted social choice: From consensus to personalized decision making. In *IJCAI*, pages 280–286, 2011.

[11] R. Potthoff and S. Brams. Proportional representation: broadening the options. *Journal of Theoretical Politics*, 10(2):pp. 147–178, 1998.

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

[13] M. A. Trick. Recognizing single-peaked preferences on a tree. *Mathematical Social Sciences*, 17(3):329–334, 1989.

Denis Cornaz
University Paris Dauphine
LAMSADE
Place du Marchal de Lattre de Tassigny
75775 Paris cedex 16, France
Email: denis.cornaz@dauphine.fr

Lucie Galand
University Paris Dauphine
LAMSADE
Place du Marchal de Lattre de Tassigny
75775 Paris cedex 16, France
Email: lucie.galand@dauphine.fr

Olivier Spanjaard
University of Pierre et Marie Curie
LIP6
4 Place Jussieu
75252 Paris Cedex 05, France
Email: olivier.spanjaard@lip6.fr