A Note on Justified Representation Under the Reverse Sequential PAV rule

Haris Aziz

Data61, CSIRO and UNSW Australia
Computer Science and Engineering, Building K17, UNSW, Sydney NSW 2052, Australia

---

Abstract

We consider a committee voting setting in which each voter approves of a subset of candidates and based on the approvals, k candidates are to be selected. A particularly compelling rule for this setting is PAV (Proportional Approval Voting) that was proposed by Thiele. Thiele also proposed two sequential versions of the PAV rule. Whereas one of the sequential rules (SeqPAV) has received much attention in the literature and has even been used in political elections, the other one that we refer to as RevSeqPAV (Reverse Sequential PAV) is relatively less studied. We show that in contrast to SeqPAV, RevSeqPAV satisfies justified representation when the number of candidates is at most $k+2$. We also show that RevSeqPAV does not satisfy justified representation in general for $k \geq 5$.

Keywords: Social choice theory, committee voting, multi-winner voting, approval voting
JEL: C63, C70, C71, and C78

---

1. Introduction

We consider a committee voting setting in which each voter approves of a subset of candidates and based on the approvals, a target k number of candidates are selected. The setting has been referred to as approval-based multi-winner voting or committee voting with approvals and has inspired a number of natural voting rules [see e.g, 2, 5, 7].

Among the various approval-based committee rules, one of the most prominent is *PAV* (Proportional Approval Voting) [see e.g, 2, 5, 4]. In *PAV*, each voter is viewed as getting an additional score of $1/j$ for getting the j-th approved candidate in the committee. The *PAV* rule returns a committee with the highest total *PAV* score for the voters. The *PAV* rule was proposed by the Danish polymath Thorvald N. Thiele in the 19th century and then rediscovered

Email address: `haris.aziz@data61.csiro.au` (Haris Aziz)

by Forrest Simmons [4]. The *PAV* rule is desirable because it is satisfies natural notions of representation. In particular it is one of known natural rules that satisfies *EJR* (extended justified representation) [1]. The property is stronger than *justified representation (JR)* [1] and *proportional justified representation (PJR)* [6] both considered as desirable properties in approval-based representation. The idea behind all the three properties is that a cohesive and large enough group deserves sufficient number of approved candidates in the winning set of candidates.

Whereas *PAV* appears to be a compelling rule that has received considerable interest, it has one drawback: it is computationally intractable [2, 7]. Its relatively more complex nature was evident to Thiele who proposed SeqPAV and RevSeqPAV that are two sequential and greedy versions of *PAV* [4]. In SeqPAV, a committee is gradually built while adding candidates that result in the largest increase in the *PAV*-score. In RevSeqPAV, we start from the set of all candidates and gradually remove candidates that result in the least decrease in the *PAV*-score until exactly k candidate are left. Both sequential versions have another advantage over PAV: they satisfy a natural monotonicity property that is called house monotonicity or committee monotonicity [4].[1]

Whereas SeqPAV is better known, RevSeqPAV has received much less attention. Interestingly, Thiele gave one possible reason for preferring RevSeqPAV over SeqPAV: the latter *"might give the first elected pretensions to be superior to their colleagues"* [4]. In this note, we explore other possible merits of RevSeqPAV especially in terms of satisfying justified representation. In particular, SeqPAV does not satisfy *JR* [1] so we study whether RevSeqPAV satisfies *JR* or not. We show that just like SeqPAV, RevSeqPAV does not satisfy *JR* in general. We complement the result by showing that RevSeqPAV has an advantage over SeqPAV in terms of satisfying *JR* if k is almost equal to the number of candidates.

2. Preliminaries

We consider a social choice setting with a set $N = \{1, \ldots, n\}$ of voters and a set C of m candidates. Each voter $i \in N$ submits an approval ballot $A_i \subseteq C$, which represents the subset of candidates that she approves of. We refer to the list $\vec{A} = (A_1, \ldots, A_n)$ of approval ballots as the *ballot profile*. We will consider *approval-based multi-winner voting rules* that take as input a quadruple $(N, C, \vec{A}, k)$, where k is a positive integer that satisfies $k \leq m$, and return a subset $W \subseteq C$ of size k, which we call the *winning set*, or *committee*.

Given a ballot profile $\vec{A} = (A_1, \ldots, A_n)$ over a candidate set C and a target committee size k, we say that a set of candidates W of size $|W| = k$ *satisfies*

[1] A rule satisfies *house monotonicity or committee monotonicity* if for any outcome W of size k, there is a possible outcome W' of size $k+1$ such that $W' \supset W$.

justified representation for $(\vec{A}, k)$ if

$$\forall X \subseteq N : |X| \geq \frac{n}{k} \text{ and } |\cap_{i\in X} A_i| \geq 1 \implies (|W \cap (\cup_{i\in X} A_i)| \geq 1)$$

A committee rule satisfy *JR* if it always returns a committee that satisfies *JR*.

2.1. PAV-score

The *PAV*-score of a voter i for a committee W is

$$H(|W \cap A_i|)$$

where

$$H(p) = \begin{cases} 0, \text{ for } p = 0 \\ \sum_{j=1}^{p} \frac{1}{j}, \text{ for } p > 0. \end{cases}$$

The *PAV*-score of a committee $W \subseteq C$ is defined as

$$\sum_{i\in N} H(|W \cap A_i|).$$

The *PAV* rule that we discussed in the introduction outputs a set $W \subseteq C$ of size k with the highest *PAV*-score.

We now define marginal contribution as used in [1]. For each candidate $w \in W$, we define $MC(w, W)$ its *marginal contribution* as the difference between the *PAV*-score of W and that of $W \setminus \{w\}$:

$$MC(w, W) = \textit{PAV-score}((W) - \textit{PAV-score}((W \setminus \{w\}).$$

Let $MC(W)$ denote the sum of marginal contributions of all candidates in W:

$$MC(W) = \sum_{w\in W} MC(w, W).$$

The following is a simple lemma that was observed in previous work [1, 3].

Lemma 1. *For any committee* W *such that* $|W| = k$, $\sum_{c\in W} MC(c, W) \leq |\{i \in N : A_i \cap W \neq \emptyset\}|$. *Moreover there exists at least one* $c \in W$ *such that* $MC(c, W) \leq |\{i \in N : A_i \cap W \neq \emptyset\}|/k \leq n/k$.

3. RevSeqPAV

We first specify RevSeqPAV formally as Algorithm 1 (RevSeqPAV). Note that removing $w \in W$ from W with the minimum $MC(w, W)$ is equivalent to removing a candidate that will decrease the *PAV*-score of the committee by the minimum amount.

We observe that unlike SeqPAV and PAV, RevSeqPAV is not an extension of AV (the rule that returns the candidate with high number of approvals). In other words, if $k = 1$, then RevSeqPAV does not return a candidate with the highest number of approvals.

Algorithm 1 RevSeqPAV

Require: $(N, \vec{A}, k)$.
Ensure: W

```
W ⟵ C.
while |W| > k do
    Remove w ∈ W from W with the minimum MC(w, W).
end while
return W.
```

Example 1. *Consider the setting in which $C = \{a, b, c\}$ and $k = 2$.*

$$|\{i \in N : b \in A_i\}| = n/2,$$

$$|\{i \in N : c \in A_i\}| = n/2,$$

and there exists no i such that $b, c \in A_i$. Suppose that

$$|\{i \in N : a \in A_i\}| < 2n/3$$

and

$$|\{i \in N : a \in A_i\} \cap \{i \in N : b \in A_i\}| = |\{i \in N : a \in A_i\} \cap \{i \in N : c \in A_i\}|.$$

In that case when we run RevSeqPAV, we get $\{b, c\}$ and when we run SeqPAV we get $\{a, b\}$ or $\{a, c\}$. If $k = 1$, then the outcome of RevSeqPAV is b or c.

Next, we examine how RevSeqPAV fares in terms of satisfying *JR*, *PJR*, or *EJR*. Our first result is promising.

Proposition 1. *RevSeqPAV satisfies JR if $|C| \leq k + 2$.*

Proof. The statement is trivial for $|C| = k$ so we first consider the case where $|C| = k + 1$. Suppose RevSeqPAV removes $c \in C$ and it results in violation of *JR*. Note that c has to be the candidate that is approved by at least $\frac{n}{k}$ voters who are unrepresented after c's removal. Hence $MC(c, C) \geq n/k$. Since $|\{i{:}A_i \cap (C \setminus \{c\})\}| \leq n - \frac{n}{k}$, by Lemma 1, there exists at least one candidate $w \neq c$ such that $MC(w, W) \leq \frac{n}{k+1} - \frac{n}{k(k+1)}$. Hence c is not removed; a contradiction.

We next consider the case where $|C| = k + 2$. Suppose for contradiction that W the outcome of RevSeqPAV violates *JR*. This means that there exists a set of voters N^* such that $|N^*| \geq n/k$, $c \in \bigcap_{i \in N^*} A_i$ and $c \notin W$. We also suppose that $d \notin W$. Let $|\{i \in N^* : d \in A_i\}| = \lambda \frac{n}{k}$ where $\lambda \in [0, 1]$. If $\lambda = 0$, then we know that c will not be removed because of the argument we have for the case $|C| = k + 1$. Therefore we can assume that $\lambda > 0$ and $\lambda |N^*|$ is an integer. We can also assume that $\lambda < 1$ because if $\lambda = 1$, then after removing one of c or d, RevSeqPAV does not remove the other by the argument for the case $|C| = k + 1$.

Note that $MC(c, C) = |N^*|\frac{\lambda}{2} + |N^*|(1-\lambda)$. If W violates JR, then c needs to be the first candidate to be removed or else d is removed first in which case we can simply assume that $|C| = k+1$ so JR cannot be violated. If c is removed first, then it follows that $MC(d, C) \geq MC(c, C) = |N^*|\frac{\lambda}{2} + |N^*|(1-\lambda)$. Since d derives only $|N^*|\frac{\lambda}{2}$ marginal contribution from the voters in N^* then it must derive the remaining marginal contribution from the voters in $N \setminus N^*$. Hence the candidates in $C \setminus \{c, d\}$ can derive total marginal contribution at most $(n - \frac{n}{k}) - |N^*|(1-\lambda)$ due to Lemma 1. Hence the candidate $w \in C \setminus \{c, d\}$ with the minimum marginal contribution among candidates in $C \setminus \{c, d\}$ has marginal contribution at most $\frac{n}{k} - \frac{n}{k^2} - \frac{n}{k}(1-\lambda)$. When c is removed from the C, the marginal contribution of d increases. It becomes $MC(d, C \setminus \{c\}) = |N^*|\frac{\lambda}{1} + |N^*|(1-\lambda) = |N^*| \geq \frac{n}{k}$. Since $MC(d, C \setminus \{c\}) > MC(w, C \setminus \{c\})$, d is *not* removed; a contradiction. ☐

The proposition above already shows that RevSeqPAV has some merit over SeqPAV because SeqPAV can fail *JR* even if $|C| = k + 1$ [1]. One possible explanation for the seemingly better performance of RevSeqPAV is that when it is removing candidates, it takes into account the synergies of candidates with each other. In Example 1, we saw that candidates b and c complement each other best and RevSeqPAV selects $\{b, c\}$. On the other hand, when SeqPAV is building a committee, it may make a wrong decision early on because synergies with the latter candidates in the committee are not examined. This leads to the question whether RevSeqPAV satisfies *JR* in general. Unfortunately, the answer is no.

Proposition 2. *RevSeqPAV does not satisfy JR if* $k \geq 5$ *and* $|C| \geq k + 7$.

Proof. Consider the profile in which $k \geq 5$, and

$$C = \{c\} \cup X \cup Y$$

where

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

and

$$Y = \{y_1, \dots, y_k\}.$$

The set of voters is

$$N = N_1 \cup N_2 \cup N_3$$

where $|N| = n$, $|N_1| = \frac{n}{k}$, $|N_2| = \frac{3n}{4k}$, and $|N_3| = n - \frac{n}{k} - \frac{3n}{4k} = n - \frac{7n}{4k}$.

We can consider n such that all the set sizes considered in the analysis are integers. In particular, for a given k, we can choose n such that n/k is a multiple of both 3 and 4.

The voters from set N_1 unanimously approve of candidate c. Half of them also approve of x_1, x_3, x_5 and the other half approve of x_2, x_4, x_6.

Each of the voters from set N_2 approve of exactly one candidate among $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ with each of the candidates in $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ get approval from 1/6-th of the voters in N_2.

The voters from N_3 all approve of each candidate in $\{y_1, \ldots, y_k\}$. Now let us trace the sequence of candidate that are removed. In case multiple candidates have the same minimum marginal contribution, we will use lexicographic tie-breaking to choose which candidate to remove.

In the first stage $W = C$; $MC(c, W) = \frac{n}{4k}$, $MC(x, W) = \frac{n}{8k} + \frac{n}{8k} = \frac{n}{4k}$ for each $x \in X$ and $MC(y, W) = \frac{n}{k} - \frac{7n}{4k^2}$ for all $y \in Y$. So if $\frac{n}{4k} \leq \frac{n}{k} - \frac{7n}{4k^2}$, we remove c and update W to $C \setminus \{c\}$. The marginal contribution of candidates in Y is unchanged whereas $MC(x, W) = \frac{n}{6k} + \frac{n}{8k}$ so we remove x_1 if $\frac{n}{6k} + \frac{n}{8k} \leq \frac{n}{k} - \frac{7n}{4k^2}$. Assuming that $MC(x, W) \leq MC(y, W)$ for $y \in Y$ and $x \in X \cap W$, we continue removing candidates x_2, x_3, x_4, x_5 until $W = Y \cup \{x_6\}$. Now $MC(x_6, W) = \frac{n}{2k} + \frac{n}{8k}$. On the other hand $MC(y, W) = \frac{n}{k} - \frac{7n}{4k^2}$ for all $y \in Y$. If we prove that

$$\frac{n}{k} - \frac{7n}{4k^2} > \frac{n}{2k} + \frac{n}{8k} \tag{1}$$

then it follows that the candidates in order of removal are $c, x_1, \ldots, x_6$. Let us analyse the inequality (1):

$$\begin{aligned}
&\frac{n}{k} - \frac{7n}{4k^2} > \frac{n}{2k} + \frac{n}{8k} \\
\iff &\frac{4k-7}{4k} > \frac{4+1}{8} \\
\iff &32k - 56 > 20k \\
\iff &12k > 56
\end{aligned}$$

The inequality indeed holds for any $k \geq 5$. Hence Y is the outcome of RevSeqPAV.

Note that committee Y violates *JR* because the set N_1 has a size at least n/k, the voters in N_1 unanimously approve of candidate c but no voter from N_1 is represented in Y. Thus RevSeqPAV does not satisfy *JR*. □

To make the example concrete, we reproduce the argument for $k = 5$ using a concrete value for n.

Example 2 (RevSeqPAV does not satisfy *JR* for $k = 5$.)**.** *Consider the profile in which $n = 3000$, $k = 5$, and*

$$C = \{c\} \cup X \cup Y$$

where

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\},$$

and

$$Y = \{y_1, \ldots, y_5\}.$$

The voter set is

$$N = N_1 \cup N_2 \cup N_3$$

where $|N_1| = 600$, $|N_2| = 450$, *and* $|N_3| = 3000 - 600 - 450 = 1950$.

The voters from set N_1 *unanimously approve of candidate* c. *Half of them also approve of* x_1, x_3, x_5 *and the other half approve of* x_2, x_4, x_6.

Then each of the voters from set N_2 *each approve of exactly one candidate among* $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ *with each of the candidates in* $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ *get approval from 75 of the 450 voters.*

The voters from N_3 *all approve of each candidate in* $\{y_1, \ldots, y_5\}$. *Now let us trace the sequence of voters who are removed. In case multiple candidates have the same minimum marginal contribution, we will use lexicographic tie-breaking to choose which candidate to remove.*

In the first stage $W = C$; $MC(c, W) = 600/4 = 150$, $MC(x, W) = 300(1/4) + 75 = 150$ *for each* $x \in X$ *and* $MC(y, W) = 1950/5 = 390$ *for all* $y \in Y$. *So we remove* c *and update* W *to* $C \setminus \{c\}$. *The marginal contribution of candidates in* Y *is unchanged whereas* $MC(x, W) = 300/3 + 75 = 175$ *so we remove* x_1. *We continue removing candidates* x_2, x_3, x_4, x_5 *until* $W = Y \cup \{x_6\}$. *Now* $MC(x_6, W) = 300 + 75 = 375$. *On the other hand* $MC(y, W) = 390$ *for all* $y \in Y$. *Therefore* x_6 *is removed and the final outcome of RevSeqPAV is* $W = Y$. *Note that committee* Y *violates JR because the set* N_1 *has a size at least* n/k *and the voters in* N_1 *unanimously approve of candidate* c. *However no voter from* N_1 *is represented in* Y. *Thus RevSeqPAV does not satisfy JR.*

Corollary 1. *RevSeqPAV does not satisfy PJR or EJR.*

The proposition shows that RevSeqPAV does not satisfy *JR* if $k \geq 5$. Interestingly, SeqPAV does satisfy *JR* as long as $k \leq 5$ [6]. We note that for small k, SeqPAV has two advantages over RevSeqPAV: it requires only a few rounds of candidate inclusions and may be more likely to satisfy *JR*. On the other hand, for k almost equal to $|C|$, RevSeqPAV has two advantages over SeqPAV: it requires only a few rounds of candidate removals and may be more likely to satisfy *JR*.

References

[1] Aziz, H., Brill, M., Conitzer, V., Elkind, E., Freeman, R., Walsh, T., 2017. Justified representation in approval-based committee voting. Social Choice and Welfare.

[2] Aziz, H., Gaspers, S., Gudmundsson, J., Mackenzie, S., Mattei, N., Walsh, T., 2015. Computational aspects of multi-winner approval voting. In: Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS). IFAAMAS, pp. 107–115.

[3] Aziz, H., Huang, S., 2017. A polynomial-time algorithm to achieve extended justified representation. Tech. Rep. 1703.10415, arXiv.org.

[4] Janson, S., 2016. Phragmén's and Thiele's election methods. Tech. Rep. arXiv:1611.08826 [math.HO], arXiv.org.

[5] Kilgour, D. M., 2010. Approval balloting for multi-winner elections. In: Laslier, J.-F., Sanver, M. R. (Eds.), Handbook on Approval Voting. Springer, Ch. 6, pp. 105–124.

[6] Sánchez-Fernández, L., Elkind, E., Lackner, M., Fernández, N., Fisteus, J. A., Basanta Val, P., Skowron, P., 2017. Proportional justified representation. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI). AAAI Press, forthcoming.

[7] Skowron, P. K., Faliszewski, P., Lang, J., 2016. Finding a collective set of items: From proportional multirepresentation to group recommendation. Artificial Intelligence 241, 191–216.