Approximation Algorithms for Optimization Problems with Justified Representation Constraints

Zhihao Jiang[1] , **Ashish Goel**[1]
[1]Stanford University
{faebdc, ashishg}@stanford.edu,

Abstract

Justified representation (JR) is a crucial fairness concept within the context of committee selection in the approval voting setting. When multiple committees satisfy JR, an additional quality measurement is required to select a committee among all JR committees. Optimization problems over all JR committees are NP-hard in many cases. In this paper, we consider approximation algorithms for optimization problems subject to JR constraints. Specifically, we explore algorithms and complexity related to minimizing total committee member costs and maximizing social welfare over JR committees. Our cost-minimization algorithm also works in the EJR$^+$ and BJR setting. Our approach employs techniques such as linear programming rounding and reduction of hypergraph problems to our context.

1 Introduction

Committee selection, or multiwinner voting, is a fundamental problem in computational social choice. The central objective is to select a subset of K candidates from a larger set of m alternatives, in a way that reflects the preferences of a population of n voters. A widely adopted model for capturing voter preferences is approval voting, where each voter submits a binary ballot indicating which candidates they approve.

A central normative goal in this setting is proportional representation. Roughly speaking, if a large enough group of voters shares similar preferences, then these preferences should be reflected in the selected committee. This idea is formalized by Aziz et al. [Aziz *et al.*, 2017], who introduce the notion of justified representation (JR). A committee satisfies JR if every group of at least n/K voters who all approve a common candidate sees at least one of their approved candidates included in the committee. Aziz et al. further propose extended justified representation (EJR), a strengthening of JR that guarantees more extensive representation for larger cohesive groups.

Subsequent work has refined and extended these proportionality guarantees. Brill and Peters [Brill and Peters, 2023] introduce EJR$^+$, a tractable enhancement of EJR that preserves many of its proportionality properties while ensuring easier verification. Fish et al. [Fish *et al.*, 2023] propose balanced justified representation (BJR), which requires that every candidate in the committee is backed by at least some approving voters. These axioms are not only normatively appealing but also practically relevant. While verifying whether a committee satisfies EJR is coNP-hard [Aziz *et al.*, 2017], JR and EJR$^+$ can be efficiently checked. Verification of BJR is also tractable under additional certificate conditions. These considerations have led to a growing interest in voting rules such as Proportional Approval Voting (PAV) [Thiele, 1895] and the Method of Equal Shares (MES) [Peters and Skowron, 2020; Peters *et al.*, 2021].

Beyond satisfying proportionality axioms, many real-world applications impose further constraints or optimization goals. For example, in participatory budgeting, each candidate may have an associated cost, and the selected committee must fit within a budget. In such contexts, it is natural to ask whether we can find committees that satisfy fairness guarantees such as JR, while also minimizing total cost or maximizing social welfare.

Our work addresses this question by studying optimization problems under relaxed proportionality constraints. We build on recent work by Elkind et al. [Elkind *et al.*, 2023], who investigate the task of finding minimum-size JR-satisfying groups. We generalize their model to a cost-based setting, where each candidate has a nonnegative cost, and the objective is to minimize the total cost of a committee that approximately satisfies JR. Specifically, we aim to find a committee that satisfies a relaxed JR condition—requiring that each group of size $n/[(1-\delta)K]$ be represented—while achieving a cost that is within $O(\log n/\delta)$ of the minimum cost required to satisfy standard JR. Our approach matches known lower bounds for JR-satisfying committee selection without relaxation.

To obtain this result, we formulate a linear programming relaxation of the cost minimization problem, and design a rounding algorithm that converts fractional solutions into valid committees. A key technical challenge is encoding JR-like constraints in the LP formulation, which we address through a set of carefully constructed conditions. The framework extends to the EJR$^+$ setting for $\delta \geq 0.5$, and to the BJR setting for any $\delta > 0$, with appropriate modifications to the LP.

We also study the complementary problem of welfare maximization. Here the objective is to select a committee with high total voter approval, subject to proportionality constraints. Prior work [Elkind *et al.*, 2022] shows that it is NP-hard to approximate the maximum social welfare of a JR-satisfying committee of size K within a factor better than $O(K^{0.5-\epsilon})$ for any $\epsilon > 0$. We consider an alternative relaxation: allowing a slightly larger committee of size $(1+\delta)K$, we show that a simple greedy algorithm guarantees welfare at least δ times the optimal value. Moreover, assuming the Unique Games Conjecture [Khot, 2002], we prove that this approximation guarantee is tight. The hardness reduction proceeds via intermediate problems on hypergraphs, including hypergraph independent set and hyperedge dominating set, under specific parameter regimes.

1.1 Problem Setting and Results

There is a set of voters $N = [n] = \{1, 2, \ldots, n\}$, and a set of candidates $C = [m]$. Each voter $v \in N$ submits a ballot $A_v \subseteq C$. If $c \in A_v$ for some $c \in C$ and $v \in N$, we say v *approves* c. For candidate $c \in C$, we define $B_c = \{v \in N : c \in A_v\}$ denoting the set of voters who approve c. Each candidate $c \in C$ is associated with a cost $f(c)$. An instance of the problem can be written as $I = (C, \mathcal{A}, f)$, where $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ is the list of ballots from all voters.

A committee is a subset of candidates, and the goal is to find a committee $W \subseteq C$. We measure the quality of a committee in the following two ways.

Definition 1 (committee cost)**.** *Given an instance $I = (C, \mathcal{A}, f)$ and a committee $W \subseteq C$. We define the cost of W as*

$$f(W) = \sum_{c \in W} f(c).$$

Definition 2 (social welfare)**.** *Given an instance $I = (C, \mathcal{A}, f)$ and a committee $W \subseteq C$. We define the social welfare of W as*

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

Besides committee quality, we also care about fairness. We consider the following criteria.

Definition 3 (JR [Aziz *et al.*, 2017])**.** *Given an instance $I = (C, \mathcal{A}, f)$ and positive integer K, we say committee W provides* justified representation *(JR), if there is no candidate $c \in C \setminus W$, group of voters $N' \subset N$ with $|N'| \geq n/K$ such that*

$$c \in \bigcap_{v \in N'} A_v \quad \textit{and} \quad A_v \cap W = \emptyset \textit{ for all } v \in N'.$$

A committee provides JR if, for any n/K voters who have a common approving candidate, at least one of them is represented, where a voter is represented if at least one candidate in her approving set is in the committee.

Let $\mathrm{JR}_K(I)$ represent the set of all committees that provide JR given I and K, which corresponds to the notion of n/K-justifying groups in [Bredereck *et al.*, 2019]. We define $\mathrm{JR}_{K,s}(I)$ as the subset of $\mathrm{JR}_K(I)$ containing all committees of size at most s, where $\mathrm{JR}_{K,K}(I)$ corresponds to the notion of JR given K in [Aziz *et al.*, 2017].

We define

$$f_{\mathrm{JR}}(I, K) = \min_{W \in \mathrm{JR}_K(I)} f(W). \tag{1}$$

Computing $f_{\mathrm{JR}}(I, K)$ is NP-hard. Elkind *et al.* showed the following hardness result.

Theorem 1 (Theorem A.1[1] in [Elkind *et al.*, 2023])**.** *For $\epsilon > 0$, the following problem is NP-hard. Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, find a committee $W \in JR_K(I)$ such that $f(W) \leq \big((1-\epsilon)\log(n)\big)\, f_{JR}(I, K)$.*

We propose an approximation algorithm with relaxation both on committee cost and JR guarantee.

Theorem 2 (proof in Section 2.1)**.** *Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, for any constant $\delta \in (0, 1)$, there is a polynomial-time algorithm that finds $W \in JR_{(1-\delta)K}(I)$, such that*

$$f(W) \leq M(\delta) \cdot \log(n) \cdot f_{JR}(I, K),$$

where $M(\delta)$ is a constant related to δ.

We also propose such an approximation algorithm for EJR^+, which is a more strict fairness guarantee than JR. We first define EJR^+. A committee provides EJR^+ if for any voter set $N' \subset N$ who have a common approving candidate c, either c is in the committee, or at least one voter in N' is represented at least $K|N'|/n$ times, where a voter is represented ℓ times if ℓ candidates in her approving set are in the committee.

Definition 4 (EJR^+ [Brill and Peters, 2023])**.** *Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, we say committee W provides EJR^+, if there is no candidate $c \in C \setminus W$, positive integer ℓ, group of voters $N' \subseteq N$ with $|N'| \geq n\ell/K$ such that*

$$c \in \bigcap_{v \in N'} A_v \quad \textit{and} \quad |A_v \cap W| < \ell \textit{ for all } v \in N'.$$

Let $\mathrm{EJR}^+_K(I)$ represent the set of all committees that provide EJR^+ given I and K. We define $\mathrm{EJR}^+_{K,s}(I)$ as the subset of $\mathrm{EJR}^+_K(I)$ containing all committees of size at most s. We define

$$f_{\mathrm{EJR}^+}(I, K) = \min_{W \in \mathrm{EJR}^+_K(I)} f(W).$$

Our approximation algorithm has the following guarantee.

Theorem 3 (proof in Section 2.2)**.** *Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, there is a polynomial-time algorithm that finds $W \in EJR^+_{K/2}(I)$, such that*

$$f(W) \leq M \cdot \log(nm) \cdot f_{EJR^+}(I, K),$$

where M is a constant.

[1] This theorem shows the problem is NP-hard even if all the candidates have same cost, *i.e.*, $f(c) = 1$ for every $c \in C$.

A recently introduced concept, BJR, extends to a more general setting where voters have cardinal utilities for candidates. The approval-based voting model we previously considered is a special case, where utilities take values of only 0 or 1.

Instead of submitting a ballot $A_v \subseteq C$, each voter $v \in N$ submits a utility function $u_v : C \to \mathbb{R}$, specifying her utility for each candidate. An instance of the BJR problem can thus be represented as $I = (C, \mathcal{U}, f)$, where $\mathcal{U} = (u_1, u_2, \dots, u_n)$ denotes the collection of utility vectors from all voters. Furthermore, in this case, a candidate may be selected multiple times, so actually $w \in \{0, 1, \dots\}^C$. For convenience, we continue to use $c \in W$ to enumerate all candidate copies in the committee.

A committee satisfies BJR if there exists an assignment of voters to committee members such that each committee member is matched with (approximately) the same number of voters, and for any sufficiently large group of voters who highly value a particular candidate, at least one voter in the group is matched to a candidate they also value highly. The formal definition of BJR is provided below.

Definition 5 (BJR [Fish *et al.*, 2023][2]). *Given an instance $I = (C, \mathcal{U}, f)$ and a positive integer K, we say committee W provides BJR, if there is a function $\omega : N \to W$, matching voters to candidates such that each candidate in W is matched to at most $\lceil n/K \rceil$ voters, for which there is no coalition $N' \subseteq N$ of size at least n/K, candidate $c \in C$, and threshold $\vartheta \in \mathbb{R}$, such that*

$$u_v(c) \geq \vartheta \text{ for all } v \in N' \text{ and } u_v(\omega(v)) < \vartheta \text{ for all } v \in N'.$$

Let $\text{BJR}_K(I)$ represent the set of all committees that provide BJR given I and K. We define $\text{BJR}_{K,s}(I)$ as the subset of $\text{BJR}_K(I)$ containing all committees of size at most s. We define

$$f_{\text{BJR}}(I, K) = \min_{W \in \text{BJR}_K(I)} f(W).$$

We also propose an approximation algorithm for BJR, which has similar guarantee as JR.

Theorem 4 (proof in Section 2.3)**.** *Given an instance $I = (C, \mathcal{U}, f)$ and a positive integer K, for any constant $\delta \in (0, 1)$, there is a polynomial-time algorithm that finds $W \in BJR_{(1-\delta)K}(I)$, such that*

$$f(W) \leq M(\delta) \cdot \log(nm) \cdot f_{BJR}(I, K),$$

where $M(\delta)$ is a constant related to δ.

Besides the minimization problem on committee cost, we also consider the maximization problem on social welfare. We define

$$\text{sw}_{\text{JR}}(I, K) = \max_{W \in \text{JR}_{K,K}(I)} \text{sw}(W).$$

Computing $\text{sw}_{\text{JR}}(I, K)$ is NP-hard. Elkind *et al.* showed the following hardness result.

Theorem 5 (Theorem 4.1 in [Elkind *et al.*, 2022])**.** *For $\epsilon > 0$, the following problem is NP-hard. Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, find a committee $W \in JR_{K,K}(I)$ such that $sw(W) \geq sw_{JR}(I, K)/K^{0.5-\epsilon}$.*

[2] A modified version of the definition in [Fish *et al.*, 2023], structured to simplify the development of approximation results.

To achieve better committee quality, we relax the constraint on committee size.

Theorem 6 (proof in Section 3.1)**.** *Given an instance $I = (C, \mathcal{A}, f)$, a positive integer K, and $\delta \in (0, 1]$ such that δK is an integer, there is a polynomial-time algorithm that finds $W \in JR_{K,(1+\delta)K}(I)$, such that*

$$sw(W) \geq \delta \cdot sw_{JR}(I, K).$$

While the proof of Theorem 6 is via a straight-forward greedy algorithm, we also show the surprising (and non-trivial) result that this simple algorithm is nearly optimal:

Theorem 7 (proof in Section 3.2)**.** *Assuming the Unique Games Conjecture, for any $\delta \in (0, 1)$ and $\epsilon > 0$, the following problem is NP-hard. Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, find $W \in JR_{K,(1+\delta)K}(I)$ such that*

$$sw(W) \geq (\delta + \epsilon) \cdot sw_{JR}(I, K).$$

1.2 Additional Related Work

In addition to the concepts we have already described (JR, EJR, and EJR$^+$), several other variants of JR have been studied, such as *proportional justified representation* (PJR) [Sánchez-Fernández *et al.*, 2017; Aziz *et al.*, 2018], *Strong Justified Representation* (SJR) [Aziz *et al.*, 2017], *average justified representation* (AJR) [Sánchez-Fernández *et al.*, 2017], *full justified representation* (FJR) [Peters *et al.*, 2021], *etc*.

As for the trade-off between fairness and committee quality, Lackner and Skowron [Lackner and Skowron, 2019; Lackner and Skowron, 2020] consider the trade-off between notions of social welfare and representation in the committee selection problem. They start with two voting rules, 1) *Approval Voting* (AV) which selects the committee that maximizes social welfare, and this voting rule cares more about committee quality. 2) *Approval Chamberlin–Courant* (CC) [Thiele, 1895; Chamberlin and Courant, 1983] which selects the committee that maximizes the number of voters that are covered, and this voting rule cares more about fairness. They evaluate other voting rules in these two perspectives, *i.e.*, social welfare and the number of covered voters. Kocot *et al.* [Kocot *et al.*, 2019] consider ordinal elections and introduce a framework where committees are evaluated not just on one goal, but on a set of goals that may have varying degrees of importance. They propose algorithms and analyze their effectiveness in finding committees that meet these multigoal criteria. Fairstein *et al.* [Fairstein *et al.*, 2022] study such trade-offs in the broader context of participatory budgeting. Maly *et al.* [Maly *et al.*, 2022] introduce a new notion of fairness in participatory budgeting, and consider the trade-off between fairness and social welfare.

Another notion of fairness is *core stability* [Droop, 1881; Fain *et al.*, 2018; Lindahl, 1958; Scarf, 1967; Thiele, 1895]. It is similar to EJR in our approval set setting, but its requirement is stronger. For EJR, a guarantee is required only for *coherent* groups, whereas the core requires it for all groups, making core stability a much harder property to satisfy. There is no polynomial time algorithm known for computing a single committee that is core stable for our setting of multiwinner selection under approval voting, but for EJR, as we mentioned before, voting rules PAV and MES satisfy EJR.

1.3 Roadmap

In Section 2, we study the cost minimization problem under the constraints of JR, EJR$^+$, and BJR, and design corresponding approximation algorithms. In Section 3, we address the problem of maximizing social welfare subject to the JR guarantee, presenting a simple approximation algorithm and proving its optimality in a certain sense.

2 Minimizing Committee Cost

In this section, we consider the problem of minimizing committee cost, with JR guarantee (Section 2.1), EJR$^+$ guarantee (Section 2.2) and BJR guarantee (Section 2.3). In both settings, our method consists of the following three steps.

1. Formulate the optimization problem as a binary integer linear programming (BILP) problem.
2. Consider the relaxation of the BILP problem, which is a linear programming (LP) problem.
3. Solve the LP problem, and then round the obtained solution to get a solution for the original BILP problem.

In Appendix C, we have more discussion on the LP rounding method, which provides additional insight.

2.1 Approximation Algorithm with JR Guarantee

We propose an approximation algorithm with properties described in Theorem 2. Recall that given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, the problem is to find a committee $W \in \text{JR}_K(I)$ that minimizes the committee cost.

We formulate the problem as a BILP problem.

For each candidate $c \in C$, define a binary variable $y_c \in \{0,1\}$ as the indicator that $c \in W$, so the objective is to minimize

$$f(W) = \sum_{c \in C} f(c) y_c.$$

For each voter $v \in N$, define a binary variable $x_v \in \{0,1\}$ as the indicator that voter v is represented, so the relation between $\{x_v\}_{v \in N}$ and $\{y_c\}_{c \in C}$ can be written as

$$x_v \leq \sum_{c \in A_v} y_c.$$

To ensure $W \in \text{JR}_K(I)$, a straight-forward LP formulation would mimic the JR definition by having a constraint for each set, that is, for any $N' \subset N$ and $c \in C$ where $|N'| \geq n/K$ and $c \in \bigcap_{v \in N'} A_v$, there is a constraint $\sum_{c' \in \bigcup_{v \in N'} A_v} x_{c'} \geq 1$. Although there are exponentially many constraints, it is still tractable via a simple separation oracle. However, it would preclude using Chernoff bounds to get approximation guarantees after randomized rounding. As a result, we use an alternate formulation of these constraints, as given below.

For candidate $c \in C$, at most $\left(\lceil \frac{n}{K} \rceil - 1\right)$ voters in B_c is not represented, so we have constraint

$$\sum_{v \in B_c} x_v \geq |B_c| - \left\lceil \frac{n}{K} \right\rceil + 1.$$

In summary, the problem of finding a committee $W \in \text{JR}_K(I)$ that minimizes $f(W)$ can be formulated as the following BILP problem (left), and we consider its LP relaxation (right).

$$\text{minimize} \quad \sum_{c \in C} f(c) \cdot y_c \tag{2a}$$

$$s.t. \quad x_v \leq \sum_{c \in A_v} y_c, \quad \forall v \in N \tag{2b}$$

$$\sum_{v \in B_c} x_v \geq |B_c| - \left\lceil \frac{n}{K} \right\rceil + 1, \quad \forall c \in C \tag{2c}$$

$$x_v \in \{0,1\}, \quad \forall v \in N \tag{2d}$$

$$y_c \in \{0,1\}, \quad \forall c \in C \tag{2e}$$

$$\text{minimize} \quad \sum_{c \in C} f(c) \cdot y_c \tag{3a}$$

$$s.t. \quad x_v \leq \sum_{c \in A_v} y_c, \quad \forall v \in N \tag{3b}$$

$$\sum_{v \in B_c} x_v \geq |B_c| - \left\lceil \frac{n}{K} \right\rceil + 1, \quad \forall c \in C \tag{3c}$$

$$0 \leq x_v \leq 1, \quad \forall v \in N \tag{3d}$$

$$0 \leq y_c \leq 1, \quad \forall c \in C \tag{3e}$$

Suppose x_v^* and y_c^* is the optimal solution of the LP relaxation. Then our rounding method computes integer solution $\hat{x}_v$ and $\hat{y}_c$, such that the committee W induced by $\hat{y}_c$ is in $\text{JR}_{(1-\delta)K}(I)$, in the following way.

- For $c \in C$, let $\hat{y}_c = 1$ with probability $\min\{1, 2\log(n) y_c^*/\delta\}$, otherwise $\hat{y}_c = 0$. This rounding procedure is independent for each $c \in C$.
- For $v \in N$, let $\hat{x}_v = 1$ if there exists $c \in A_v$ such that $\hat{y}_c = 1$, otherwise $\hat{x}_v = 0$.

This integer solution satisfies the constraints 2b-2e in the BILP, except the JR constraint 2c. We use $W_{\hat{y}}$ to denote the committee induced by $\{\hat{y}_c\}_{c \in C}$, meaning $W_{\hat{y}}$ contains candidate $c \in C$ if $\hat{y}_c = 1$. Although it may be false that $W_{\hat{y}} \in \text{JR}_K(I)$, we show a weaker guarantee $W_{\hat{y}} \in \text{JR}_{(1-\delta)K}(I)$ for any constant $\delta \in (0,1)$. We start with the following lemma.

Lemma 1. *For $v \in N$, if $x_v^* \geq \delta$, then $\mathbf{Pr}(\hat{x}_v = 0) \leq \frac{1}{n^2}$.*

Proof. We assume for any $c \in A_v$, $2\log(n) y_c^*/\delta \leq 1$, otherwise $\hat{x}_v = 1$ for sure.

Since $x_v \geq \delta$, which implies $\sum_{c \in A_v} y_c^* \geq \delta$, and we have

$$\begin{aligned}\mathbf{Pr}(\hat{x}_v = 0) = \prod_{c \in A_v} \mathbf{Pr}(\hat{y}_c = 0) = \prod_{c \in A_v} \left(1 - \frac{2\log(n) \cdot y_c^*}{\delta}\right) \\ \leq \left(1 - \frac{2\log(n)}{|A_v|}\right)^{|A_v|} \leq e^{-2\log(n)} = \frac{1}{n^2},\end{aligned}$$

where the inequality connecting the first line and the second line is based on the AM–GM inequality. □

Applying the Union Bound over $v \in N$, we have

- With probability at least $1 - 1/n$, for any $x_v \geq \delta$, $\hat{x}_v = 1$.

For each $c \in C$, since $|B_c| - \sum_{v \in B_c} x_v \leq \lceil n/K \rceil - 1$, there are at most $(\lceil n/K \rceil - 1)/(1-\delta)$ voters whose corresponding variable x_v^* is smaller than δ, so we have

- With probability at least $1 - 1/n$, for any $c \in C$,

$$|B_c| - \sum_{v \in B_c} \hat{x}_v \leq \frac{\lceil \frac{n}{K} \rceil - 1}{1-\delta} < \left\lceil \frac{n}{(1-\delta)K} \right\rceil. \quad (4)$$

This fact concludes $W_{\hat{y}} \in \text{JR}_{(1-\delta)K}(I)$ with probability at least $1 - 1/n$. As for the committee cost, we have the following observation.

$$\mathbb{E}\left[f(W_{\hat{y}})\right] = \mathbb{E}\left[\sum_{c \in C} f(c)\hat{y}_c\right] \leq \frac{2\log(n)}{\delta} \cdot \sum_{c \in C} f(c) y_c^* \leq \frac{2\log(n)}{\delta} \cdot f_{\text{JR}}(I, K).$$

By Markov's inequality, this observation implies

$$\mathbf{Pr}\left(f(W_{\hat{y}}) > \frac{3\log(n)}{\delta} \cdot f_{\text{JR}}(I,K)\right) < \frac{2}{3}.$$

Combining this inequality with Inequality 4, by the Union Bound, we have

Corollary 1. *With probability at least* 0.3*, we have* $W_{\hat{y}} \in JR_{(1-\delta)K}(I)$ *and* $f(W_{\hat{y}}) < \frac{3\log(n)}{\delta} \cdot f_{JR}(I,K)$.

Our algorithm repeatedly round $\{y_c^*\}_{c \in C}$ and $\{x_v^*\}_{v \in N}$ into $\{\hat{y}_c\}_{c \in C}$ and $\{\hat{x}_v\}_{v \in N}$ for $O(\log(n))$ times, and pick the reduced committee $W_{\hat{y}}$ satisfies $W_{\hat{y}} \in \text{JR}_{(1-\delta)K}(I)$ with smallest $f(W_{\hat{y}})$. Then with probability $1 - O(1/n)$, we are able to find $W_{\hat{y}} \in \text{JR}_{(1-\delta)K}(I)$ satisfying $f(W_{\hat{y}}) < \frac{3\log(n)}{\delta} \cdot f_{\text{JR}}(I,K)$. These arguments complete the proof of Theorem 2.

2.2 Approximation Algorithm with EJR$^+$ Guarantee

We propose an approximation algorithm with properties described in Theorem 3. Recall that given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, the problem is to find a committee $W \in \text{EJR}_K^+(I)$ that minimizes the committee cost.

The algorithm and analysis follow a framework similar to that in Section 2.1, but the LP formulation is more elaborate and the analysis requires more details.

We first formulate the problem as a BILP problem.

For each candidate $c \in C$, define a binary variable $y_c \in \{0,1\}$ as the indicator that $c \in W$, so the objective is to minimize

$$f(W) = \sum_{c \in C} f(c) y_c.$$

For each voter $v \in N$ and positive integer $\ell \in [K]$, define a binary variable $x_{v,\ell} \in \{0,1\}$ as the indicator that voter $|A_v \cap W| \geq \ell$, so the relation between $\{x_v\}_{v \in N}$ and $\{y_c\}_{c \in C}$ can be written as

$$\ell \cdot x_{v,\ell} \leq \sum_{c \in A_v} y_c.$$

To ensure $W \in \text{EJR}_K^+(I)$, for candidate $c \in C$ and positive integer $\ell \in [K]$, we need

$$y_c = 1 \quad \text{OR} \quad \left|\left\{v \in B_c : x_{v,\ell} = 0\right\}\right| \leq \left\lceil \frac{n\ell}{K} \right\rceil - 1,$$

and we can write it as the following constraint.

$$\sum_{v \in B_c} x_{v,\ell} \geq \left(|B_c| - \left\lceil \frac{n\ell}{k} \right\rceil + 1\right)(1 - y_c).$$

In summary, the problem of finding a committee $W \in \text{EJR}_K^+(I)$ that minimizes $f(W)$ can be formulated as the following BILP problem.

$$\begin{aligned}
\text{minimize} \quad & \sum_{c \in C} f(c) y_c && (5)\\
s.t. \quad & \ell \cdot x_{v,\ell} \leq \sum_{c \in A_v} y_c, \quad \forall v \in N, \ell \in [K] && (6)\\
& \sum_{v \in B_c} x_{v,\ell} \geq \left(|B_c| - \left\lceil \frac{n\ell}{K} \right\rceil + 1\right)(1 - y_c), && (7)\\
& \qquad \forall c \in C, \ell \in [K] && (8)\\
& x_{v,\ell} \in \{0,1\}, \quad \forall v \in N, \ell \in [K] && (9)\\
& y_c \in \{0,1\}, \quad \forall c \in C && (10)
\end{aligned}$$

If we directly replace $x_{v,\ell} \in \{0,1\}$ and $y_c \in \{0,1\}$ by $0 \leq x_{v,\ell} \leq 1$ and $0 \leq y_c \leq 1$ respectively, the rounding steps may face some problems. However, a modified formulation with constraints derived from a different perspective allows for successful rounding. We defer the full details and proofs to Appendix A.

2.3 Approximation Algorithm with BJR Guarantee

We propose an approximation algorithm with properties described in Theorem 4. Recall that given an instance $I = (C, \mathcal{U}, f)$ and a positive integer K, the problem is to find a committee $W \in \text{BJR}_K(I)$ that minimizes the committee cost.

We formulate the problem as a BILP problem.

For each candidate $c \in C$, define a binary variable $y_c \in \{0,1\}$ as the indicator that $c \in W$, so the objective is to minimize

$$f(W) = \sum_{c \in C} f(c) y_c.$$

Recall that the definition of BJR involves a matching. Let $z_{v,c}$ denote the assignment of voter v to candidate c within this matching, which must satisfy

$$\sum_{c} z_{v,c} \leq 1, \quad \forall v \in N,$$

$$\sum_{v} z_{v,c} \leq \frac{n}{K} y_c, \quad \forall c \in C.$$

For each voter $v \in N$ and threshold ϑ, define a binary variable $x_{v,\vartheta} \in \{0,1\}$ as the indicator that voter v has utility less than ϑ in the matching. Note that the utilities take at most nm distinct values, so there are at most $nm+1$ essential choices for ϑ, and we use Θ to denote the set of ϑ's. Then, the following relations hold.

$$x_{v,\vartheta} \geq 1 - \sum_{c:u_v(c)\geq\vartheta} z_{v,c}, \quad \forall v \in N, \vartheta \in \Theta.$$

To ensure $W \in \mathrm{BJR}_K(I)$, we have constraint

$$\sum_{v:u_v(c)\geq\vartheta} x_{v,\vartheta} \leq \left\lceil \frac{n}{K} \right\rceil - 1, \quad \forall c \in C, \vartheta \in \Theta.$$

In summary, the problem of finding a committee $W \in \mathrm{BJR}_K(I)$ that minimizes $f(W)$ can be formulated as the following BILP problem.

$$\text{minimize} \quad \sum_{c\in C} f(c)\cdot y_c \tag{11a}$$

$$s.t. \quad \sum_{c} z_{v,c} \leq 1, \quad \forall v \in N \tag{11b}$$

$$\sum_{v} z_{v,c} \leq \frac{n}{K} y_c, \quad \forall c \in C \tag{11c}$$

$$x_{v,\vartheta} \geq 1 - \sum_{c:u_v(c)\geq\vartheta} z_{v,c}, \quad \forall v \in N, \vartheta \in \Theta \tag{11d}$$

$$\sum_{v:u_v(c)\geq\vartheta} x_{v,\vartheta} \leq \frac{n}{K} - 1, \quad \forall c \in C, \vartheta \in \Theta \tag{11e}$$

$$z_{v,c} \in \{0,1\}, \quad \forall v \in N, c \in C \tag{11f}$$

$$x_{v,\vartheta} \in \{0,1\}, \quad \forall v \in N, \vartheta \in \Theta \tag{11g}$$

$$y_c \in \{0,1,\dots\}, \quad \forall c \in C \tag{11h}$$

We consider its LP relaxation.

$$\text{minimize} \quad \sum_{c\in C} f(c)\cdot y_c \tag{12a}$$

$$s.t. \quad \sum_{c} z_{v,c} \leq 1, \quad \forall v \in N \tag{12b}$$

$$\sum_{v} z_{v,c} \leq \frac{n}{K} y_c, \quad \forall c \in C \tag{12c}$$

$$x_{v,\vartheta} \geq 1 - \sum_{c:u_v(c)\geq\vartheta} z_{v,c}, \quad \forall v \in N, \vartheta \in \Theta \tag{12d}$$

$$\sum_{v:u_v(c)\geq\vartheta} x_{v,\vartheta} \leq \frac{n}{K} - 1, \quad \forall c \in C, \vartheta \in \Theta \tag{12e}$$

$$z_{v,c} \geq 0, \quad \forall v \in N, c \in C \tag{12f}$$

$$x_{v,\vartheta} \geq 0, \quad \forall v \in N, \vartheta \in \Theta \tag{12g}$$

$$y_c \geq 0, \quad \forall c \in C \tag{12h}$$

The main difference between this subsection and Section 2.1 is the LP formulation, and the remaining proof of rounding guarantee is very similar, we defer the proof to Appendix B.

Algorithm 1

Input: An instance $I = (C, \mathcal{A}, f)$, a positive integer K, and $\delta \in (0,1)$ such that δK is an integer
Output: Committee W

1: Let W_1 be an arbitrary committee in $\mathrm{JR}_{K,K}(I)$. (Lemma 2)
2: Let W_2 be the set of δK candidates in $C \setminus W_1$ with largest social welfare ($|B_c|$ for candidate c).
3: **return** $W = W_1 \cup W_2$

3 Maximization of Social Welfare

In this section, we consider the maximization problem of social welfare, with JR guarantee. Our positive result is a straight-forward approximation algorithm, which finds a committee W in $\mathrm{JR}_{K,(1+\delta)K}(I)$, such that $\mathrm{sw}(W) \geq \delta \cdot \mathrm{sw}_{\mathrm{JR}}(I,K)$. Our hardness result shows this straight-forward algorithm is optimal in some sense, assuming Unique Game Conjecture.

3.1 Algorithm

In this subsection, we propose an algorithm to prove Theorem 6. We start with the following lemma.

Lemma 2 ([Aziz *et al.*, 2017]). *Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, there is an algorithm that finds a committee $W \in JR_{K,K}(I)$.*

Next, we state our algorithm.

Next, we analyze this algorithm. Since $W_1 \in \mathrm{JR}_{K,K}(I)$ and $|W_2| = \delta K$, we have $W \in \mathrm{JR}_{K,(1+\delta)K}(I)$. As for $\mathrm{sw}(W)$, since W_2 contains δK candidates in $C \setminus W_1$ with largest social welfare, W contains δK candidates in C with largest social welfare, so we have $\mathrm{sw}(W) \geq \delta \cdot \mathrm{sw}_{\mathrm{JR}}(I,K)$. These arguments complete the proof of Theorem 6.

The algorithm above also works for EJR and EJR$^+$, because we can always select K candidates to satisfy EJR or EJR$^+$, and then select another δK candidates with the largest social welfare.

3.2 Hardness

In this subsection, we prove Theorem 7. Assuming the Unique Game Conjecture, we reduce the unique label cover problem to our social welfare maximization problem through the following route.

unique label cover $\to$ hypergraph independent set
$\to$hyperedge dominating set $\to$ our problem

We first briefly introduce hypergraph. A hypergraph is a generalization of a graph where an edge (usually referred to as a hyperedge to distinguish it from an edge in a graph) can join any number of vertices. A hypergraph $H = (V, E)$ consists of a vertex set V and a hyperedge set E, and a hyperedge $e \in E$ is a subset of the vertex set V, with e containing at least 2 vertices. If $v \in e$, we say hyperedge e *covers* vertex v. We also have the following definitions of hypergraph.

Definition 6 (r-uniform hypergraph)**.** *Hypergraph $H = (V, E)$ is a r-uniform hypergraph, if the size of every hyperedge is r, that is, for any $e \in E$, $|e| = r$.* [3]

Definition 7 (independent set)**.** *Given a hypergraph $H = (V, E)$, a set of vertices $S \subset V$ forms an independent set if for every $e \in E$, $e \nsubseteq S$. We use $\alpha(H)$ to denote the size of the maximum independent set of $H = (V, E)$.*

Definition 8 (hyperedge dominating set)**.** *Given a hypergraph $H = (V, E)$, we say hyperedges $e_1, e_2 \in E$ are* adjacent *if $e_1 \cap e_2 \neq \emptyset$. A set of hyperedges $T \subset E$ forms an edge dominating set if, for any $e \in E$, there exists $e' \in T$ that is adjacent to e. We use $\beta(H)$ to denote the size of the minimum edge dominating set of H.*

Khot and Regev [Khot and Regev, 2008] reduce the unique label cover problem to the hypergraph independent set problem, and show the hardness of the hypergraph independent set problem assuming the Unique Games Conjecture. We skip the introduction to the unique label cover problem and directly use their result.

Lemma 3 ([Khot and Regev, 2008])**.** *Assuming the Unique Games Conjecture, for any $\epsilon > 0$ and positive integer $r \geq 2$, the following problem is NP-hard. Given a r-uniform hypergraph $H = (V, E)$, distinguish between*

1. *(YES) $\alpha(H) \geq (1 - \frac{1}{r} - \epsilon)|V|$,*
2. *(NO) $\alpha(H) \leq \epsilon|V|$.*

Next, we show the minimum edge dominating set problem is also hard, by reducing the hypergraph independent set problem to it.

Lemma 4. *Assuming the Unique Games Conjecture, for any $\epsilon > 0$ and positive integer $r \geq 2$, the following problem is NP-hard. Given a r-uniform hypergraph $H = (V, E)$, distinguish between*

1. *(YES) $\beta(H') \leq \left(\frac{1}{r(r-1)} + \epsilon\right)|V'|$,*
2. *(NO) $\beta(H') \geq \frac{1-\epsilon}{r+3}|V'|$.*

We defer the proof to Appendix D.1

Finally, we show our social welfare maximization problem is also hard, by reducing the hyperedge dominating set problem to it.

Lemma 5. *Assuming the Unique Games Conjecture, for any $\epsilon > 0$, positive integer $r \geq 2$, and $\delta > 0$, the following problem is NP-hard. Given an instance $I = (C, \mathcal{A}, f)$, distinguish between*

1. *(YES) $sw_{JR}(I, K) \geq \left(\frac{r-2}{r(r-1)} - \epsilon\right) t\sqrt{t}$,*
2. *(NO) $\max_{W \in JR_{K,(1+\delta)K}(I)} \leq \left(\frac{\delta}{r} + \frac{\epsilon r+3}{r^2}\right) t\sqrt{t} + (\delta + 3)t$,*

where

- *recall that n is the number of voters and $\mathcal{A} = (A_1, \ldots, A_n)$,*
- *$K = n/r$,*
- *t is the positive real number satisfying $t + \sqrt{t} = n$.*

We defer the proof to Appendix D.2.

Since r can be arbitrarily large and ϵ can be arbitrarily small, we have

Corollary 2. *Assuming the Unique Games Conjecture, for any $\delta \in (0, 1)$ and $\epsilon > 0$, the following problem is NP-hard. Given an instance $I = (C, \mathcal{A}, f)$ and a positive integer K, find $W \in JR_{K,(1+\delta)K}(I)$ such that*

$$sw(W) \geq (\delta + \epsilon) \cdot sw_{JR}(I, K).$$

If this corollary does not apply, then the two cases in Lemma 5 are distinguishable. This corollary is identical to Theorem 7.

[3] So a graph is a 2-uniform hypergraph.

References

[Aziz *et al.*, 2017] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017.

[Aziz *et al.*, 2018] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez-Fernández, and Piotr Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Bredereck *et al.*, 2019] Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, and Rolf Niedermeier. An experimental view on committees providing justified representation. In *IJCAI*, pages 109–115, 2019.

[Brill and Peters, 2023] Markus Brill and Jannik Peters. Robust and verifiable proportionality axioms for multiwinner voting. *arXiv preprint arXiv:2302.01989*, 2023.

[Chamberlin and Courant, 1983] John R Chamberlin and Paul N Courant. Representative deliberations and representative decisions: Proportional representation and the borda rule. *American Political Science Review*, 77(3):718–733, 1983.

[Droop, 1881] Henry Richmond Droop. On methods of electing representatives. *Journal of the Statistical Society of London*, 44(2):141–202, 1881.

[Elkind *et al.*, 2022] Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. The price of justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4983–4990, 2022.

[Elkind *et al.*, 2023] Edith Elkind, Piotr Faliszewski, Ayumi Igarashi, Pasin Manurangsi, Ulrike Schmidt-Kraepelin, and Warut Suksompong. Justifying groups in multiwinner approval voting. *Theoretical Computer Science*, 969:114039, 2023.

[Fain *et al.*, 2018] Brandon Fain, Kamesh Munagala, and Nisarg Shah. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 575–592, 2018.

[Fairstein *et al.*, 2022] Roy Fairstein, Reshef Meir, Dan Vilenchik, and Kobi Gal. Welfare vs. representation in participatory budgeting. *arXiv preprint arXiv:2201.07546*, 2022.

[Fish *et al.*, 2023] Sara Fish, Paul Gölz, David C Parkes, Ariel D Procaccia, Gili Rusak, Itai Shapira, and Manuel Wüthrich. Generative social choice. *arXiv preprint arXiv:2309.01291*, 2023.

[Goemans *et al.*, 2009] Michel X Goemans, Nicholas JA Harvey, Satoru Iwata, and Vahab Mirrokni. Approximating submodular functions everywhere. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 535–544. SIAM, 2009.

[Grötschel *et al.*, 2012] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[Khot and Regev, 2008] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[Khot, 2002] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002.

[Kocot *et al.*, 2019] Maciej Kocot, Anna Kolonko, Edith Elkind, Piotr Faliszewski, and Nimrod Talmon. Multigoal committee selection. International Joint Conferences on Artificial Intelligence, 2019.

[Lackner and Skowron, 2019] Martin Lackner and Piotr Skowron. A quantitative analysis of multi-winner rules. In *IJCAI*, pages 407–413, 2019.

[Lackner and Skowron, 2020] Martin Lackner and Piotr Skowron. Utilitarian welfare and representation guarantees of approval-based multiwinner rules. *Artificial Intelligence*, 288:103366, 2020.

[Lindahl, 1958] Erik Lindahl. Just taxation—a positive solution. In *Classics in the theory of public finance*, pages 168–176. Springer, 1958.

[Maly *et al.*, 2022] Jan Maly, Simon Rey, Ulle Endriss, and Martin Lackner. Effort-based fairness for participatory budgeting. *arXiv preprint arXiv:2205.07517*, 2022.

[Peters and Skowron, 2020] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 793–794, 2020.

[Peters *et al.*, 2021] Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. Proportional participatory budgeting with additive utilities. *Advances in Neural Information Processing Systems*, 34:12726–12737, 2021.

[Sánchez-Fernández *et al.*, 2017] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Scarf, 1967] Herbert E Scarf. The core of an n person game. *Econometrica: Journal of the Econometric Society*, pages 50–69, 1967.

[Svitkina and Fleischer, 2011] Zoya Svitkina and Lisa Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. *SIAM Journal on Computing*, 40(6):1715–1737, 2011.

[Thiele, 1895] Thorvald N Thiele. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, 1895:415–441, 1895.

A Omitted proof in Section 2.2

Recall that we have the following LP.

$$\text{minimize} \quad \sum_{c\in C} f(c) y_c \tag{13}$$

$$s.t. \quad \ell \cdot x_{v,\ell} \le \sum_{c\in A_v} y_c, \quad \forall v \in N, \ell \in [K] \tag{14}$$

$$\sum_{v\in B_c} x_{v,\ell} \ge \left(|B_c| - \left\lceil \frac{n\ell}{K} \right\rceil + 1\right)(1-y_c), \tag{15}$$

$$\forall c \in C, \ell \in [K] \tag{16}$$

$$x_{v,\ell} \in \{0,1\}, \quad \forall v \in N, \ell \in [K] \tag{17}$$

$$y_c \in \{0,1\}, \quad \forall c \in C \tag{18}$$

Next, we consider the relaxation of this BILP. If we directly replace $x_{v,\ell} \in \{0,1\}$ and $y_c \in \{0,1\}$ by $0 \le x_{v,\ell} \le 1$ and $0 \le y_c \le 1$ respectively, the rounding steps may face some problems. We consider the following example.

Example 1. *Suppose t is a large even integer, we construct instance $I = (C, \mathcal{A}, f)$ and K in the following way.*

- $n = t^2$, $m = n + 1 = t^2 + 1$, *and* $K = n/2$.
- $A_i = \{i, m\}$ *for* $i \in [n]$.
- *For* $c \in C$, *we have*

$$f(c) = \begin{cases} 1, & c \le t(t-1), \\ t^3, & \textit{otherwise}. \end{cases}$$

For committee $W \in EJR^+_K(I)$, either $m \in W$ or $[m-1] \subset W$, implying $W = \{m\}$ has minimum cost among committees in $EJR^+_K(I)$, so $f_{EJR^+}(I, K) = t^3$.

However, if we run the straight-forward LP relaxation, one feasible solution is

$$y^*_c = \begin{cases} 1, & c \le t(t-1), \\ 0, & t(t-1) \le c \le t^2, \\ \frac{1}{t}, & c = m. \end{cases}$$

This solution provides objective function value $\sum_{c\in C} f(c) y_c < 2t^2$, so the integrality gap goes to infinity as t goes to infinity.

In the example above, A small positive value of $f(m)$ destroys the effect of constraint 16 when $c = m$ and $\ell = 1$.

To avoid this happening, we adopt stronger constraints, without excluding valid integer solutions. Roughly speaking, our method is to set tighter upper bounds for variables, in the following ways.

1. In Constraint 16, for specific $c \in C$ and $\ell \in [K]$, if $y_c > 0$, we aim to also upper bound $x_{v,\ell}$ by $(1 - y_c)$. If this is the case, the upper bounds for the variable $x_{v,\ell}$ will vary depending on the specific $c \in C$. To address this issue, we split the variable $x_{v,\ell}$ into m copies (becomes $x_{v,c,\ell}$), each with a different upper bound.
2. For $x_{v,c,\ell}$, as Constraint 14, we have $\ell \cdot x_{v,c,\ell} \leq \sum_{c' \in A_v} y_{c'}$. We aim to also upper bound $y_{c'}$ by $(1-y_c)$. Using a similar trick, we split $y_{c'}$ into m copies (becomes $z_{c',c}$), each with a different upper bound.

Our improved BILP is

$$\begin{aligned}
&\text{minimize} \quad \sum_{c \in C} f(c) y_c && \text{(19a)}\\
&s.t. \quad z_{c',c} \leq y_{c'}, \quad \forall c \in C, c' \in C && \text{(19b)}\\
&\quad z_{c',c} \leq 1 - y_c, \quad \forall c \in C, c' \in C && \text{(19c)}\\
&\quad \ell \cdot x_{v,c,\ell} \leq \sum_{c' \in A_v} z_{c',c}, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(19d)}\\
&\quad x_{v,c,\ell} \leq 1 - y_c, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(19e)}\\
&\quad \sum_{v \in B_c} x_{v,c,\ell} \geq \left(|B_c| - \left\lceil \frac{n\ell}{K} \right\rceil + 1 \right)(1 - y_c),\\
&\qquad\qquad \forall c \in C, \ell \in [K] && \text{(19f)}\\
&\quad x_{v,c,\ell} \in \{0,1\}, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(19g)}\\
&\quad y_c \in \{0,1\}, \quad \forall c \in C && \text{(19h)}
\end{aligned}$$

And its relaxation is

$$\begin{aligned}
&\text{minimize} \quad \sum_{c \in C} f(c) y_c && \text{(20a)}\\
&s.t. \quad z_{c',c} \leq y_{c'}, \quad \forall c \in C, c' \in C && \text{(20b)}\\
&\quad z_{c',c} \leq 1 - y_c, \quad \forall c \in C, c' \in C && \text{(20c)}\\
&\quad \ell \cdot x_{v,c,\ell} \leq \sum_{c' \in A_v} z_{c',c}, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(20d)}\\
&\quad x_{v,c,\ell} \leq 1 - y_c, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(20e)}\\
&\quad \sum_{v \in B_c} x_{v,c,\ell} \geq \left(|B_c| - \left\lceil \frac{n\ell}{K} \right\rceil + 1 \right)(1 - y_c),\\
&\qquad\qquad \forall c \in C, \ell \in [K] && \text{(20f)}\\
&\quad x_{v,c,\ell} \geq 0, \quad \forall v \in N, c \in C, \ell \in [K] && \text{(20g)}\\
&\quad 0 \leq y_c \leq 1, \quad \forall c \in C && \text{(20h)}
\end{aligned}$$

Suppose $x^*_{v,c,\ell}, y^*_c$ and $z^*_{c',c}$ is the optimal solution of the LP relaxation. Then our rounding method computes integer solution $\hat{y}_c$ (we do not use integer solutions for x^* and z^* throughout the entire proof) in the following way.

- For $c \in C$, let $\hat{y}_c = 1$ with probability $\min\{1, 12\log(nm) y^*_c\}$, otherwise $\hat{y}_c = 0$. This rounding procedure is independent for each $c \in C$.

We use $W_{\hat{y}}$ to denote the committee induced by $\{\hat{y}_c\}_{c \in C}$, meaning $W_{\hat{y}}$ contains candidate $c \in C$ if $\hat{y}_c = 1$. Although it may be false that $W_{\hat{y}} \in \text{EJR}^+_K(I)$, we show a weaker guarantee $W_{\hat{y}} \in \text{EJR}^+_{K/2}(I)$.

We start with the following lemma.

Lemma 6. *For $v \in N$, $c \in C$, $\ell \in [k]$, if $y^*_c \leq 0.5$ and $x^*_{v,c,\ell} \geq 0.5(1 - y^*_c)$, then*

$$\mathbf{Pr}\left(\sum_{c' \in A_v} \hat{y}_{c'} < 0.5\ell \right) \leq \frac{1}{n^2 m^2}.$$

Proof. If $x^*_{v,c,\ell} \geq 0.5(1 - y^*_c)$, then

$$\sum_{c' \in A_v} \min\{y^*_{c'}, 1 - y^*_c\} \geq \sum_{c' \in A_v} z^*_{c',c} \geq 0.5\ell(1 - y^*_c).$$

Define $S = \{c' \in A_v : 12\log(nm) y^*_{c'} \geq 1\}$. Then

$$\sum_{c' \in A_v \setminus S} y^*_{c'} \geq (0.5\ell - |S|)(1 - y^*_c),$$

which implies

$$\begin{aligned}
\mathbb{E}\left[\sum_{c' \in A_v \setminus S} \hat{y}_{c'} \right] \geq & 12\log(nm)(0.5\ell - |S|)(1 - y^*_c)\\
\geq & 6\log(nm)(0.5\ell - |S|).
\end{aligned}$$

We assume $|S| < 0.5\ell$, otherwise $\mathbf{Pr}\left(\sum_{c' \in A_v} \hat{y}_{c'} < 0.5\ell \right) = 0$, which finishes the proof.

Applying Chernoff Bound, we have

$$\begin{aligned}
&\mathbf{Pr}\left(\sum_{c' \in A_v} \hat{y}_{c'} < 0.5\ell \right)\\
=&\mathbf{Pr}\left(\sum_{c' \in A_v \setminus S} \hat{y}_{c'} < 0.5\ell - |S| \right)\\
\leq& \exp\left\{ -\frac{1}{2}\left(1 - \frac{1}{6\log(nm)}\right)^2 \mathbb{E}\left[\sum_{c' \in A_v \setminus S} \hat{y}_{c'} \right] \right\}\\
\leq& \frac{1}{n^2 m^2}.
\end{aligned}$$

□

Applying the Union Bound over $v \in N$, $c \in C$ and $\ell \in [K]$, we have

- With probability at least $1 - 1/n$, for any $v \in V$, $c \in C$, $\ell \in [k]$, if $y^*_c \leq 0.5$ and $x^*_{v,c,\ell} \geq 0.5(1 - y^*_c)$, then $\sum_{c' \in A_v} \hat{y}_{c'} \geq 0.5\ell$.

For each $c \in C$ and $\ell \in [K]$, note that we have the following two properties

1. $\sum_{v \in B_c} x^*_{v,c,\ell} \geq \left(|B_c| - \left\lceil \frac{n\ell}{K} \right\rceil + 1 \right)(1 - y^*_c)$,

2. $x^*_{v,c,\ell} \le (1 - y^*_c)$ for any $v \in N$,

so there are at most $2 \cdot \left(\lceil n\ell/K \rceil - 1\right)$ voters whose corresponding variable $x^*_{v,c,\ell}$ is smaller than $0.5(1 - y^*_c)$, by Lemma 6, we have

- With probability at least $1 - 1/n$, for any $c \in C$, at least one of the following holds.
    1. $y^*_c > 0.5$, implying $\hat{y}_c = 1$.
    2. $\left|\left\{v \in B_c : \sum_{c' \in A_v} \hat{y}_{c'} < 0.5\ell\right\}\right| \le 2 \cdot \left(\lceil n\ell/K \rceil - 1\right)$.

    It implies $W_{\hat{y}} \in \text{EJR}^+_{K/2}(I)$.

This property concludes $W_{\hat{y}} \in \text{EJR}^+_{K/2}(I)$ with probability at least $1 - 1/n$. As for the committee cost, we have the following observation.

$$\begin{aligned}\mathbb{E}\left[f(W_{\hat{y}})\right] = \mathbb{E}\left[\sum_{c \in C} f(c)\hat{y}_c\right] &\le 12\log(nm) \cdot \sum_{c \in C} f(c) y^*_c \\ &\le 12\log(nm) \cdot f_{\text{EJR}^+}(I, K).\end{aligned}$$

By Markov's inequality, this observation implies

$$\mathbf{Pr}\left(f(W_{\hat{y}}) > 13\log(nm) \cdot f_{\text{JR}}(I, K)\right) < \frac{12}{13}. \tag{21}$$

Combining the properties above, by the Union Bound, we have

Corollary 3. *With probability at least $1/15$, we have $W_{\hat{y}} \in$ EJR$^+_{K/2}(I)$ and $f(W_{\hat{y}}) < 13\log(nm) \cdot f_{JR}(I, K)$.*

Our algorithm repeatedly round $\{y^*_c\}_{c \in C}$ into $\{\hat{y}_c\}_{c \in C}$ for $O(\log(n))$ times, and pick the reduced committee $W_{\hat{y}}$ satisfies $W_{\hat{y}} \in \text{EJR}^+_{K/2}(I)$ with smallest $f(W_{\hat{y}})$. Then with probability $1 - O(1/n)$, we are able to find $W_{\hat{y}} \in \text{EJR}^+_{K/2}(I)$ satisfying $f(W_{\hat{y}}) < 13\log(nm) \cdot f_{\text{JR}}(I, K)$. These arguments complete the proof of Theorem 3.

B Omitted proof in Section 2.3

Recall that we have the following LP.

$$\text{minimize} \quad \sum_{c \in C} f(c) \cdot y_c \tag{22a}$$
$$s.t. \quad \sum_{c} z_{v,c} \le 1, \quad \forall v \in N \tag{22b}$$
$$\sum_{v} z_{v,c} \le \frac{n}{K} y_c, \quad \forall c \in C \tag{22c}$$
$$x_{v,\vartheta} \ge 1 - \sum_{c: u_v(c) \ge \vartheta} z_{v,c}, \quad \forall v \in N, \vartheta \in \Theta \tag{22d}$$
$$\sum_{v: u_v(c) \ge \vartheta} x_{v,\vartheta} \le \frac{n}{K} - 1, \quad \forall c \in C, \vartheta \in \Theta \tag{22e}$$
$$z_{v,c} \in \{0,1\}, \quad \forall v \in N, c \in C \tag{22f}$$
$$x_{v,\vartheta} \in \{0,1\}, \quad \forall v \in N, \vartheta \in \Theta \tag{22g}$$
$$y_c \in \{0,1,\dots\}, \quad \forall c \in C \tag{22h}$$

And its relaxation is

$$\text{minimize} \quad \sum_{c \in C} f(c) \cdot y_c \tag{23a}$$
$$s.t. \quad \sum_{c} z_{v,c} \le 1, \quad \forall v \in N \tag{23b}$$
$$\sum_{v} z_{v,c} \le \frac{n}{K} y_c, \quad \forall c \in C \tag{23c}$$
$$x_{v,\vartheta} \ge 1 - \sum_{c: u_v(c) \ge \vartheta} z_{v,c}, \quad \forall v \in N, \vartheta \in \Theta \tag{23d}$$
$$\sum_{v: u_v(c) \ge \vartheta} x_{v,\vartheta} \le \frac{n}{K} - 1, \quad \forall c \in C, \vartheta \in \Theta \tag{23e}$$
$$z_{v,c} \ge 0, \quad \forall v \in N, c \in C \tag{23f}$$
$$x_{v,\vartheta} \ge 0, \quad \forall v \in N, \vartheta \in \Theta \tag{23g}$$
$$y_c \ge 0, \quad \forall c \in C \tag{23h}$$

Define $\hat{y}_c = \frac{3\log(nm)}{\delta} y_c$ and round it to an integer such that its expectation remains unchanged.

For each v, c, define $z'_{v,c} = \frac{\hat{y}_c}{y_c} z_{v,c}$, and let $\hat{z}_{v,c} \in \{0,1\}$ such that $\mathbb{E}[\hat{z}_{v,c}] = \min\{1, z'_{v,c}\}$ and

$$\mathbf{Pr}\left(\sum_{v} \hat{z}_{v,c} \le \frac{n}{K}\hat{y}_c\right) = 1.$$

Since $\hat{y}_c$ variables are independent, $\{\hat{z}_{v,c}\}_{c \in S}$ are also independent for any v, and set $S \subseteq C$.

Lemma 7. *For any v and $S \subseteq C$,*

$$\mathbf{Pr}\left(\bigwedge_{c \in S} (\hat{z}_{v,c} = 0)\right) < \exp\left(-\frac{3\log(nm)}{\delta}\sum_{c \in S} z_{v,c}\right).$$

Proof. We want to bound the probability that all of the variables $\hat{z}_{v,c}$ are 0 for every c in a subset S.

Since each $\hat{z}_{v,c}$ is independently zero with probability $1 - \frac{3\log(nm)}{\delta} z_{v,c}$, the overall probability is the product of these terms:

$$\mathbf{Pr}\left(\bigwedge_{c \in S} (\hat{z}_{v,c} = 0)\right) = \prod_{c \in S}\left(1 - \frac{3\log(nm)}{\delta} z_{v,c}\right).$$

Now, by using the inequality that the geometric mean is less than or equal to the arithmetic mean, we can write:

$$\prod_{c \in S}\left(1 - \frac{3\log(nm)}{\delta} z_{v,c}\right) \le \left(1 - \frac{3\log(nm)}{\delta} \cdot \frac{1}{|S|}\sum_{c \in S} z_{v,c}\right)^{|S|}.$$

Using the standard bound $(1 - x)^a \le e^{-ax}$, we conclude:

$$\left(1 - \frac{3\log(nm)}{\delta} \cdot \frac{1}{|S|}\sum_{c \in S} z_{v,c}\right)^{|S|} \le \exp\left(-\frac{3\log(nm)}{\delta}\sum_{c \in S} z_{v,c}\right).$$

□

Now define $\hat{x}_{v,\vartheta} \in \{0,1\}$ such that it equals 0 if $\sum_{c:u_v(c)>\vartheta} \hat{z}_{v,c} \geq 1$. If $x_{v,\vartheta} < 1-\delta$, then $\hat{x}_{v,\vartheta} = 0$ with probability at least $1 - 1/(n^3 m^3)$.

By a union bound, with probability at least $1 - 1/(nm^2)$, $x_{v,\vartheta} < 1-\delta$ implies $\hat{x}_{v,\vartheta} = 0$ for all v, ϑ, so the total unsatisfied voters satisfies:

$$\sum_{v:u_v(c)\geq\vartheta} \hat{x}_{v,c',\vartheta} \leq \frac{n}{(1-\delta)K}.$$

Furthermore, by Markov's Inequality,

$$\mathbf{Pr}\left(\sum_{c\in C} \hat{y}_c > \frac{9\log(nm)}{\delta}\sum_{c\in C} y_c\right) < 1/3,$$
$$\mathbf{Pr}\left(\sum_{c\in C} f(c)\hat{y}_c > \frac{9\log(nm)}{\delta}\sum_{c\in C} f(c)y_c\right) < 1/3.$$

Combining these properties with Lemma 7, we have the following corollary.

Corollary 4. *With probability at least* $1/3 - 1/(nm^2)$*, the algorithm outputs a committee* $W \in \mathit{BJR}_{(1-\delta)K}(I)$ *achieves cost*

$$f(W) \leq \frac{9\log(nm)}{\delta} \cdot f_{\mathit{BJR}}(I,K).$$

Our algorithm repeatedly rounds the fractional variables to integers and checks whether the resulting committee satisfies BJR. This verification is feasible because the rounded variables include the matching. The algorithm terminates as soon as a committee satisfying BJR is found. Moreover, the algorithm runs in expected polynomial time. These arguments complete the proof of Theorem 4.

C Discussion: Limitation of Our LP Rounding Methods

In this section, we do not present formal theorems, but we delve deeper into the LP Rounding method. We explore why our method fails for exact JR guarantee, and discuss an alternative LP formulation of the problem.

We consider an instance $I = (C, \mathcal{A}, f)$ as follows, where our goal is to find a committee W in $\mathrm{JR}_K(I)$ that minimizes $f(W)$.

1. Let $m = 3$, meaning there are three candidates.
2. Let $n = 2d$ be a large even number.
3. The list of ballots is

$$A_v = \begin{cases} \{1\}, & v \leq d, \\ \{2\}, & d+1 \leq v \leq n-1, \\ \{3\}, & v = n. \end{cases}$$

4. The cost function is $f(c) = 1$ for any $c \in C$.
5. Let $K = 2$.

In this instance, to ensure JR, candidate 1 must be selected. Furthermore, committee $W = \{1\}$ is in $\mathrm{JR}_K(I)$, so it is the optimal committee, and the cost is $f(W) = 1$.

However, consider the LP in Section 2.1, which is restated below.

$$\begin{aligned} \text{minimize} \quad & \sum_{c\in C} f(c)\cdot y_c \\ s.t. \quad & x_v \leq \sum_{c\in A_v} y_c, \quad \forall v \in N \\ & \sum_{v\in B_c} x_v \geq |B_c| - \left\lceil \frac{n}{K} \right\rceil + 1, \quad \forall c \in C \\ & 0 \leq x_v \leq 1, \quad \forall v \in N \\ & 0 \leq y_c \leq 1, \quad \forall c \in C \end{aligned}$$

The optimal solution is $y_1^* = 1/d$, $x_v^* = 1/d$ for $v \in [d]$, and other variables has value 0. The corresponding objective value is $1/d$. As a result, the integrality gap of the LP is very large. This implies that to develop an algorithm capable of finding $W \in \mathrm{JR}_K(I)$ such that $f(W) \leq o(n) \cdot f_{\mathrm{JR}}(I,K)$, modifications to the LP formulation are necessary.

We consider the following LP.

$$\text{minimize} \quad \sum_{c\in C} f(c)\cdot y_c \tag{24}$$
$$s.t. \quad \sum_{c\in\bigcup_{v\in N'} A_v} y_c \geq 1, \quad \forall N' \in \mathcal{N} \tag{25}$$
$$0 \leq y_c \leq 1, \quad \forall c \in C \tag{26}$$

where

$$\mathcal{N} = \left\{ N' \subset N : |N'| = \frac{n}{K} \text{ and } \bigcap_{v\in N'} A_v \neq \emptyset \right\}.$$

This LP is actually a more natural formulation, because its constraints directly follow the definition of JR. However, this LP has exponentially many constraints. A typical method to deal with it is separation oracle [Grötschel *et al.*, 2012]. Given an LP, the separation oracle is able to decide whether a given solution is feasible. Specifically, for the above LP, the separation oracle reads the instance $I = (C, \mathcal{A}, f)$ and vector $\{y_c\}_{c\in C}$, and decides if Constraints 25 are satisfied. One way to construct the separation oracle is to solve the following problem.

- Define function $g(N') = \sum_{c\in\bigcup_{v\in N'} A_v} y_c$, so $g(\cdot)$ is a submodular function.
- For each $c \in C$, check whether $\min_{N'\subseteq B_c:|N'|=K} g(N')$.

In general, for a submodular function $h(\cdot) : 2^{[n]} \to \mathbb{R}_{\geq 0}$, it is NP-hard to find $S \subseteq [n]$ of size at least K such that $h(S) \leq O(n^{0.5-\epsilon}) \min_{S'\subseteq[n]:|S'|\geq K} h(S')$, for any $\epsilon > 0$ [Goemans *et al.*, 2009; Svitkina and Fleischer, 2011]. So the LP 24-26 may be hard to solve.

D Omitted Proofs in Section 3

D.1 Proof of Lemma 4

Proof. By using the same values of r and ϵ, we construct the hypergraph $H' = (V', E')$ from the hypergraph $H = (V, E)$ in Lemma 3 as follows.

1. Let $V' = V \cup V''$, where $|V''| = \frac{1}{r-1}\left(\frac{1}{r}+\epsilon\right)|V|$.[4]
2. Let $E'' = \left\{e \subset V' : |e| = r \text{ and } e \cap V'' \neq \emptyset\right\}$.
3. Let $E' = E \cup E''$.

Although $|E''| = |V''| \cdot \binom{|V|}{r-1}$, note that r is a constant, so this reduction runs in polynomial time.

For the (YES) case, suppose $\alpha(H) \geq (1 - 1/r - \epsilon)|V|$. Let $S \subset V$ be an independent set of size $(1 - 1/r - \epsilon)|V|$ in H. For each $v \in V''$, pick a hyperedge $e \in E''$ that covers v and $(r-1)$ vertices in $V \setminus S$, ensuring that the chosen $|V''|$ hyperedges collectively do not cover any vertex more than once, so they covered all the $\left(\frac{1}{r}+\epsilon\right)$ vertices in $V \setminus S$ as well as all vertices in V'', implying these $|V''|$ hyperedges forms a hyperedge dominating set. Then we have

$$\beta(H') \leq |V''| = \frac{1}{r-1}\left(\frac{1}{r}+\epsilon\right)|V| \leq \left(\frac{1}{r(r-1)}+\epsilon\right)|V'|.$$

For the (NO) case, suppose $\alpha(H) \leq \epsilon|V|$. If T is a hyperedge dominating set in H', and let S_T be the set of vertices that are covered by at least one hyperedge in T, where $|S_T| \leq r|T|$, then $V' \setminus S_T$ is an independent set in H', which implies $V \setminus S_T$ is an independent set in H, so we have

$$\begin{aligned}\epsilon|V| \geq \alpha(H) \geq |V \setminus S_T| \geq |V| - |S_T| \geq & |V| - r|T| \\ \geq & |V| - r\beta(H'),\end{aligned}$$

which implies

$$\beta(H') \geq \frac{1-\epsilon}{r}|V| \geq \frac{1-\epsilon}{r+3}|V'|.$$

□

D.2 Proof of Lemma 5

Proof. By using the same values of r and ϵ, we construct the instance $I = (C, \mathcal{A}, f)$ from the hypergraph $H' = (V', E')$ in Lemma 4 as follows.

1. Define $t = |V'|$, and let voter set $N = V' \cup V''$, where V'' is a set of additional voters of size $\sqrt{t}$, so $n = |N| = t + \sqrt{t}$. Furthermore, let $K = n/r$.
2. Define candidate set $C = E' \cup E''$, where E'' is a set of additional candidates of size $(1+\delta)K$.
3. For $v \in V'$, let $A_v = \{e \in E' : v \in e\}$, meaning voter v approves candidate e if e covers v in H'.
4. For $v \in V''$, let $A_v = E''$.
5. For this problem, the cost function $f(\cdot)$ does not matter.

For the (YES) case, suppose $\beta(H') \leq \left(\frac{1}{r(r-1)}+\epsilon\right)|V'|$. Let T be an hyperedge dominating set of size $\left(\frac{1}{r(r-1)}+\epsilon\right)|V'|$ in H'. Consider a committee W consists of T and arbitrary $\big(K - |T|\big)$ candidates in E''. Then we have $W \in \text{JR}_{K,K}(I)$, implying

$$\begin{aligned}\text{sw}_{\text{JR}}(I,K) \geq \text{sw}(W) \geq & \sqrt{t}\,\big(K - |T|\big) \\ = & \sqrt{t}\left(\frac{n}{r} - \left(\frac{1}{r(r-1)}+\epsilon\right)|V'|\right) \\ \geq & \left(\frac{r-2}{r(r-1)} - \epsilon\right)t\sqrt{t}.\end{aligned}$$

For the (NO) case, suppose $\beta(H') \geq \frac{1-\epsilon}{r+3}|V'|$. For every $W \in \text{JR}_K(I)$, $\big(W \cap E'\big)$ is a hyperedge dominating set in H', otherwise there is a hyperedge $e \in E'$ with vertices covered by e not covered by W, contradicting $W \in \text{JR}_K(I)$. As a result, for every $W \in \text{JR}_{K,(1+\delta)K}(I) \subset \text{JR}_K(I)$, we have

$$\begin{aligned}sw(W) \leq & 2\left(\frac{1-\epsilon}{r+3}t\right) + \sqrt{t}\left((1+\delta)K - \frac{1-\epsilon}{r+3}t\right) \\ \leq & \left(\frac{\delta}{r} + \frac{\epsilon r + 3}{r^2}\right)t\sqrt{t} + (\delta+3)t,\end{aligned}$$

which implies

$$\max_{W \in \text{JR}_{K,(1+\delta)K}(I)} \leq \left(\frac{\delta}{r} + \frac{\epsilon r + 3}{r^2}\right)t\sqrt{t} + (\delta+3)t.$$

□

[4] Note that $|V''|$ may not be an integer. However, since we are only interested in cases where the graph size is very large, rounding these fractional values to the nearest integer does not impact the correctness of the proof. To maintain clarity, we ignore these details in the proof.