האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

APPROXIMABILITY AND INAPPROXIMABILITY OF DODGSON AND YOUNG ELECTIONS

by

ARIEL D. PROCACCIA, MICHAL FELDMAN
and JEFFREY S. ROSENSCHEIN

Discussion Paper # 466 October 2007

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Approximability and Inapproximability of Dodgson and Young Elections

Ariel D. Procaccia* Michal Feldman† Jeffrey S. Rosenschein‡

October 7, 2007

Abstract

The voting rules proposed by Dodgson and Young are both designed to find the candidate closest to being a Condorcet winner, according to two different notions of proximity; the score of a given candidate is known to be hard to compute under both rules. In this paper, we put forward an LP-based randomized rounding algorithm which yields an $\mathcal{O}(\log m)$ approximation ratio for the Dodgson score, where m is the number of candidates. Surprisingly, we show that the seemingly simpler Young score is $\mathcal{NP}$-hard to approximate by any factor.

1 Introduction

The discipline of voting theory deals with the following setting: a set of voters each rank a set of candidates; one candidate is to be elected. The big question is: which candidate best reflects the social good? The French philosopher and mathematician Marie Jean Antoine Nicolas de Caritat, marquis de Condorcet, suggested the following intuitive criterion: the winner should be a candidate which beats every other candidate in a *pairwise election*, that is: a candidate which is preferred to any other candidate by a majority of the voters. Sadly, it is fairly easy to see that the preferences of the majority may be cyclic, hence a *Condorcet winner* does not necessarily exist. This unfortunate phenomenon is known as the *Condorcet paradox* (see Black [4]).

In order to circumvent this result, several researchers have proposed to choose a candidate which is "as close as possible" to a Condorcet winner. Different notions of proximity can be chosen, and yield different voting rules. One such suggestion was advocated by Charles Dodgson, better known as Lewis Caroll, author of "Alice's Adventures in Wonderland" [4]. The *Dodgson score* of a candidate, with respect to given voters' preferences, is the minimal number of exchanges between adjacent candidates in the voters' rankings one has to introduce in order to make the given candidate a Condorcet winner. A *Dodgson winner* is any candidate with minimal Dodgson score.

Young [20] raised a second option: measuring the distance by voters. Specifically, the *Young score* of a candidate is the size of the largest subset of voters such that, if only these ballots are taken into account, the given candidate becomes a Condorcet winner. A *Young winner* is any candidate with maximal Young score. Alternatively, one can perceive a Young winner as the candidate that becomes a Condorcet winner by removing the least number of voters.

Though these two voting rules sound appealing and straightforward, they are notoriously complicated to resolve. As early as 1989, Bartholdi, Tovey and Trick [2] have shown that computing the Dodgson score

*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, email: `arielpro@cs.huji.ac.il`.

†School of Business Administration and Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, email: `mfeldman@huji.ac.il`.

‡School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, email: `jeff@cs.huji.ac.il`.

is $\mathcal{NP}$-complete, and that pinpointing a Dodgson winner is $\mathcal{NP}$-hard. This important paper was one of the first to introduce complexity-theoretic considerations to social choice theory. Hemaspaandra et al. [9] refined the abovementioned result by showing that the Dodgson winner problem is complete for Θ_2^p, the class of problems which can be solved by $\mathcal{O}(\log n)$ queries to an $\mathcal{NP}$ set. Consequently, Rothe et al. [18] proved that the Young winner problem is also complete for Θ_2^p.

The abovementioned complexity results give rise to the agenda of *approximately* calculating a candidate's score, under the Dodgson and Young schemes. This is clearly an interesting computational problem, as an application area of algorithmic techniques.

However, from the point of view of social choice theory, it is not immediately apparent that an approximation of a voting rule is satisfactory, since an "incorrect" candidate—in our case one which is not closest to a Condorcet winner—can be elected. Nevertheless, we argue that the employment of such an approximation is strongly motivated. Indeed, at least in the case of the Dodgson and Young rules, the winner is an "approximation" in the first place, in instances where no Condorcet winner exists. Moreover, the approximation algorithm is equivalent to a new voting rule, which is guaranteed to elect a candidate which is not far from being a Condorcet winner. In other words, a perfectly sensible definition of a "socially good" winner, given the circumstances, is simply the candidate which is chosen by the approximation algorithm. Note that the approximation algorithm can be designed to satisfy the Condorcet criterion, i.e. always elect a Condorcet winner if one exists (this is always true for an approximation of the Dodgson score, as the score of a Condorcet winner is 0, and is indeed the case here).

Related work. The agenda of approximating voting rules was recently pursued by Ailon et al. [1], Coppersmith et al. [7], and Kenyon-Mathieu and Schudy [12]. These works deal, directly or indirectly, with the Kemeny rank aggregation rule, which chooses a ranking of the candidates instead of a single winning candidate. The Kemeny rule picks the ranking which has the maximum number of agreements with the voters' individual rankings regarding the correct order of pairs of candidates. Ailon et al. improve the trivial 2-approximation algorithm to an involved randomized algorithm which gives an 11/7-approximation; Kenyon-Mathieu and Schudy further improve the approximation, and obtain a PTAS. Coppersmith et al. show that the Borda ranking is a 5-approximation of the Kemeny ranking. Interestingly, Klamler [13] discusses the relation between the Kemeny rule and an extension of Dodgson's rule. However, Klamler shows that the candidate ranked first by Kemeny can appear anywhere in the Dodgson ranking. This implies that approximation algorithms for Kemeny cannot be leveraged to approximate Dodgson.

Two recent works have directly put forward algorithms for the Dodgson winner problem [10, 15]. Both papers independently build upon the same basic idea: if the number of voters is significantly larger than the number of candidates, and one looks at a uniform distribution over the preferences of the voters, with high probability one obtains an instance on which it is trivial to compute the Dodgson score of a given candidate. This directly gives rise to an algorithm with the property which Homan and Hemaspaandra [10] call *frequently self-knowingly correct*: the algorithm knows when it is definitely correct, and the algorithm is able to give a definite answer with high probability (under the assumption on the number of voters and candidates). However, this is not an approximation algorithm in the usual sense, since the algorithm *a priori* gives up on certain instances, whereas an approximation algorithm is judged by its worst-case guarantees. In addition, this algorithm would be useless if the number of candidates is not small compared to the number of voters.[1]

Betzler et al. [3] have investigated the parameterized computational complexity of the Dodgson and Young rules. The authors have devised a fixed parameter algorithm for exact computation of the Dodgson score, where the fixed parameter is the "edit distance", i.e. the number of exchanges. Specifically, if k is an

[1]This would normally not happen in political elections, but can certainly be the case in many other settings. For instance, consider a group of agents trying to reach an agreement on a joint plan, when multiple alternative plans are available.

upper bound on the Dodgson score of a given candidate, n is the number of voters, and m the number of candidates, the algorithm runs in time $\mathcal{O}(2^k \cdot nk + nm)$. Notice that in general it may hold that $k = \Omega(nm)$. In contrast, computing the Young score is $W[2]$-complete; this implies that there is no algorithm which computes the Young score exactly, and whose running time polynomial in n, m and only exponential in k, where the parameter k is the number of remaining votes. These results complement ours nicely, as we shall also demonstrate that computing Dodgson score is in a sense easier than computing Young score, albeit in the context of approximation.

More distantly related to our work is research which is concerned with exactly resolving hard-to-compute voting rules by heuristic methods. Typical examples include works regarding the Kemeny rule [6] and the Slater rule [5]. Another more remotely related field of research is concerned with finding approximate, efficient representations of voting rules, by eliciting as little information as possible; this line of research employs techniques from learning theory [17, 16].

Our results. Our results are two-fold. In the context of approximating the Dodgson score, we devise an $\mathcal{O}(\log m)$ randomized approximation algorithm, where m is the number of candidates. Our algorithm is based on solving the linear program proposed by Bartholdi et al. [2] and using randomized rounding. We also show that the integrality gap of this LP formulation, that is the worst-case ratio between the costs of the optimal integral solution and optimal fractional solution, is at least $2 - \epsilon$ for every $\epsilon > 0$. Finally, we show that the analysis of our randomized rounding algorithm is asymptotically tight.

The problem of calculating the Young score seems simpler at first glance. Therefore, our result with respect to this problem is quite surprising: it is $\mathcal{NP}$-hard to approximate the Young score by any factor. Specifically, we show that it is $\mathcal{NP}$-hard to distinguish between the case where the Young score of a given candidate is 0, and the case where the score is greater than 0.

Open questions: An obvious open question is whether a better approximation ratio for the Dodgson score can be achieved. In particular, we were unable to bridge the chasm between our $\mathcal{O}(\log m)$ upper bound and the integrality gap of only 2. Is there a better LP-based algorithm? Furthermore, it would be interesting to know whether our randomized algorithm can be derandomized.

Structure of the paper. In Section 2, we introduce some notations and definitions. In Section 3, we present our randomized rounding approximation algorithm for Dodgson score and analyze it. In Section 4, we prove that the Young score is inapproximable.

2 Preliminaries

Let $N = \{1, \dots, n\}$ be a set of *players* (which we shall refer to as *voters*), and let A be the set of *alternatives* (which we shall refer to as *candidates*). We denote $|A| = m$, and denote the candidates themselves by letters, such as $a \in A$. Each player $i \in N$ holds a binary relation R^i over A which satisfies reflexivity, antisymmetry, transitivity and totality. Informally, R^i is a ranking of the candidates. Let $L = L(A)$ be the set of all rankings over A; we have that each $R^i \in L$. We denote $R^N = \langle R^1, \dots, R^n \rangle \in L^N$, and refer to this vector as a *preference profile*. For sets of alternatives $B_1, B_2 \subseteq A$, we write $B_1 R^i B_2$ if for all $a \in B_1, b \in B_2$, aR^ib.

Let $a, b \in A$. Denote $P(a, b) = \{i \in N : aR^ib\}$. We say that *$a$ beats b in a pairwise election* if $P(a, b) > n/2$. A *Condorcet winner* is a candidate which beats every other candidate in a pairwise election.

The *Dodgson score* of a given candidate c^*, with respect to a given preference profile R^N, is the least number of exchanges between adjacent candidates in R^N needed to make c^* a Condorcet winner. For instance, let $N = \{1, 2, 3\}$, $A = \{a, b, c\}$, and let R^N be given by:

R^1	R^2	R^3
a	b	a
b	a	c
c	c	b

In this example, the Dodgson score of a is 0 (a is a Condorcet winner), b's score is 1, and c's is 3. Bartholdi et al. [2] have shown that computing the Dodgson score is an $\mathcal{NP}$-complete problem.

The *Young score* of c^* with respect to R^N is the size of the largest subset of voters such that c^* is a Condorcet winner for these voters. This is the definition given by Young himself [20], and used in subsequent works [18]. If for every nonempty subset of voters c^* is not a Condorcet winner, its Young score is 0. In the above example, a's Young score is 3, b's is 1, and c's is 0.

Notice that, equivalently, a Young winner is a candidate such that one has to remove the least number of voters in order to make it a Condorcet winner. However, in the context of approximation the two definitions are not equivalent; we employ the former (original, prevalent) definition.

As the Young winner problem is known to be intractable [18], the Young score problem must also be hard; otherwise, we would be able to calculate the scores of all the candidates efficiently, and identify the candidates with minimal score.

3 Approximating Dodgson score

In this section, we present the main result of the paper: an LP-based randomized rounding algorithm which gives an $\mathcal{O}(\log m)$ approximation for the Dodgson score of a candidate.

As a warm-up, we start by considering some trivial combinatorial algorithms. Recall that in order to compute the Dodgson score of a given candidate under some preference profile, we must perform the minimal number of exchanges between adjacent candidates. In fact, clearly the only type of exchanges to be considered are the ones which move the given candidate upward in some ranking, at the expense of some other candidate. In other words, we can simply talk about the number of positions each voter pushes the given candidate.

An approximation algorithm which immediately comes to mind is the following greedy algorithm.

Algorithm 1:
Input: A candidate c^* whose Dodgson score we wish to estimate, and a preference profile $R^N \in L^N$.
Output: An approximation of the Dodgson score of c^*.
The algorithm:

1. Let A' be the candidates which are not beaten by c^* in a pairwise election under R^N.
2. While $A' \neq \emptyset$:
   - Choose some $a \in A'$ arbitrarily.
   - Perform the minimal number of exchanges needed to make c^* beat a in a pairwise election.
   - Recalculate A'.
3. Return the number of exchanges performed.

Notice that step 2 in the while loop can be efficiently carried out. Indeed, let $d(a)$ be the *deficit* of c^* with respect to a, i.e. the number of voters c^* has to win over in order to beat a in a pairwise election. Then it is sufficient to simply choose the $d(a)$ voters which require the smallest number of exchanges in order to place c^* above a, and perform these exchanges.

Claim 3.1. *Algorithm 1 is an m-approximation algorithm for Dodgson score.*

Proof. Consider the given preference profile R^N; let $a \in A$ be the candidate which requires the maximum number t of exchanges in order to have c^* beat a in a pairwise election. The Dodgson score of c^* is at least t. On the other hand, each iteration of the algorithm's while loop clearly performs at most t exchanges, and there are at most m iterations. $\square$

Unfortunately, it is also easily seen that there are examples on which Algorithm 1 gives an $\Omega(m)$ approximation. We now turn our attention to a second simple combinatorial algorithm. The input and output of the algorithm are the same as before.

Algorithm 2:

1. Let A' be the candidates which are not beaten by c^* in a pairwise election under R^N.
2. While $A' \neq \emptyset$:
   - Move c^* upward by one position in the preferences of all the voters (unless c^* is already ranked highest).
   - Recalculate A'.
3. Return the number of exchanges performed.

Claim 3.2. *Algorithm 2 is an n-approximation algorithm for Dodgson score.*

Proof. Consider the minimal sequence of exchanges which makes c^* a Condorcet winner, and denote the length of this sequence (which is, in fact, c^*'s Dodgson score) by t. For every $i \in N$, denote by s_i^* the position of c^* as a result of this sequence in the preferences of voter i (where m is the top ranking position and 1 is the lowest ranking). Let s_i be the position of c^* in voter i's ranking after t iterations of the algorithm's while loop. It is self evident that for all $i \in N$, $s_i \geq s_i^*$. Therefore, after at most t iterations c^* certainly becomes a Condorcet winner, and the algorithm halts. We conclude that the number of exchanges the algorithm makes is at most $t \cdot n$. $\square$

Algorithm 2's worst-case approximation ratio is also $\Omega(n)$. Indeed, it is easy to find an example where c^* needs only one exchange to become a Condorcet winner, but a single iteration of the algorithm leads to $\Omega(n)$ exchanges.

3.1 The Randomized Rounding algorithm

Bartholdi et al. [2] give a surprising integer linear programming (ILP) formulation for the Dodgson score. The number of constraints and variables in their program depends purely on the number of candidates. Therefore, if the number of candidates is constant, the program is solvable in polynomial time using the algorithm of Lenstra [14]. However, if the number of candidates is not constant, the LP is of gargantuan size.[2]

Fortunately, it is straightforward to modify the abovementioned ILP to obtain a program of polynomial size. As before, let $c^* \in A$ be the candidate whose score we wish to compute. Let the variables of the program be $x_{ij} \in \{0,1\}$ for all $i \in N$ and $j = 1, \dots, m-1$; $x_{ij} = 1$ if and only if c^* is moved upward by j positions in the ranking of voter i. Define constants $e_{ij}^a \in \{0,1\}$, for all $i \in N$, $j = 1, \dots, m$, and $a \in A \setminus \{c^*\}$, which depend on the given preference profile; $e_{ij}^a = 1$ iff moving c^* upward by j positions in the ranking of

[2] Note that there is also an efficient solution if the number of voters n is constant; indeed, brute force search requires checking $\mathcal{O}(m^n)$ possibilities.

voter i makes c^* gain an *additional* vote against a (note that $e^a_{ij} = 0$ for all j if $c^* R^i a$). Once again, let $d(a)$ be the deficit of c^* with respect to a, i.e. the number of voters c^* must gain in order to defeat a in a pairwise election (if a is beaten from the outset, $d(a) = 0$). The ILP which computes the Dodgson score of c^* is given by:

$$\begin{aligned}
&\mathbf{min} \sum_{i,j} j \cdot x_{ij} \ \mathbf{such\ that} \\
&\quad \forall i \in N,\ \sum_j x_{ij} = 1 \\
&\quad \forall a \in A \setminus \{c^*\},\ \sum_{i,j} x_{ij} e^a_{ij} \geq d(a) \\
&\quad \forall i, j,\ x_{ij} \in \{0, 1\}
\end{aligned} \tag{1}$$

This ILP can be relaxed by asking merely that $0 \leq x_{ij} \leq 1$ for all i and j. The resulting linear program (LP) can be solved efficiently [19].

We are now ready to present our randomized rounding algorithm. Its input and output are as before.

Randomized Rounding Algorithm:

1. Solve the relaxed LP given by (1) to obtain a solution $\vec{x}$.
2. For $k = 1, \ldots, c \cdot \log m$ (where $c > 0$ is a constant to be chosen later)
   - For all $i \in N$, randomly and independently (from other players and other iterations) choose a value X^k_i, such that $X^k_i = j$ with probability x_{ij}.
3. For all $i \in N$, set $X^{max}_i = \max_k X^k_i$.
4. Let $\mathcal{X}'$ be the solution which moves c^* upward in the ranking of i by X^{max}_i positions; return $\text{cost}(\mathcal{X}') = \sum_{i \in N} X^{max}_i$.

Before commencing our formal analysis of the algorithm, we remark that if c^* is a Condorcet winner from the outset, clearly the algorithm will calculate a score of 0 (with probability 1). Therefore, if we defined a new (randomized) voting rule, which elects the candidate with minimal score according to the algorithm, this voting rule would satisfy the Condorcet criterion.

3.2 Analysis of the Algorithm

In this subsection we shall prove:

Theorem 3.3. *For any input c^* and R^N, the randomized rounding algorithm returns a $4c \cdot \log m$ approximation of the Dodgson score with probability at least 1/2.*

Proof. Fix some iteration k of the algorithm's for loop. Let $X_i = X^k_i$, $i \in N$, be independent discrete random variables such that $X_i = j$ with probability x_{ij}. Consider the sequence of exchanges induced by the variables X_i, i.e. each voter $i \in N$ moves c^* upward by j places with probability x_{ij}. As a result of the constraint $\forall i \in N$, $\sum_j x_{ij} = 1$, these are legal random variables. Moreover, let $\mathcal{X}$ be the chosen sequence of exchanges, and denote the LP's optimal fractional solution by $\text{OPT}_f = \sum_{i,j} j \cdot x_{ij}$; it holds that

$$\mathbb{E}[\text{cost}(\mathcal{X})] = \mathbb{E}\left[\sum_i X_i\right] = \text{OPT}_f. \tag{2}$$

Now, fix some candidate $a \neq c^*$. We wish to bound the probability that c^* does not beat a after the exchanges given by $\mathcal{X}$ are made in R^N.

Let Y_i, $i \in N$, be independent Bernoulli trials, such that $Y_i = 1$ iff $aR^i c^*$, and c^* is moved above a in voter i's ranking. In other words, $Y_i = 1$ if voter i becomes an additional voter which ranks c^* above a as a result of the exchanges. We want to provide a lower bound on $\Pr[\sum_i Y_i < d(a)]$. Denote

$$p_i = \sum_{j:\ e^a_{ij}=1} x_{ij}.$$

Notice that $Y_i = 1$ with probability p_i, so $\mathbb{E}[\sum_i Y_i] = \sum_i p_i$. Moreover, by the constraint $\forall a \in A \setminus \{c^*\}$, $\sum_{i,j} x_{ij} e^a_{ij} \geq d(a)$, we have that $\sum_i p_i \geq d(a)$. We now employ a deceivingly intuitive but highly nontrivial result:

Lemma 3.4 (Jogdeo and Samuels [11]). *Let $Y_1, \ldots, Y_n$ be independent heterogeneous Bernoulli trials. Suppose that $\mathbb{E}[\sum_i Y_i]$ is an integer. Then*

$$\Pr\left[\sum_i Y_i < \mathbb{E}\left[\sum_i Y_i\right]\right] < 1/2.$$

Since $d(a)$ is an integer, and $\mathbb{E}[\sum_i Y_i] = \sum_i p_i \geq d(a)$, it follows from the lemma that:

$$\Pr[a \text{ not beaten in } \mathcal{X}] = \Pr\left[\sum_i Y_i < d(a)\right] < 1/2.$$

At this point, we choose the value of the constant c to be such that $2^{c \log m} \geq 4m$. Note that if $m \geq 4$, we can choose $c \leq 2$. As in the algorithm, set $X_i^{max} = \max_k X_i^k$. Denote by $\mathcal{X}'$ the induced sequence of exchanges. It holds that a is not beaten in a pairwise election under $\mathcal{X}'$ only if a is not beaten under the exchanges obtained in each one of the $c \cdot \log m$ individual iterations. Therefore,

$$\Pr[a \text{ not beaten in } \mathcal{X}'] < \left(\frac{1}{2}\right)^{c \cdot \log m} \leq \frac{1}{4m}.$$

By the union bound we get:[3]

$$\Pr[c^* \text{ is not a Condorcet winner in } \mathcal{X}'] \leq m \cdot \frac{1}{4m} = 1/4. \tag{3}$$

$X_i^1, \cdots, X_i^{c \log m}$ are i.i.d random variables; it holds that $X_i^{max} = \max_k X_i^k \leq \sum_k X_i^k$, and thus

$$\mathbb{E}[X_i^{max}] \leq \mathbb{E}[\sum_k X_i^k] = c \cdot \log m \cdot \mathbb{E}[X_i^1].$$

Therefore, by the linearity of expectation,

$$\begin{aligned}
\mathbb{E}[\text{cost}(\mathcal{X}')] &= \mathbb{E}\left[\sum_i X_i^{max}\right] \\
&\leq c \cdot \log m \cdot \mathbb{E}[\sum_i X_i^1] \\
&= c \cdot \log m \cdot \mathbb{E}[\text{cost}(\mathcal{X})] \\
&= c \cdot \log m \cdot \text{OPT}_f \\
&\leq c \cdot \log m \cdot \text{OPT},
\end{aligned}$$

[3] Strictly speaking, we can use $m - 1$ instead of m.

where OPT is the Dodgson score of c^*, i.e. the optimal integral solution to the ILP (1).

By Markov's inequality we have that

$$\Pr[\text{cost}(\mathcal{X}') > \text{OPT} \cdot 4 \cdot c \cdot \log m] \leq 1/4. \tag{4}$$

We now apply the union bound once again on (3) and (4), and obtain that with probability at least 1/2, c^* is a Condorcet winner under $\mathcal{X}'$ and, at the same time, $\text{cost}(\mathcal{X}') \leq \text{OPT} \cdot 4 \cdot c \cdot \log m$. This complete the proof of Theorem 3.3. □

It is possible to verify in polynomial time whether the algorithm's output is, at the same time, a valid solution (i.e. c^* is a Condorcet winner) and a $4c \cdot \log m$-approximation. Therefore, it is possible to repeat the algorithm from scratch to improve the probability of success. The expected number of repetitions is at most 2.

3.3 An Integrality Gap of $2 - \epsilon$

In this subsection we shall demonstrate that using an algorithm which solves the relaxed LP (1) and rounds it, one cannot hope to obtain an approximation ratio better than 2. More formally, we shall show that the *integrality gap* of the LP (1), i.e. the worst-case ratio between the optimal integral solution OPT and the optimal fractional solution OPT_f, is at least $2 - \epsilon$, for any given ϵ. Interestingly, the example used to establish the integrality gap will be directly used to show that the analysis of our randomized rounding algorithm is tight.

Let m be the number of candidates, and let $A' \subset A$ be a subset of candidates of size s: $A' = \{a_1, \ldots, a_s\}$. Let $A'' = A \setminus (A' \cup \{c^*\})$; we shall assume that $|A''|$ is divisible by s. Let $A''_1, \ldots, A''_s$ be a partition of A'' such that for all $t = 1, \ldots, s$,

$$|A''_t| = \frac{|A''|}{s} = \frac{m - s - 1}{s}.$$

We shall consider a set of voters N, $|N| = 2s - 2$, partitioned into two subsets, $N' = \{1, \ldots, s\}$, and $N'' = \{s + 1, \ldots, 2s - 2\}$. Let us define the preference profile R^N. The voters in N' vote as follows:

R^1	. . .	R^s
$A' \setminus \{a_1\}$	. . .	$A' \setminus \{a_s\}$
$A'' \setminus A''_1$	. . .	$A'' \setminus A''_s$
c^*	. . .	c^*
a_1	. . .	a_s
A''_1	. . .	A''_s

Meanwhile, the voters in N'' all rank c^* on top, and the other candidates in some arbitrary order. We have that for all $a \in A \setminus \{c^*\}$, $d(a) = 1$, since $s - 1$ voters prefer c^* to a and $s - 1$ voters prefer a to c^*. Now, in any optimal integral solution, c^* must be moved above $A'' \setminus A''_t$ by at least two voters (i.e at least two values of t), in order for c^* to beat all the candidates in A'. Thus,

$$\text{OPT} \geq 2 \cdot |A'' \setminus A''_t| = 2 \cdot \frac{s - 1}{s} \cdot (m - s - 1).$$

On the other hand, consider the fractional solution where for each $i \in N'$, c^* is moved to the top of voter i's ranking with probability $\frac{1}{s-1}$, and stays in place with probability $1 - \frac{1}{s-1}$. Formally, for $i = 1, \ldots, s$,

$$x_{i, m-2-\frac{m-s-1}{s}} = \frac{1}{s - 1},$$

and $x_{i0} = 1 - \frac{1}{s-1}$. Since for all $a \in A \setminus \{c^*\}$, $s-1$ voters in N' rank a above c^*, this gives a valid solution; it is also straightforward that this solution is optimal. By summing on all voters in N', we conclude that

$$\mathrm{OPT}_f = \frac{s}{s-1} \cdot \left(m - 2 - \frac{m-s-1}{s} \right).$$

Hence,

$$\frac{\mathrm{OPT}}{\mathrm{OPT}_f} \geq 2 \cdot \left(\frac{s-1}{s} \right)^2 \cdot \frac{m-s-1}{\left(m-2-\frac{m-s-1}{s}\right)}.$$

If we take, for instance, $s \approx \log m$, this expression is arbitrarily close to 2 as m grows larger.

Let us now leverage this example to demonstrate the asymptotic tightness of the analysis given in Subsection 3.2. Let $r = m - 2 - \frac{m-s-1}{s}$. The optimal integral solution satisfies $\mathrm{OPT} \leq 2r$, as it is possible to make c^* a Condorcet winner by moving it to the top of the rankings of two of the voters in N'.

On the other hand, the optimal fractional solution described above induces Bernoulli random variables X_i^k, for $i = 1, \ldots, s$, which assume the value r with probability $\frac{1}{s-1}$, and 0 with probability $1 - \frac{1}{s-1}$.

Now, let $s = c \cdot \log m + 1$; we have that for $i = 1, \ldots, s$, $X_i^{max} = r$ if and only if there exists $k \in \{1, \ldots, c \cdot \log m\}$ such that $X_i^k = r$; this happens with probability

$$1 - \left(1 - \frac{1}{s-1}\right)^{s-1} \geq 1 - \frac{1}{e},$$

where the constant e is the base of the natural logarithm. Therefore, $\mathbb{E}[X_i^{max}] \geq \left(1 - \frac{1}{e}\right) \cdot r$, and consequently $\mathbb{E}(\mathcal{X}') \geq s \cdot \left(1 - \frac{1}{e}\right) \cdot r$. We conclude that

$$\frac{\mathbb{E}[\mathrm{cost}(\mathcal{X}')]}{\mathrm{OPT}} \geq \left(1 - \frac{1}{e}\right) \cdot \frac{1}{2} \cdot s = \Omega(\log m).$$

4 Inapproximability of the Young Score

Recall that the Young score of a given candidate $c^* \in A$ is the size of the maximal subset of voters for which c^* is a Condorcet winner.

It is straightforward to obtain a simple ILP for the Young score problem. As before, let $c^* \in A$ be the candidate whose Young score we wish to compute. Let the variables of the program be $x_i \in \{0,1\}$ for all $i \in N$; $x_i = 1$ iff voter i is included in the subset of voters for c^*. Define constants $e_i^a \in \{-1, 1\}$ for all $i \in N$ and $a \in A \setminus \{c^*\}$, which depend on the given preference profile; $e_i^a = 1$ iff voter i ranks c^* higher than a. The ILP which computes the Young score of c^* is given by:

$$\begin{aligned} &\mathbf{max} \sum_i x_i \textbf{ such that} \\ &\forall a \in A \setminus c^*,\ \sum_i x_i e_i^a \geq 1 \\ &\forall i,\ x_i \in \{0,1\} \end{aligned} \tag{5}$$

The ILP (5) for the Young score is seemingly simpler than the one for the Dodgson score, given as (1). This might seem to indicate the the problem can be easily approximated by similar techniques. Therefore, the following result is quite surprising.

Theorem 4.1. *It is $\mathcal{NP}$-hard to approximate the Young score by any factor.*

This result becomes more self-evident when we notice that the Young score is nonmonotonic as an optimization problem. Indeed, given a subset of voters which make c^* a Condorcet winner, it is not necessarily the case that a smaller subset of the voters would satisfy the same property. This stands in contrast to many approximable optimization problems, in which a solution which is worse than a valid solution is also a valid solution. Consider the famous Set Cover problem, for instance: if one adds more subsets to a valid cover, one obtains a valid cover. The same goes for the Dodgson score problem: if a sequence of exchanges makes c^* a Condorcet winner, introducing more exchanges on top of the existing ones would not undo this fact.

In order to prove the inapproximability of the Young score, we define the following problem.

`NonEmptySubset`
Instance: A candidate c^*, and a preference profile $R^N \in L^N$.
Question: Is there a nonempty subset of voters $S \subseteq N$, $S \neq \emptyset$, for which c^* is a Condorcet winner?

In order to prove Theorem 4.1, it is sufficient to prove that NonEmptySubset is $\mathcal{NP}$-hard. Indeed, this implies that it is $\mathcal{NP}$-hard to distinguish whether the Young score of a given candidate is 0 or greater than 0, which directly entails that the score cannot be approximated.

Lemma 4.2. NonEmptySubset *is $\mathcal{NP}$-complete.*

Proof. The problem is clearly in $\mathcal{NP}$; a witness is given by a nonempty set of voters for which c^* is a Condorcet winner.

In order to show $\mathcal{NP}$-hardness, we present a polynomial-time reduction from the $\mathcal{NP}$-hard Exact Cover by 3-Sets (X3C) problem [8] to our problem. An instance of the X3C problem includes a finite set of elements U, $|U| = n$ (where n is divisible by 3), and a collection $\mathcal{C}$ of 3-element subsets of U, $\mathcal{C} = \{C_1, \ldots, C_k\}$, such that for every $1 \leq i \leq k$, $C_i \subseteq U$ and $|C_i| = 3$. The question is whether the collection $\mathcal{C}$ contains an *exact cover* for U, i.e., a subcollection $\mathcal{C}^* \subseteq \mathcal{C}$ of size $n/3$ such that every element of U occurs in exactly one subset in $\mathcal{C}'$.

We next give the details of the reduction from X3C to NonEmptySubset. Given an instance of X3C, defined by the set U and a collection of 3-element sets $\mathcal{C}$, we construct the following instance of NonEmptySubset.

Define the set of candidates as $A = U \cup \{a\} \cup \{c^*\}$. Let the set of voters be $N = N' \cup N''$, where N' and N'' are defined as follows. The set N' is composed of k voters, corresponding to the k subsets in $\mathcal{C}$, such that for all $i \in N'$, voter i prefers the candidates in $U \setminus C_i$ to c^*, and prefers c^* to all the candidates in $C_i \cup \{a\}$ (i.e., $U \setminus C_i \; R^i \; c^*$, and $R^i \; C_i \cup \{a\}$).

Subset N'' is composed of $\frac{n}{3} - 1$ voters who prefer a to c^* and c^* to U (i.e., for all $i \in N''$, $a \; R^i \; c^* \; R^i \; U$).

We next show that there is an exact cover in the given instance iff there is nonempty subset of voters for which c^* is a Condorcet winner in the constructed instance.

Sufficiency: Let $\mathcal{C}^*$ be an exact cover by 3-sets of U, and let $N^* \subseteq N'$ be the subset of voters corresponding to the $\frac{n}{3}$ subsets $C_i \in \mathcal{C}^*$. We show that c^* is a Condorcet winner for $S = N^* \cup N''$. Since $\mathcal{C}'$ is an exact cover, for all $b \in U$ there exists exactly one voter in N^* that prefers c^* to b and $\frac{n}{3} - 1$ voters in N^* that prefer b to c^*. In addition, all $\frac{n}{3} - 1$ voters in N'' prefer c^* to b. Therefore, c^* beats b in a pairwise election.

It remains to show that c^* beats a in a pairwise election. This is true since all $\frac{n}{3}$ voters in N^* prefer c^* to a, and there are only $\frac{n}{3} - 1$ voters in N'' who prefer a over c^*. It follows that c^* is a Condorcet winner for $N^* \cup N''$.

Necessity: Assume the given instance of X3C has no exact cover. We have to show that there is no subset of voters for which c^* is a Condorcet winner. Let $S \subseteq N$, $S \neq \emptyset$, and let $N^* = S \cap N'$. We distinguish between three cases.

Case 1: $|N^*| = 0$. It must hold that $S \cap N'' \neq \emptyset$. In this case, c^* loses to a in a pairwise election, since all the voters in N'' prefer a to c^*.

Case 2: $0 < |N^*| \leq \frac{n}{3}$. Since there is no exact cover, the corresponding sets C_i cannot cover U. Thus there exists $b \in U$ that is ranked lower than c^* by all voters in N^*. In order for c^* to beat b in a pairwise election, S must include at least $|N^*|+1$ voters from N''. However, this means that a beats c^* in a pairwise election (since a is ranked lower than c^* by $|N^*|$ voters, and higher than c^* by at least $|N^*|+1$ voters). It follows that c^* is not a Condorcet winner for S.

Case 3: $|N^*| > \frac{n}{3}$. Let us award each candidate $b \in A \setminus \{c^*\}$ a point for each voter which ranks it above c^*, and subtract a point for each voter which ranks it below c^*. c^* is a Condorcet winner iff the score of every other candidate, counted this way, is negative. This implies that c^* is a Condorcet winner only if for every subset $K \subseteq A$ of candidates, the total score of the candidates in K is at most $-|K|$.

We shall calculate the total score of the candidates in U from the voters in N^*. Every voter in N^* prefers c^* to 3 candidates in U and prefers $n-3$ candidates in U to c^*. Thus, every voter in N^* contributes $(n-3)-3 = n-6$ points to the total score of U. Summing over all the voters in N^*, we have that the total score of U from N^* is $|N^*|(n-6)$. By $|N^*| > \frac{n}{3}$, we have that

$$|N^*|(n-6) \geq \left((\frac{n}{3}-1)+2\right)(n-6) = (\frac{n}{3}-1)n - 6.$$

Recall that every voter in N'' prefers c^* to all candidates in U. However, since $|N''| = \frac{n}{3}-1$, voters from N'' can only subtract $(\frac{n}{3}-1)n$ from the total score of U. We conclude that the total score of U is at least -6. Since we can assume that $|U| = n > 6$,[4] c^* cannot beat all the candidates in U in pairwise elections. This concludes the proof. □

5 Acknowledgments

The authors wish to thank Nati Linial for pointing their attention to Jogdeo and Samuels [11]. Ariel Procaccia is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

References

[1] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. In *Proceedings of the 37th annual ACM symposium on Theory of computing (STOC)*, pages 684–693, 2005.

[2] J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.

[3] N. Betzler, J. Guo, and R. Niedermeier. Parameterized computational complexity of Dodgson and Young elections. Manuscript, 2007.

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

[5] V. Conitzer. Computing Slater rankings using similarities among candidates. In *In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006.

[6] V. Conitzer, A. Davenport, and J. Kalagnanam. Improved bounds for computing Kemeny rankings. In *In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006.

[7] D. Coppersmith, L. Fleischer, and A. Rudra. Ordering by weighted number of wins gives a good ranking for weighted tournaments. In *Proceedings of the 17th annual ACM-SIAM symposium on Discrete algorithm (SODA)*, pages 776–782, 2006.

[4]X3C is obviously tractable for a constant n, as one can examine all the families $\mathcal{C}' \subseteq \mathcal{C}$ of constant size in polynomial time.

[8] M. R. Garey and D. S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[9] E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Exact analysis of Dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM*, 44(6):806–825, 1997.

[10] C. Homan and L. A. Hemaspaandra. Guarantees for the success frequency of an algorithm for finding Dodgson election winners. In *Proceedings of the 31st International Symposium on Mathematical Foundations of Computer Science (MFCS)*, 2006.

[11] K. Jogdeo and S. Samuels. Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. *Annals of Mathematical Statistics*, 39:1191–1195, 1968.

[12] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *Proceedings of the 39th annual ACM symposium on Theory of computing (STOC)*, pages 95–103, 2007.

[13] C. Klamler. The Dodgson ranking and its relation to Kemeny's method and Slater's rule. *Social Choice and Welfare*, 23(1):91–102, 2004.

[14] H. W. Lenstra. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8:538–548, 1983.

[15] J. C. McCabe-Dansted, G. Pritchard, and A. M. Slinko. Approximability of Dodgson's rule. In *Proceedings of the 1st International Workshop on Computational Social Choice (COMSOC)*, pages 331–344, 2006.

[16] A. D. Procaccia, A. Zohar, Y. Peleg, and J. S. Rosenschein. Learning voting trees. In *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI)*, pages 110–115, 2007.

[17] A. D. Procaccia, A. Zohar, and J. S. Rosenschein. Automated design of voting rules by learning from examples. In *Proceedings of the 1st International Workshop on Computational Social Choice (COMSOC)*, pages 436–449, 2006.

[18] J. Rothe, H. Spakowski, and J. Vogel. Exact complexity of the winner problem for young elections. *Theory of Computing Systems*, 36(4):375–386, 2003.

[19] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*. Springer, 2nd edition, 2001.

[20] H. P. Young. Extending Condorcet's rule. *Journal of Economic Theory*, 16:335–353, 1977.