Weighted Voting Via No-Regret Learning

Nika Haghtalab
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
nhaghtal@cs.cmu.edu

Ritesh Noothigattu
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
riteshn@cs.cmu.edu

Ariel D. Procaccia
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
arielpro@cs.cmu.edu

Abstract

Voting systems typically treat all voters equally. We argue that perhaps they should not: Voters who have supported good choices in the past should be given higher weight than voters who have supported bad ones. To develop a formal framework for desirable weighting schemes, we draw on *no-regret learning*. Specifically, given a voting rule, we wish to design a weighting scheme such that applying the voting rule, with voters weighted by the scheme, leads to choices that are almost as good as those endorsed by the best voter in hindsight. We derive possibility and impossibility results for the existence of such weighting schemes, depending on whether the voting rule and the weighting scheme are deterministic or randomized, as well as on the social choice axioms satisfied by the voting rule.

1 Introduction

In most elections, voters are entitled to equal voting power. This principle underlies the *one person, one vote* doctrine, and is enshrined in the United States Supreme Court ruling in the *Reynolds v. Sims* (1964) case.

But there are numerous voting systems in which voters do, in fact, have different *weights*. Standard examples include the European Council, where (for certain decisions) the weight of each member country is proportional to its population; and corporate voting procedures where stockholders have one vote per share. Some historical voting systems are even more pertinent: Sweden's 1866 system weighted voters by wealth, giving especially wealthy voters as many as 5000 votes; and a Belgian system, used for a decade at the end of the 19th Century, gave (at least) one vote to each man, (at least) two votes to each educated man, and three votes to men who were both educated and wealthy [Congleton, 2011].

The last two examples can be seen as (silly, from a modern viewpoint) attempts to weight voters by *merit*, using wealth and education as measurable proxies thereof. We believe that the basic idea of weighting voters by merit does itself have merit. But we propose to measure a voter's merit by the *quality of his past votes*. That is, a voter who has supported good choices in the past should be given higher weight than a voter who has supported bad ones.

This high-level scheme is, arguably, most applicable to *repeated aggregation of objective opinions*. For example, consider a group of engineers trying to decide which prototype to develop, based on an objective measure of success such as projected market share. If an engineer supported a certain prototype and it turned out to be a success, she should be given higher weight compared to her peers in future decisions; if it is a failure, her weight should lower. Similar examples include a group of investors selecting companies to invest in; and a group of decision makers in a movie studio choosing movie scripts to produce. Importantly, the recently launched, not-for-profit website RoboVote.org already provides public access to voting tools for precisely these situations, albeit using methods that always treat all voters equally [Procaccia et al., 2016].

Our goal in this paper, therefore, is to augment existing voting methods with weights, in a way that keeps track of voters' past performance, and guarantees good choices over time. The main conceptual problem we face is the development of a formal framework in which one can reason about desirable weighting schemes.[1] To address this problem, we build on the *no-regret learning* literature, but depart from the classic setting in several ways — some superficial, and some fundamental.

Specifically, instead of experts, we have a set of n *voters*. In each round, each voter reveals a *ranking* over a set of alternatives, and the loss of each alternative is determined. In addition, we are given a (possibly randomized) *voting rule*, which receives weighted rankings as input, and outputs the winning alternative. The voting rule is not part of our design space; it is exogenous and fixed throughout the process. The *loss* of a voter in round t is given by assigning his ranking all the weight (equivalently, imagining that all voters have that ranking), applying the voting rule, and measuring the loss of the winning alternative (or the expected loss, if the rule is randomized). As in the classic setting, our benchmark is the best *voter* in hindsight (but we also discuss the stronger benchmark of best *voter weights* in hindsight in Section 6).

At first glance, it may seem that our setting easily reduces to the classic one, by treating voters as experts. But our loss is computed by applying the given voting rule to the entire profile of weighted rankings, and therein lies the rub.[2] This leads to our main research question: *For which voting rules is there a weighting scheme such that the difference between our average per-round loss and that of the best voter goes to zero as the number of rounds goes to infinity?*

In Section 4, we devise no-regret weighting schemes for any voting rule, under two classic feedback models — *full information* and *partial information*. While these results make no assumptions on the voting rule, the foregoing weighting schemes heavily rely on randomization. By contrast, deterministic weighting schemes seem more desirable, as they are easier to interpret and explain. In Section 5, therefore, we restrict our attention to deterministic weighting schemes. We find that if the voting rule is itself deterministic, it admits a no-regret weighting scheme if and only if it is *constant on unanimous profiles*. Because this property is not satisfied by any reasonable rule, the theorem should be interpreted as a strong impossibility result. We next consider randomized voting rules, and find that they give rise to much more subtle results, which depend on the properties of the voting rule in question.

2 Preliminaries

Our work draws on social choice theory and online learning. In this section we present important concepts and results from each of these areas in turn.

2.1 Social Choice

We consider a set $[n] \triangleq \{1, \ldots, n\}$ of *voters* and a set A of m *alternatives*. A *vote* $\sigma : A \to [m]$ is a linear ordering — a ranking or permutation — of the alternatives. That is, for any vote σ and alternative a, $\sigma(a)$ denotes the position of alternative a in vote σ. For any $a, b \in A$, $\sigma(a) < \sigma(b)$ indicates that alternative a is preferred to b under vote σ. We also denote this preference by $a \succ_\sigma b$. We denote the set of all $m!$ possible votes over A by $\mathcal{L}(A)$.

A *vote profile* $\boldsymbol{\sigma} \in \mathcal{L}(A)^n$ denotes the votes of n voters. Furthermore, given a vote profile $\boldsymbol{\sigma} \in \mathcal{L}(A)^n$ and a weight vector $\mathbf{w} \in \mathbb{R}^n_{\geq 0}$, we define the *anonymous vote profile corresponding to* $\boldsymbol{\sigma}$ *and* $\mathbf{w}$,

[1] In that sense, our work is related to papers in *computational social choice* [Brandt et al., 2016] that study weighted voting, in the context of manipulation, control, and bribery in elections [Conitzer et al., 2007, Zuckerman et al., 2009, Faliszewski et al., 2009, 2015].

[2] For the same reason, our work is quite different from papers on online learning algorithms for ranking, where the algorithm chooses a ranking of objects at each stage, and suffers a loss based on the "relevance" of the ranking [Radlinski et al., 2008, Chaudhuri and Tewari, 2015].

denoted $\boldsymbol{\pi} \in [0,1]^{|\mathcal{L}(A)|}$, by setting

$$\pi_\sigma \triangleq \frac{1}{\|\mathbf{w}\|_1} \sum_{i=1}^{n} w_i \mathbb{1}_{(\sigma_i = \sigma)}, \quad \forall \sigma \in \mathcal{L}(A).$$

That is, $\boldsymbol{\pi}$ is an $|\mathcal{L}(A)|$-dimensional vector such that for each vote $\sigma \in \mathcal{L}(A)$, π_σ is the fraction of the total weight on σ. When needed, we use $\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}$ to clarify the vote profile and weight vector to which the anonymous vote profile corresponds to. Note that $\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}$ only contains the anonymized information about $\boldsymbol{\sigma}$ and $\mathbf{w}$, i.e., the anonymous vote profile remains the same even when the identities of the voters change.

To aggregate the (weighted) votes into a distribution over alternatives, we next introduce the concept of (anonymous) voting rules. Let $\Delta(\mathcal{L}(A))$ be the set of all possible anonymous vote profiles. Similarly, let $\Delta(A)$ denote the set of all possible distributions over A. An anonymous *voting rule* is a function $f : \Delta(\mathcal{L}(A)) \rightarrow \Delta(A)$ that takes as input an anonymous vote profile $\boldsymbol{\pi}$ and returns a distribution over the alternatives indicated by a vector $f(\boldsymbol{\pi})$, where $f(\boldsymbol{\pi})_a$ is the probability that alternative a is the winner under $\boldsymbol{\pi}$. We say that a voting rule f is *deterministic* if for any $\boldsymbol{\pi} \in \Delta(\mathcal{L}(A))$, $f(\boldsymbol{\pi})$ has support of size 1, i.e., there is a unique winner.

An anonymous voting rule f is called *strategyproof* if, informally, voters can never achieve a better outcome by misreporting their preferences (see Appendix A for formal definitions). While strategyproofness is a natural property to be desired in a voting rule, the celebrated Gibbard-Satterthwaite Theorem [Gibbard, 1973, Satterthwaite, 1975] shows that non-dictatorial strategyproof deterministic voting rules do not exist.[3] Subsequently, Gibbard [1977] extended this result to a characterization of strategyproof *randomized* voting rules. The next proposition is a direct corollary of his result for the case of anonymous rules.

Proposition 2.1. *Any strategyproof randomized rule is a distribution over a collection of the following types of rules:*

1. *Anonymous Unilaterals: g is an anonymous unilateral if there exists a function $h : \mathcal{L}(A) \rightarrow A$ for which $g(\boldsymbol{\pi}) = \sum_{\sigma \in \mathcal{L}(A)} \pi_\sigma \mathbf{e}_{h(\sigma)}$.*

2. *Duple: g is a duple rule if $|\{a \mid \exists \boldsymbol{\pi}$ such that $g(\boldsymbol{\pi})_a \neq 0\}| \leq 2$.*

Examples of strategyproof randomized voting rules include *randomized positional scoring rules* and the *randomized Copeland* rule, which were previously studied in this context [Conitzer and Sandholm, 2006, Procaccia, 2010]. The reader is referred to Appendix A for more details.

2.2 Online Learning

We next describe the general setting of online learning, also known as learning from experts. We consider a game between a *learner* and an *adversary*. There is a set of actions (a.k.a experts) $\mathcal{X}$ available to the learner, a set of actions $\mathcal{Y}$ available to the adversary, and a loss function $f : \mathcal{X} \times \mathcal{Y} \rightarrow [0,1]$ that is known to both parties. In every time step $t \in [T]$, the learner chooses a distribution, denoted by a vector $\mathbf{p}^t \in \Delta(\mathcal{X})$, over the actions in $\mathcal{X}$, and the adversary chooses an action y^t from the set $\mathcal{Y}$. The learner then receives a loss of $f(x^t, y^t)$ for $x^t \sim \mathbf{p}^t$. At this point, the learner receives some feedback regarding the action of the adversary. In the *full information* setting, the learner observes y^t before proceeding to the next time step. In the *partial information* setting, the learner only observes the loss $f(x^t, y^t)$.

The *regret* of the algorithm is defined as the difference between its total expected loss and that of the best fixed action in hindsight. The goal of the learner is to minimize its expected regret, that is, minimize

$$\mathbb{E}[Reg_T] \triangleq \mathbb{E}\left[\sum_{t=1}^{T} f(x^t, y^t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} f(x, y^t)\right],$$

where the expectation is taken over the choice of $x^t \sim \mathbf{p}^t$, and any other random choices made by the algorithm and the adversary. An online algorithm is called a *no-regret* algorithm if $\mathbb{E}[Reg_T] \in o(T)$. In words, the average regret of the learner must go to 0 as $T \rightarrow \infty$. In general, deterministic algorithms, for which $\|\mathbf{p}^t\|_\infty = 1$, can suffer linear regret, because the adversary can choose a

[3] The theorem also requires a range of size at least 3.

sequence of actions $y^1, \ldots, y^T$ on which the algorithm makes sub-optimal decisions at every round. Therefore, randomization is one of the key aspects of no-regret algorithms.

Many online no-regret algorithms are known for the full information and the partial information settings. In particular, the HEDGE algorithm [Freund and Schapire, 1995] is one of the earliest results in this space for the full information setting. At time $t+1$, HEDGE picks each action x with probability $p_x^{t+1} \propto \exp(-\eta F^t(x))$, for $F^t(x) = \sum_{s=1}^t f(x, y^s)$ and $\eta = \Theta(\sqrt{2\ln(|\mathcal{X}|)\,/\,T})$.

Proposition 2.2 (Freund and Schapire [1995]). HEDGE *has regret* $\mathbb{E}[Reg_T] \leq O\left(\sqrt{T\ln(|\mathcal{X}|)}\right)$.

For the partial information setting, the EXP3 algorithm of Auer et al. [2002] can be thought of as a variant of the HEDGE algorithm with importance weighting. In particular, at time $t+1$, EXP3 picks each action x with probability $p_x^{t+1} \propto \exp(-\eta \tilde{F}^t(x))$, for $\eta = \Theta(\sqrt{2\ln(|\mathcal{X}|)\,/\,T|\mathcal{X}|})$ and

$$\tilde{F}^t(x) = \sum_{s=1}^{t} \frac{\mathbb{1}_{(x^s = x)} f(x, y^s)}{p_x^s}. \tag{1}$$

In other words, EXP3 is similar to HEDGE, except that instead of taking into account the total loss of an action, $F^t(x)$, it takes into account an *estimate* of the loss, $\tilde{F}^t(x)$.

3 Problem Formulation

In this section, we formulate the question of how one can *design a weighting scheme that effectively weights the rankings of voters based on the history of their votes and the performance of the selected alternatives*.

We consider a setting where n voters participate in a sequence of elections that are decided by a known voting rule f. In each election, voters submit their rankings over a different set of m alternatives so as to elect a winner. Given an adversarial sequence of voters' rankings $\boldsymbol{\sigma}^{1:T}$ and alternative losses $\boldsymbol{\ell}^{1:T}$ over a span of T elections, the best voter is the one whose rankings lead to the election of the winners with smallest loss overall. We call this voter *the best voter in hindsight*. (See Section 6 for a discussion of a stronger benchmark: best *weight vector* in hindsight.)

When the sequence of elections is not known a priori, the best voter is not known either. In this case, the weighting scheme has to take an online approach to *weighting the voters' rankings*. That is, at each time step $t \leq T$, the weighting scheme chooses a weight vector $\mathbf{w}^t$, possibly at random, to weight the rankings of the voters. After the election is held, the weighting scheme receives some feedback regarding the quality of the alternatives in that election, typically in the form of the loss of the elected alternative or that of all alternatives. Using the feedback, the weighting scheme then re-weights the voters' rankings based on their performance so far. Our goal is to design a weighting scheme that weights the rankings of the voters at each time step, and elects winners with overall expected loss that is almost as small as that of the best voter in hindsight. We refer to the expected difference between these losses as the expected *regret*. That is,

$$\mathbb{E}[Reg_T] \triangleq \mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t}, \boldsymbol{\ell}^t) - \min_i \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t)\right],$$

where the expectation is taken over any additional source of randomness in the adversarial sequence or the algorithm. In particular, we seek a weighting scheme for which the average expected regret goes to zero as the time horizon T goes to infinity, at a rate that is polynomial in the number of voters and alternatives. That is, we wish to achieve $\mathbb{E}[Reg_T] = \mathrm{poly}(n, m) \cdot o(T)$. This is our version of a *no-regret* algorithm.

The type of the feedback is an important factor in designing a weighting scheme. Analogously to the online learning models described in Section 2.2, we consider two types of feedback, *full information* and *partial information*. In the full information case, after a winner is selected at time t, the quality of all alternatives and rankings of the voters at that round are revealed to the weighting scheme. Note that this information is sufficient for computing the loss of each voter's rankings so far. This would be the case, for example, if the alternatives are companies to invest in. On the other hand, in the partial information setting only the loss of the winner is revealed. This type of feedback is appropriate when the alternatives are product prototypes: we cannot know how successful an undeveloped prototype would have been, but obviously we can measure the success of a prototype that was selected for development. More formally, in the full information setting the choice of $\mathbf{w}^{t+1}$ can depend on $\boldsymbol{\sigma}^{1:t}$ and $\boldsymbol{\ell}^{1:t}$, while in the partial information setting it can only depend on $\boldsymbol{\sigma}^{1:t}$ and $\ell^s_{a^s}$ for $s \le t$, where a^s is the alternative that won the election at time s.

No doubt the reader has noted that the above problem formulation is closely related to the general setting of online learning. Using the language of online learning introduced in Section 2.2, the weight vector $\mathbf{w}^t$ corresponds to the learner's action x^t, the vote profile and alternative losses $(\boldsymbol{\sigma}^t, \boldsymbol{\ell}^t)$ correspond to the adversary's action y^t, the expected loss of the weighting scheme $L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t)$ corresponds to the loss of the learning algorithm $f(x^t, y^t)$, and the best-in-hindsight voter — or weight vector $\mathbf{e}_i$ — refers to the best-in-hindsight action.

4 Randomized Weights

In this section, we develop no-regret algorithms for the full information and partial information settings. We essentially require no assumptions on the voting rule, but also impose no restrictions on the weighting scheme. In particular, the weighting scheme may be randomized, that is, the weights can be sampled from a distribution over weight vectors. This allows us to obtain general positive results.

As we just discussed, our setting is closely related to the classic online learning setting. Here, we introduce an algorithm analogous to HEDGE that works in the full information setting of Section 3 and achieves no-regret guarantees.

Algorithm 1: Full information setting, using randomized weights.

Input: Adversarial sequences $\boldsymbol{\sigma}^{1:T}$ and $\boldsymbol{\ell}^{1:T}$, and parameter $\eta = \sqrt{2 \ln n / T}$
for $t = 1, \ldots, T$ **do**
  Play weight vector $\mathbf{e}_i$ with probability $p^t_i \propto \exp\left(-\eta \sum_{s=1}^{t-1} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^s,\mathbf{e}_i}, \boldsymbol{\ell}^s)\right)$.
  Observe $\boldsymbol{\ell}^t$ and $\boldsymbol{\sigma}^t$.
end

Theorem 4.1. *For any anonymous voting rule f and n voters, Algorithm 1 has regret $O(\sqrt{T \ln(n)})$ in the full information setting.*

Proof Sketch. At a high level, this algorithm only considers weight vectors that correspond to a single voter. At every time step, the algorithm chooses a distribution over such weight vectors and applies the voting rule to one such weight vector that is drawn at random from this distribution. This is equivalent to applying the HEDGE algorithm to a set of actions, each of which is a weight vector that corresponds to a single voter. In addition, the loss of the benchmark weighting scheme is the smallest loss that one can get from following one such weight vector. Therefore, the theorem follows from Proposition 2.2. □

Let us now address the partial information setting. One may wonder whether the above approach, i.e., reducing our problem to online learning and using a standard algorithm, directly extends to the partial information setting (with the EXP3 algorithm). The answer is that it does not. In particular, in the classic setting of online learning with partial information feedback, the algorithm can compute the estimated loss of the action it just played, that is, the algorithm can compute $f(x^t, y^t)$. In our problem setting, however, the weighting scheme only observes $\boldsymbol{\sigma}^t$ and $\ell^t_{a^t}$ for the specific alternative a^t that was elected at this time. Since the losses of other alternatives remain unknown, the weighting scheme cannot even compute the expected loss of the specific voter i^t it selected at time t, i.e., $L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_{i^t}}, \boldsymbol{\ell}^t)$. Therefore, we cannot directly use the EXP3 algorithm by imagining that the voters are actions, as we do not obtain the partial information feedback that the algorithm requires. Nevertheless, we can design a new algorithm inspired by EXP3.

Theorem 4.2. *For any anonymous voting rule f and n voters, Algorithm 2 has regret $O(\sqrt{Tn \ln(n)})$ in the partial information setting.*

The theorem's proof is relegated to Appendix B. In a nutshell, we show that certain properties, which are necessary for the performance of EXP3, still hold in our setting. Specifically, Lemma 4.3 (proof in Appendix B.1) asserts that $\tilde{\boldsymbol{\ell}}^t$ creates an unbiased estimator of the expected loss of the weighting

Algorithm 2: Partial information setting, using randomized weights.

Input: An adversarial sequences of $\boldsymbol{\sigma}^{1:T}$ and $\boldsymbol{\ell}^{1:T}$, and parameter $\eta = \sqrt{2\ln n/Tn}$.

Let $\tilde{\mathbf{L}}^0 = \mathbf{0}$.
for $t = 1, \ldots, T$ **do**
  for $i = 1, \ldots, n$ **do**
    Let $p_i^t \propto \exp(-\eta \tilde{L}_i^{t-1})$.
  end
  Play weight vector $\mathbf{e}_{i^t}$ from distribution $\mathbf{p}^t$, and observe the vote profile $\boldsymbol{\sigma}^t$, the alternative $a^t \sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}})$, and its loss $\ell^t_{a^t}$.
  Let $\tilde{\boldsymbol{\ell}}^t$ be the vector such that $\tilde{\ell}^t_{i^t} = \ell^t_{a^t}/p^t_{i^t}$ and $\tilde{\ell}^t_i = 0$ for $i \neq i^t$.
  Let $\tilde{\mathbf{L}}^t = \tilde{\mathbf{L}}^{t-1} + \tilde{\boldsymbol{\ell}}^t$.
end

scheme. Moreover, it states that for any voter i^*, $\tilde{L}^t_{i^*}$ is an unbiased estimator for the loss that the weighting scheme would have received if it followed the rankings of voter i^* throughout the sequence of elections. Lemma 4.4 (proof in Appendix B.2) then establishes that the variance of this estimator is small.

Lemma 4.3. *For any t, any i^*, $i^t \sim \mathbf{p}^t$, and $a^t \sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}})$, we have*

$$\mathbb{E}_{i^t, a^t}\left[\sum_{i=1}^n p_i^t \tilde{\ell}_i^t\right] = \mathbb{E}_{i^t}\left[L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}}, \boldsymbol{\ell}^t)\right] \quad \textit{and} \quad \mathbb{E}_{i^t, a^t}\left[\tilde{L}^T_{i^*}\right] = \sum_{t=1}^T L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^*}}, \boldsymbol{\ell}^t).$$

Lemma 4.4. *For any t, $i^t \sim \mathbf{p}^t$, and $a^t \sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}})$, we have*

$$\mathbb{E}_{i^t, a^t}\left[\sum_{i=1}^n p_i^t (\tilde{\ell}_i^t)^2\right] \leq n.$$

5 Deterministic Weights

One of the key aspects of the weighting schemes we used in the previous section is randomization. In such weighting schemes, the weights of the voters not only depend on their performance so far, but also on the algorithm's coin flips. In practice, voters would most likely prefer weighting schemes that depend only on their past performance, and are therefore easier to interpret.

In this section, we focus on designing weighting schemes that are deterministic in nature. Formally, a *deterministic weighting scheme* is an algorithm that at time step $t+1$ deterministically chooses one weight vector $\mathbf{w}^{t+1}$ based on the history of play, i.e., sequences $\boldsymbol{\sigma}^{1:t}$, $\boldsymbol{\ell}^{1:t}$, and $a^{1:t}$. In this section, we seek an answer to the following question: *For which voting rules is there a no-regret deterministic weighting scheme?* In contrast to the results established in the previous section, we find that the properties of the voting rule play an important role here. In the remainder of this section, we show possibility and impossibility results for the existence of such weighting schemes under randomized and deterministic voting rules.

We begin our search for deterministic weighting schemes by considering deterministic voting rules. Note that in this case the winning alternatives are induced deterministically by the weighting scheme, so the weight vector $\mathbf{w}^{t+1}$ should be deterministically chosen based on the sequences $\boldsymbol{\sigma}^{1:t}$ and $\boldsymbol{\ell}^{1:t}$. We establish an impossibility result: Essentially no deterministic weighting scheme is no-regret for a deterministic voting rule. Specifically, we show that a deterministic no-regret weighting scheme exists for a deterministic voting rule if and only if the voting rule is constant on unanimous profiles.

Definition 5.1. *A voting rule f is* constant on unanimous profiles *if and only if for all $\sigma, \sigma' \in \mathcal{L}(A)$, $f(\mathbf{e}_\sigma) = f(\mathbf{e}_{\sigma'})$, where $\mathbf{e}_\sigma$ denotes the anonymous vote profile that has all of its weight on ranking σ.*

Theorem 5.2. *For any deterministic voting rule f, a deterministic weighting scheme with regret $o(T)$ exists if and only if f is constant on unanimous profiles. This is true in both the full information and partial information settings.*

Proof. We first prove that for any voting rule that is constant on unanimous profiles there exists a deterministic weighting scheme that is no-regret. Consider such a voting rule f and a simple deterministic weighting scheme that uses weight vector $\mathbf{w}^t = \mathbf{e}_1$ for every time step $t \leq T$ (so it does not use feedback — whether full or partial — at all). Note that at each time step t and for any voter $i \in [n]$,

$$f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}) = f(\mathbf{e}_{\sigma_1^t}) = f(\mathbf{e}_{\sigma_i^t}) = f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}),$$

where the second transition holds because f is constant on unanimous profiles. As a result, $L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) = L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}, \boldsymbol{\ell}^t)$. In words, the total loss of the weighting scheme is the same as the total loss of any individual voter — this weighting scheme has 0 regret.

Next, we prove that if f is not constant on unanimous profiles then for any deterministic weighting scheme there is an adversarial sequence of $\boldsymbol{\sigma}^{1:T}$ and $\boldsymbol{\ell}^{1:T}$ that leads to regret of $\Omega(T)$, even in the full information setting. Take any such voting rule f and let $\tau, \tau' \in \mathcal{L}(A)$ be such that $f(\mathbf{e}_\tau) \neq f(\mathbf{e}_{\tau'})$. At time t, the adversary chooses $\boldsymbol{\sigma}^t$ and $\boldsymbol{\ell}^t$ based on the deterministic weight vector $\mathbf{w}^t$ as follows: The adversary sets $\boldsymbol{\sigma}^t$ to be such that $\sigma_1^t = \tau$ and $\sigma_j^t = \tau'$ for all $j \neq 1$. Let alternative a^t be the winner of profile $\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}$, i.e., $f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}) = \mathbf{e}_{a^t}$. The adversary sets $\ell_{a^t}^t = 1$ and $\ell_x^t = 0$ for all $x \neq a^t$. Therefore, the weighting scheme incurs a loss of 1 at every step, and its total loss is

$$\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) = \sum_{t=1}^{T} \ell_{a^t}^t = T.$$

Let us consider the total loss that the ranking of any individual voter incurs. By design, for any $j > 1$, $f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_1}) = f(\mathbf{e}_\tau) \neq f(\mathbf{e}_{\tau'}) = f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_j})$. Therefore, for at least one voter $i \in [n]$, $f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}) \neq \mathbf{e}_{a^t}$. Note that such a voter receives loss of 0, so the combined loss of all voters is at most $n-1$. Over all time steps, the total combined loss of all voters is at most $T(n-1)$. As a result, the best voter incurs a loss of at most $\frac{(n-1)T}{n}$, i.e., the average loss. We conclude that the regret of the weighting scheme is

$$Reg_T = \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) - \min_{i\in[n]} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}, \boldsymbol{\ell}^t) \geq T - \frac{(n-1)T}{n} = \frac{T}{n}.$$

□

Theorem 5.2 indicates that we need to allow randomness (either in the weighting scheme or in the voting rule) if we wish to have no-regret guarantees. As stated before, we would like to have a deterministic weighting scheme so that the weights of voters are not decided by coin flips. This leaves us with no choice other than having a randomized voting rule. Nonetheless, one might argue in favor of having a deterministic voting rule and a randomized weighting scheme, claiming that it is equivalent because the randomness has simply been shifted from the voting rule to the weights. To that imaginary critic we say that allowing the voting rule to be randomized makes it possible to achieve strategyproofness (see Section 2.1), which cannot be satisfied by a deterministic voting rule.

We next show that for any voting rule that is a distribution over unilaterals there exist deterministic weighting schemes that are no-regret. An important family of strategyproof randomized voting rules — randomized positional scoring rules (see Appendix A) — can be represented as distributions over unilaterals, hence the theorem allows us to design a no-regret weighting scheme for any randomized positional scoring rule.

The weighting schemes that we use build on Algorithms 1 and 2 directly. In more detail, we consider deterministic weighting schemes that at time t use weight vector $\mathbf{p}^t$ and a randomly drawn candidate $a^t \sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{p}^t})$, where $\mathbf{p}^t$ is computed according to Algorithms 1 or 2. The key insight behind these weighting schemes is that, if f is a distribution over unilaterals, we have

$$\mathbb{E}_{i\sim\mathbf{p}^t}[f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i})] = f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{p}^t}), \tag{2}$$

where the left-hand side is a vector of expectations. That is, the outcome of the voting rule $f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{p}^t})$ can be alternatively implemented by applying the voting rule on the ranking of voter i that is drawn at random from the distribution $\mathbf{p}^t$. This is exactly what Algorithms 1 and 2 do. Therefore, the deterministic weighting schemes induce the same distribution over alternatives at every time step as their randomized counterparts, and achieve the same regret. The next theorem, whose full proof appears in Appendix C.1, formalizes this discussion.

Theorem 5.3. *For any voting rule that is a distribution over unilaterals, there exist deterministic weighting schemes with regret of $O(\sqrt{T\ln(n)})$ and $O(\sqrt{Tn\ln(n)})$ in the full-information and partial-information settings, respectively.*

The theorem states that there exist no-regret deterministic weighting schemes for any voting rule that is a distribution over unilaterals. It is natural to ask whether being a distribution over unilaterals is, in some sense, also a necessary condition. While we do not give a complete answer to this question, we are able to identify a sufficient condition for *not* having no-regret deterministic weighting schemes.

To this end, we introduce a classic concept. Alternative $a \in A$ is a *Condorcet winner* in a given vote profile if for every $b \in A$, a majority of voters rank a above b. A deterministic rule is *Condorcet consistent* if it selects a Condorcet winner whenever one exists in the given vote profile; see Appendix A for formal definitions. We extend the notion of Condorcet consistency to randomized rules.

Definition 5.4. *For a set of alternatives A such that $|A| = m$, a randomized voting rule $f : \Delta(\mathcal{L}(A)) \to \Delta(A)$ is* probabilistically Condorcet consistent with gap $\delta(m)$ *if for any anonymous vote profile $\boldsymbol{\pi}$ that has a Condorcet winner a, and for all alternatives $x \in A \setminus \{a\}$, $f(\boldsymbol{\pi})_a \geq f(\boldsymbol{\pi})_x + \delta(m)$.*

In words, a randomized voting rule is probabilistically Condorcet consistent if the Condorcet winner has strictly higher probability of being selected than any other alternative, by a gap of $\delta(m)$. As an example, a significant strategyproof randomized voting rule — the randomized Copeland rule, defined in Appendix A — is probabilistically Condorcet consistent with gap $\delta(m) = \Omega(1/m^2)$.

Theorem 5.5. *For a set of alternatives A such that $|A| = m$, let f be a probabilistically Condorcet consistent voting rule with gap $\delta(m)$, and suppose there are n voters for $n \geq 2(\frac{3}{2\delta(m)} + 1)$. Then any deterministic weighting scheme will suffer regret of $\Omega(T)$ under f (in the worst case), even in the full information setting.*

The theorem's proof is relegated to Appendix C.2. It is interesting to note that Theorems 5.3 and 5.5 together imply that distributions over unilaterals are not probabilistically Condorcet consistent. This is actually quite intuitive: Distributions over unilaterals are "local" in that they look at each voter separately, whereas Condorcet consistency is a global property. In fact, these theorems can be used to prove — in an especially convoluted and indirect way — a simple result from social choice theory [Moulin, 1983]: No positional scoring rule is Condorcet consistent!

6 Discussion

We conclude by discussing several conceptual points.

A natural, stronger benchmark. In our model (see Section 3), we are competing with the best voter in hindsight. But our action space consists of *weight vectors*. It is therefore natural to ask whether we can compete with the best weight vector in hindsight (hereinafter, the *stronger benchmark*). Clearly the stronger benchmark is indeed at least as hard, because the best voter i^* corresponds to the weight vector $\mathbf{e}_{i^*}$. Therefore, our impossibility results for competing against the best voter in hindsight (Theorems 5.2 and 5.5) extend to the stronger benchmark. Moreover, voting rules that are distributions over unilaterals demonstrate a certain linear structure where the outcome of the voting rule nicely decomposes across individual voters. Under such voting rules, the benchmark of best weights in hindsight is equivalent to the benchmark of best voter in hindsight. Therefore, Theorem 5.3 also holds for the stronger benchmark, and, in summary, each and every result of Section 5 extends to the stronger benchmark. By contrast, Theorems 4.1 and 4.2 do not hold for the stronger benchmark; the question of identifying properties of voting rules (beyond distributions over unilaterals) that admit *randomized* no-regret weighting schemes under the stronger benchmark remains open. We describe the stronger benchmark in more detail, and formalize the above arguments, in Appendix D.

Changing the sets of alternatives and voters over time. We wish to emphasize that the set of alternatives at each time step, i.e., in each election, can be completely different. Moreover, the *number* of alternatives could be different. In fact, our positive results do not even depend on the number of alternatives m, so we can simply set m to be an upper bound. By contrast, we do need the set of voters to stay fixed throughout the process, but this is consistent with our motivating examples (e.g., a group of partners in a small venture capital firm would face different choices at every time step, but the composition of the group rarely changes).

Optimizing the voting rule. Throughout the paper, the voting rule is exogenous. One might ask whether it makes sense to optimize the choice of voting rule itself, in order to obtain good no-regret learning results. Our answer is "yes and no". On the one hand, we believe our results do give some guidance on choosing between voting rules. For example, from this viewpoint, one might prefer randomized Borda (which admits no-regret algorithms under a deterministic weighting scheme) to randomized Copeland (which does not). On the other hand, many considerations are factored into the choice of voting rule: social choice axioms, optimization of additional objectives [Procaccia et al., 2016, Boutilier et al., 2015, Elkind et al., 2009, Conitzer and Sandholm, 2005], and simplicity. It is therefore best to think of our approach as *augmenting* voting rules that are already in place.

References

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227:190–213, 2015.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

S. Chaudhuri and A. Tewari. Online ranking with top-1 feedback. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 129–137, 2015.

R. Congleton. *Perfecting Parliament: Constitutional Reform, Liberalism, and the Rise of Western Democracy*. Cambridge University Press, 2011.

V. Conitzer and T. Sandholm. Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 145–152, 2005.

V. Conitzer and T. Sandholm. Nonexistence of voting rules that are usually hard to manipulate. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, pages 627–634, 2006.

V. Conitzer, T. Sandholm, and J. Lang. When are elections with few candidates hard to manipulate? *Journal of the ACM*, 54(3):1–33, 2007.

E. Elkind, P. Faliszewski, and A. Slinko. On distance rationalizability of some voting rules. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 108–117, 2009.

P. Faliszewski, E. Hemaspaandra, , and H. Schnoor. Copeland voting: Ties matter. In *Proceedings of the 7th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 983–990, 2008.

P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. How hard is bribery in elections? *Journal of Artificial Intelligence Research*, 35:485–532, 2009.

P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. Weighted electoral control. *Journal of Artificial Intelligence Research*, 52:507–542, 2015.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the 2nd European Conference on Computational Learning Theory (EuroCOLT)*, pages 23–37, 1995.

A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–602, 1973.

A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45:665–681, 1977.

H. Moulin. *The Strategy of Social Choice*, volume 18 of *Advanced Textbooks in Economics*. North-Holland, 1983.

A. D. Procaccia. Can approximation circumvent Gibbard-Satterthwaite? In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pages 836–841, 2010.

A. D. Procaccia, N. Shah, and Y. Zick. Voting rules as error-correcting codes. *Artificial Intelligence*, 231:1–16, 2016.

F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 784–791, 2008.

M. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.

M. Zuckerman, A. D. Procaccia, and J. S. Rosenschein. Algorithms for the coalitional manipulation problem. *Artificial Intelligence*, 173(2):392–412, 2009.

A Voting Rules

This appendix provides additional background on social choice theory. It is not required to understand the rest of the paper, but may be helpful in putting our results in context.

A.1 Examples of Anonymous Voting Rules

One class of anonymous voting rules uses the positions of the individual alternatives in order to determine the winners. These rules, collectively called *positional scoring rules*, are defined by a scoring vector $\mathbf{s}$ such that $s_1 \geq s_2 \geq \cdots \geq s_m \geq 0$. Given a vote σ, the score of alternative $a \in A$ in σ is the score of its position in σ, i.e., $s_{\sigma(a)}$. Given an anonymous vote profile $\boldsymbol{\pi}$, the score of an alternative is its overall score in the rankings of $\boldsymbol{\pi}$, that is,

$$s\text{-score}_{\boldsymbol{\pi}}(a) \triangleq \sum_{\sigma \in \mathcal{L}(A)} \pi_\sigma s_{\sigma(a)}.$$

A *deterministic positional scoring rule* chooses the alternative with the highest score, i.e., $f(\boldsymbol{\pi}) = \mathbf{e}_{a^*}$, where $a^* \in \arg\max_{a \in A} s\text{-score}_{\boldsymbol{\pi}}(a)$ (tie breaking may be needed). On the other hand, a *randomized positional scoring rule* chooses each alternative with probability proportional to its score, i.e., $f(\boldsymbol{\pi})_a \propto s\text{-score}_{\boldsymbol{\pi}}(a)$ for all $a \in A$. Examples of positional scoring rules include *plurality* with $\mathbf{s} = (1, 0, \ldots, 0)$, *veto* with $\mathbf{s} = (1, \ldots, 1, 0)$, and *Borda* with $\mathbf{s} = (m-1, m-2, \ldots, 0)$.

Another class of anonymous voting rules uses pairwise comparisons between the alternatives to determine the winners. We are especially interested in the *Copeland* rule, which assigns a score to each alternative based on the number of pairwise majority contests it wins. In an anonymous vote profile $\boldsymbol{\pi}$, we denote by $a >_{\boldsymbol{\pi}} b$ the event that a beats b in a pairwise competition, i.e., a is preferred to b in rankings in $\boldsymbol{\pi}$ that collectively have more than half the weight. More formally, $\sum_{\sigma \in \mathcal{L}(A)} \pi_\sigma \mathbb{1}_{(a \succ_\sigma b)} > 1/2$. We also write $a =_{\boldsymbol{\pi}} b$ if they are tied, i.e., $\sum_{\sigma \in \mathcal{L}(A)} \pi_\sigma \mathbb{1}_{(a \succ_\sigma b)} = 1/2$. The Copeland score[4] of an alternative is defined by

$$\text{C-score}_{\boldsymbol{\pi}}(a) \triangleq \left|\{b \in A \mid a >_{\boldsymbol{\pi}} b\}\right| + \frac{1}{2} \cdot \left|\{b \in A \mid a =_{\boldsymbol{\pi}} b\}\right|.$$

The *deterministic Copeland rule* chooses the alternative that has the highest Copeland score (possibly breaking ties), and the *randomized Copeland rule* chooses each alternative with probability proportional to its Copeland score.

These notations allow us to formally define the notion of Condorcet consistency (informally introduced in Section 5). We say that $a \in A$ is a *Condorcet winner* in the vote profile $\boldsymbol{\pi}$ if $a >_{\boldsymbol{\pi}} b$ for all $b \in A \setminus \{a\}$. A voting rule is *Condorcet consistent* if it selects a Condorcet winner whenever one exists in the given vote profile. Note that the Copeland score of a Condorcet winner is $m-1$, whereas the Copeland score of any other alternative must be strictly smaller, so a Condorcet winner (if one exists) indeed has maximum Copeland score.

A.2 Strategyproofness, More Formally

An anonymous deterministic voting rule f is called *strategyproof* if for any voter $i \in [n]$, any two vote profiles $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ for which $\sigma_j = \sigma'_j$ for all $j \neq i$, and any weight vector $\mathbf{w}$, it holds that either $a = a'$ or $a \succ_{\sigma_i} a'$, where a and a' are the winning alternatives in $f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}})$ and $f(\boldsymbol{\pi}_{\boldsymbol{\sigma}',\mathbf{w}})$ respectively. In words, whenever a voter reports σ'_i instead of σ_i, the outcome does not improve according to the true ranking σ_i.

To extend this definition to randomized rules, we require some additional definitions. Given a *loss function* over the alternatives denoted by a vector $\boldsymbol{\ell} \in [0,1]^m$, the expected loss of the alternative chosen by the rule f under an anonymous vote profile $\boldsymbol{\pi}$ is

$$L_f(\boldsymbol{\pi}, \boldsymbol{\ell}) \triangleq \mathbb{E}_{a \sim f(\boldsymbol{\pi})}[\ell_a] = f(\boldsymbol{\pi}) \cdot \boldsymbol{\ell}.$$

[4] Some refer to this variant of Copeland as $\text{Copeland}_{1/2}$ [Faliszewski et al., 2008].

The higher the loss, the worse the alternative. We say that the loss function $\boldsymbol{\ell}$ is *consistent* with vote $\sigma \in \mathcal{L}(A)$ if for all $a, b \in A$, $a \succ_\sigma b \Leftrightarrow \ell_a < \ell_b$. An anonymous randomized rule f is *strategyproof* if for any voter $i \in [n]$, any two vote profiles $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ for which $\sigma_j = \sigma'_j$ for all $j \neq i$, any weight vector $\mathbf{w}$, and any loss function $\boldsymbol{\ell}$ that is consistent with σ_i, we have $L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}, \boldsymbol{\ell}) \leq L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}',\mathbf{w}}, \boldsymbol{\ell})$.

As noted in Section 2.1, randomized positional scoring rules, and the randomized Copeland rule, are known to be strategyproof. To see why they satisfy Gibbard's necessary condition (Proposition 2.1), a randomized positional scoring rule with score vector $\mathbf{s}$ is a distribution with probabilities proportional to $s_1, \ldots, s_m$ over anonymous unilateral rules $g_1, \ldots, g_m$, where each g_i corresponds to the function $h_i(\sigma)$ that returns the alternative ranked at position i of σ. Similarly, the randomized Copeland rule is a uniform distribution over duples $g_{a,b}$ for any two different $a, b \in A$, where $g_{a,b}(\boldsymbol{\pi}) = \mathbf{e}_a$ if $a >_{\boldsymbol{\pi}} b$, $g_{a,b}(\boldsymbol{\pi}) = \mathbf{e}_b$ if $b >_{\boldsymbol{\pi}} a$, and $(g_{a,b}(\boldsymbol{\pi}))_a = (g_{a,b}(\boldsymbol{\pi}))_b = 1/2$ if $a =_{\boldsymbol{\pi}} b$.

B Omitted Proof from Section 4

B.1 Proof of Lemma 4.3

For ease of notation, we suppress t when it is clear from the context. First note that $\tilde{\boldsymbol{\ell}}$ is zero in all of its elements, except for $\tilde{\ell}_{i^t}$. So,

$$\sum_{i=1}^{n} p_i \tilde{\ell}_i = p_{i^t} \tilde{\ell}_{i^t} = p_{i^t} \frac{\ell_{a^t}}{p_{i^t}} = \ell_{a^t}.$$

Therefore, we have

$$\mathbb{E}_{i^t,a^t}\left[\sum_{i=1}^{n} p_i \tilde{\ell}_i\right] = \mathbb{E}_{i^t,a^t}\left[\ell_{a^t}\right] = \mathbb{E}_{i^t}\left[L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{e}_{i^t}}, \boldsymbol{\ell})\right].$$

For clarity of presentation, let $\tilde{\boldsymbol{\ell}}^{i,a}$ be an alternative representation of $\tilde{\boldsymbol{\ell}}$ when $i^t = i$ and $a^t = a$. Note that $\ell^{i,a}_{i^*} \neq 0$ only if $i^* = i$. We have

$$\begin{aligned}
\mathbb{E}_{i^t,a^t}\left[\tilde{L}^T_{i^*}\right] &= \sum_{t=1}^{T} \mathbb{E}_{i^t,a^t}\left[\tilde{\ell}^{i^t,a^t}_{i^*}\right] = \sum_{t=1}^{T}\sum_{i=1}^{n} p^t_i\, \mathbb{E}_{a\sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i})}\left[\tilde{\ell}^{i,a}_{i^*}\right] = \sum_{t=1}^{T} p^t_{i^*}\, \mathbb{E}_{a\sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_{i^*}})}\left[\frac{\ell^t_a}{p^t_{i^*}}\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{a\sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_{i^*}})}\left[\ell^t_a\right] = \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_{i^*}}, \boldsymbol{\ell}^t).
\end{aligned}$$

□

B.2 Proof of Lemma 4.4

For ease of notation, we suppress t when it is clear from the context. Since $\tilde{\boldsymbol{\ell}}$ is zero in all of its elements, except for $\tilde{\ell}_{i^t}$, we have

$$\sum_{i=1}^{n} p_i (\tilde{\ell}_i)^2 = p_{i^t}(\tilde{\ell}_{i^t})^2 = p_{i^t}\left(\frac{\ell_{a^t}}{p_{i^t}}\right)^2 = \frac{(\ell_{a^t})^2}{p_{i^t}}.$$

Therefore,

$$\mathbb{E}_{i^t,a^t}\left[\sum_{i=1}^{n} p_i(\tilde{\ell}_i)^2\right] = \mathbb{E}_{i^t,a^t}\left[\frac{(\ell_{a^t})^2}{p_{i^t}}\right] = \sum_{i=1}^{n} p_i\, \mathbb{E}_{a\sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{e}_i})}\left[\frac{(\ell_a)^2}{p_i}\right] = \sum_{i=1}^{n} \mathbb{E}_{a\sim f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{e}_i})}\left[(\ell_a)^2\right] \leq n.$$

□

B.3 Proof of Theorem 4.2

We use a potential function, given by $\Phi^t \triangleq -\frac{1}{\eta}\ln\left(\sum_{i=1}^{n}\exp(-\eta\tilde{L}^{t-1}_i)\right)$. We prove the claim by analyzing the expected increase in this potential function at every time step. Note that

$$\Phi_{t+1} - \Phi_t = -\frac{1}{\eta}\ln\left(\frac{\sum_{i=1}^{n}\exp(-\eta\tilde{L}^{t-1}_i - \eta\tilde{\ell}^t_i)}{\sum_{i=1}^{n}\exp(-\eta\tilde{L}^{t-1}_i)}\right) = -\frac{1}{\eta}\ln\left(\sum_{i=1}^{n} p^t_i \exp(-\eta\tilde{\ell}^t_i)\right). \quad (3)$$

Taking the expected increase in the potential function over the random choices of i^t and a^t for all $t = 1, \dots, T$, we have

$$\begin{aligned}
\mathbb{E}\left[\Phi_{T+1} - \Phi_1\right] &= \sum_{t=1}^{T} \mathbb{E}_{i^t, a^t}\left[\Phi_{t+1} - \Phi_t\right] \\
&\geq \sum_{t=1}^{T} \mathbb{E}_{i^t, a^t}\left[-\frac{1}{\eta} \ln\left(\sum_{i=1}^{n} p_i^t \left(1 - \eta \tilde{\ell}_i^t + \frac{1}{2}\left(\eta \tilde{\ell}_i^t\right)^2\right)\right)\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{i^t, a^t}\left[-\frac{1}{\eta} \ln\left(1 - \eta\left(\sum_{i=1}^{n} p_i^t \tilde{\ell}_i^t - \frac{\eta}{2}\sum_{i=1}^{n} p_i^t \left(\tilde{\ell}_i^t\right)^2\right)\right)\right] \\
&\geq \sum_{t=1}^{T} \mathbb{E}_{i^t, a^t}\left[\sum_{i=1}^{n} p_i^t \tilde{\ell}_i^t - \frac{\eta}{2}\sum_{i=1}^{n} p_i^t \left(\tilde{\ell}_i^t\right)^2\right] \\
&\geq \mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}}, \boldsymbol{\ell}^t)\right] - \frac{\eta T n}{2},
\end{aligned} \tag{4}$$

where the second transition follows from Equation (3) because for all $x \geq 0$, $e^{-x} \leq 1 - x + \frac{x^2}{2}$, the fourth transition follows from $\ln(1-x) \leq -x$ for all $x \in \mathbb{R}$, and the last transition holds by Lemmas 4.3 and 4.4. On the other hand, $\Phi_1 = -\frac{1}{\eta}\ln n$ and for any i^*,

$$\Phi_{T+1} \leq -\frac{1}{\eta}\ln\left(\exp(-\eta \tilde{L}_{i^*}^T)\right) = \tilde{L}_{i^*}^T.$$

Therefore,

$$\mathbb{E}\left[\Phi_{T+1} - \Phi_1\right] \leq \mathbb{E}\left[\tilde{L}_{i^*}^T + \frac{1}{\eta}\ln n\right] = \mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^*}}, \boldsymbol{\ell}^t) + \frac{1}{\eta}\ln n\right]. \tag{5}$$

We can now prove the theorem by using Equations (4) and (5), and the parameter value $\eta = \sqrt{2\ln n / Tn}$:

$$\mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_{i^t}}, \boldsymbol{\ell}^t) - \min_{i \in [n]} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t)\right] \leq \frac{1}{\eta}\ln n + \frac{\eta T n}{2} \leq \sqrt{2Tn\ln n}.$$

□

C Omitted Proofs from Section 5

C.1 Proof of Theorem 5.3

Let f be a distribution over unilaterals $g_1, \dots, g_k$ with corresponding probabilities $q_1, \dots, q_k$. Also, let $h_j : \mathcal{L}(A) \to A$ denote the function corresponding to g_j, for $j \in [k]$. We first prove Equation (2). For ease of exposition we suppress t in the notations, when it is clear from the context. Furthermore, let $\boldsymbol{\pi}^i = \boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}$. It holds that

$$\mathbb{E}_{i \sim \mathbf{p}^t}\left[f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i})\right] = \sum_{i=1}^{n} p_i^t f(\boldsymbol{\pi}^i) = \sum_{i=1}^{n} p_i^t \sum_{j=1}^{k} q_j \sum_{\tau \in \mathcal{L}(A)} \pi_\tau^i \mathbf{e}_{h_j(\tau)} = \sum_{i=1}^{n} p_i^t \sum_{j=1}^{k} q_j \mathbf{e}_{h_j(\sigma_i)},$$

where the last equality follows by the fact that $\pi_{\sigma_i}^i = 1$ and $\pi_\tau^i = 0$ for any $\tau \neq \sigma_i$. Moreover, let $\boldsymbol{\pi} = \boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{p}^t}$, then

$$f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{p}^t}) = \sum_{j=1}^{k} q_j \sum_{\tau \in \mathcal{L}(A)} \pi_\tau \mathbf{e}_{h_j(\tau)} = \sum_{j=1}^{k} q_j \sum_{\tau \in \mathcal{L}(A)} \mathbf{e}_{h_j(\tau)} \sum_{i=1}^{n} p_i^t \mathbb{1}_{(\sigma_i = \tau)} = \sum_{i=1}^{n} p_i^t \sum_{j=1}^{k} q_j \mathbf{e}_{h_j(\sigma_i)}.$$

Now that we have established Equation (2), we use it to conclude that

$$\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{p}^t}, \boldsymbol{\ell}^t) - \min_{i \in [n]} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t) = \mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t) - \min_{i \in [n]} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t)\right],$$

where the expectation is taken over choice of $i \sim \mathbf{p}^t$ for all t. Therefore, the deterministic weighting schemes that use weight vector $\mathbf{p}^t$ achieve the same regret bounds as those established in Theorems 4.1 and 4.2. □

C.2 Proof of Theorem 5.5

We start by proving the following technical lemma.

Lemma C.1. *Let $x_1, x_2, \cdots x_n$ be n real numbers such that $x_i \geq x_{i+1}$ for all $i \in [n-1]$, and denote $S = \sum_{i=1}^{n} x_i$. Then for any $j \in [n]$, $\sum_{i=1}^{j} x_i \geq j\frac{S}{n}$.*

Proof. Assume for the sake of contradiction that there exists $j \in [n-1]$ such that $\sum_{i=1}^{j} x_i < j\frac{S}{n}$. It follows that there is $i \in [j]$ such that $x_i < \frac{S}{n}$. In addition, it must be the case that $\sum_{i=j+1}^{n} x_i > (n-j)\frac{S}{n}$, which implies that there is $i' \in \{j+1, \ldots, n\}$ such that $x_{i'} > \frac{S}{n}$. This contradicts the fact that $x_i \geq x_{i'}$. $\square$

Proof of Theorem 5.5. Fix an arbitrary deterministic weighting scheme. We will show that the loss of this weighting scheme is strictly higher than the average loss of the voters (for appropriately chosen vote profiles and loss functions) at every time step t, which directly leads to linear regret.

Consider an arbitrary time step $t \leq T$, and let $\mathbf{w}^t$ denote the weights chosen by the weighting scheme. To construct the vote profile $\boldsymbol{\sigma}^t$, the adversary first partitions the voters into two sets N_1^t and N_2^t, as follows: It sorts the weights $\mathbf{w}^t$ in non-increasing order, and then it adds voters to N_1^t by their sorted weight (largest to smallest) until

$$W_1^t \triangleq \sum_{i \in N_1^t} w_i^t > \frac{1}{2}\|\mathbf{w}^t\|_1,$$

that is, until the voters in N_1^t have more than half the total weight. The remaining voters form set N_2^t.

Now, let $\tau^{x,y} \in \mathcal{L}(A)$ denote a ranking that places x at the top (i.e., $\tau^{x,y}(x) = 1$) and y in second place (i.e., $\tau^{x,y}(y) = 2$). Let a and b be two alternatives such that $f(\mathbf{e}_{\tau^{b,a}})_b - f(\mathbf{e}_{\tau^{b,a}})_a \geq f(\mathbf{e}_{\tau^{a,b}})_a - f(\mathbf{e}_{\tau^{a,b}})_b$, i.e., the gap between the probabilities of picking the top two alternatives in $\mathbf{e}_{\tau^{b,a}}$ is at least the corresponding gap in $\mathbf{e}_{\tau^{a,b}}$. The adversary sets the vote profile $\boldsymbol{\sigma}^t$ such that $\sigma_i^t = \tau^{a,b}$ for all $i \in N_1^t$ and $\sigma_i^t = \tau^{b,a}$ for all $i \in N_2^t$. Also, it sets the loss function $\boldsymbol{\ell}^t$ to be $\ell_a^t = 1$, $\ell_b^t = 0$, and $\ell_x^t = 1/2$ for all $x \in A \setminus \{a, b\}$.

Observe that for all $i \in N_1^t$, $a \succ_{\sigma_i} x$ for all $x \in A \setminus \{a\}$. Since the total weight of voters in N_1^t is more than $1/2$, a is a Condorcet winner in $\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t}$. Therefore, because f is probabilistically Condorcet consistent with gap $\delta(m)$, it holds that

$$f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_a \geq f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_b + \delta(m).$$

It follows that the loss of the weighting scheme is

$$\begin{aligned} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t}, \boldsymbol{\ell}^t) &= 1 \cdot f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_a + \frac{1}{2} \cdot (1 - f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_a - f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_b) \\ &= \frac{1}{2} + \frac{1}{2}\left(f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_a - f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{w}^t})_b\right) \\ &\geq \frac{1}{2} + \frac{1}{2}\delta(m). \end{aligned} \tag{6}$$

Similarly, the loss of voter i is

$$L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t) = L_f(\mathbf{e}_{\sigma_i^t}, \boldsymbol{\ell}^t) = \frac{1}{2} + \frac{1}{2}\left(f(\mathbf{e}_{\sigma_i^t})_a - f(\mathbf{e}_{\sigma_i^t})_b\right). \tag{7}$$

Let $\mathbf{q}^1$ denote $f(\mathbf{e}_{\tau^{a,b}})$, i.e. the distribution over the alternatives for the votes of voters in N_1^t, and let $\mathbf{q}^2$ denote $f(\mathbf{e}_{\tau^{b,a}})$, i.e. the distribution over the alternatives for the votes of voters in N_2^t. Using these notations and Equation (7), the loss of a voter $i \in N_1^t$ is

$$L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t) = \frac{1}{2} + \frac{1}{2}\left(q_a^1 - q_b^1\right),$$

and the loss of a voter $i \in N_2^t$ is

$$L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t, \mathbf{e}_i}, \boldsymbol{\ell}^t) = \frac{1}{2} + \frac{1}{2}\left(q_a^2 - q_b^2\right) = \frac{1}{2} - \frac{1}{2}(q_b^2 - q_a^2).$$

Hence, the average loss over all voters is

$$\begin{aligned} L^t_{avg} &= \frac{|N^t_1|\left(\frac{1}{2}+\frac{1}{2}\left(q^1_a-q^1_b\right)\right)+(n-|N^t_1|)\left(\frac{1}{2}-\frac{1}{2}(q^2_b-q^2_a)\right)}{n} \\ &= \frac{1}{2}+\frac{1}{2n}\left(|N^t_1|(q^1_a-q^1_b)-(n-|N^t_1|)(q^2_b-q^2_a)\right). \end{aligned}$$

But we chose a and b such that $q^1_a - q^1_b \leq q^2_b - q^2_a$. We conclude that

$$\begin{aligned} L^t_{avg} &\leq \frac{1}{2}+\frac{1}{2n}\left(|N^t_1|(q^2_b-q^2_a)-(n-|N^t_1|)(q^2_b-q^2_a)\right) \\ &= \frac{1}{2}+\frac{1}{2}(q^2_b-q^2_a)\frac{(2|N^t_1|-n)}{n}. \end{aligned} \tag{8}$$

Our goal is to derive an upper bound on the expression $\frac{1}{2}(q^2_b-q^2_a)\frac{(2|N^t_1|-n)}{n}$. Specifically, we wish to prove that

$$\frac{1}{2}(q^2_b-q^2_a)\frac{(2|N^t_1|-n)}{n} \leq \frac{\delta(m)}{3}. \tag{9}$$

We do this by examining two cases.

Case 1: $W^t_1 \geq \left(\frac{1}{2}+\frac{\delta(m)}{3}\right)\|\mathbf{w}^t\|_1$**.** Informally, this is the case when the weights of N^t_1 overshot $\|\mathbf{w}^t\|_1/2$ by a fraction of at least $\delta(m)/3$. This means that the last voter added to N^t_1 has a weight of at least $W^t_1 - \frac{\|\mathbf{w}^t\|_1}{2}$. Since the weights were added in non-increasing order, it follows that each voter in N^t_1 has a weight of at least $W^t_1 - \frac{\|\mathbf{w}^t\|_1}{2}$. Therefore,

$$W^t_1 = \sum_{i\in N^t_1} w^t_i \geq \sum_{i\in N^t_1}\left(W^t_1 - \frac{\|\mathbf{w}^t\|_1}{2}\right) = |N^t_1|\left(W^t_1 - \frac{\|\mathbf{w}^t\|_1}{2}\right),$$

or equivalently,

$$|N^t_1| \leq \frac{1}{1-\frac{\|\mathbf{w}^t\|_1}{2W^t_1}}. \tag{10}$$

We have also assumed that $\frac{W^t_1}{\|\mathbf{w}^t\|_1} \geq \left(\frac{1}{2}+\frac{\delta(m)}{3}\right)$. Using Equation (10), we obtain

$$|N^t_1| \leq \frac{1}{1-\frac{1}{1+\frac{2\delta(m)}{3}}} = \frac{3}{2\delta(m)}+1. \tag{11}$$

Let us now examine the expression on the left-hand side of Equation (9). Note that b is a Condorcet winner in $\mathbf{e}_{\tau^{b,a}}$. Hence, $q^2_b \geq q^2_a + \delta(m)$, and, in particular, $q^2_b - q^2_a > 0$. In addition, we have assumed that $n \geq 2(\frac{3}{2\delta(m)}+1)$, which implies (by Equation (11)) that $n \geq 2|N^t_1|$. It follows that

$$\frac{1}{2}(q^2_b-q^2_a)\frac{(2|N^t_1|-n)}{n} \leq 0 \leq \frac{\delta(m)}{3},$$

thereby establishing Equation (9) for this case.

Case 2: $W^t_1 < \left(\frac{1}{2}+\frac{\delta(m)}{3}\right)\|\mathbf{w}^t\|_1$**.** Since N^t_1 contains voters who have the largest $|N^t_1|$ weights, Lemma C.1 implies that

$$W^t_1 = \sum_{i\in N^t_1} w^t_i \geq |N^t_1|\frac{\|\mathbf{w}^t\|_1}{n}.$$

We have also assumed that $W^t_1 < (\frac{1}{2}+\frac{\delta(m)}{3})\|\mathbf{w}^t\|_1$. Combining the last two inequalities, we obtain

$$|N^t_1| < n\left(\frac{1}{2}+\frac{\delta(m)}{3}\right). \tag{12}$$

Let us examine, once again, the left-hand side of Equation (9). Recall that $q_b^2 - q_a^2 > 0$, because b is a Condorcet winner in $\tau^{b,a}$. So, if $2|N_1^t| - n \leq 0$, then Equation (9) clearly holds, as in Case 1. And if $2|N_1^t| - n > 0$, the equation also holds, because

$$\frac{1}{2}(q_b^2 - q_a^2)\frac{(2|N_1^t| - n)}{n} \leq \frac{1}{2} \cdot 1 \cdot \frac{(2|N_1^t| - n)}{n} = \frac{|N_1^t|}{n} - \frac{1}{2} < \frac{\delta(m)}{3},$$

where the last inequality follows from Equation (12).

To complete the proof, we combine Equations (6), (8), and (9), to obtain

$$L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) \geq L_{avg}^t + \frac{\delta(m)}{6}.$$

The best voter in hindsight incurs loss that is at most as high as the average voter. Therefore, the overall regret is

$$\begin{aligned} Reg_T &= \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) - \min_i \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}, \boldsymbol{\ell}^t) \\ &\geq \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) - \sum_{t=1}^{T} L_{avg}^t \\ &\geq T\frac{\delta(m)}{6}. \end{aligned}$$

In words, the weighting scheme suffers linear regret. □

D The Stronger Benchmark: Best Weights in Hindsight

In this section, we discuss our results as they apply to the stronger benchmark of competing with the best voter weights in hindsight.

Our goal is to design a weighting scheme that weights the rankings of the voters at each time step, and elects winners with overall expected loss that is almost as small as that of the *best voter weights in hindsight*. We refer to the expected difference between these losses as the expected *regret* with respect to the best weight in hindsight benchmark. That is,

$$\mathbb{E}[Reg_T] \triangleq \mathbb{E}\left[\sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}^t}, \boldsymbol{\ell}^t) - \min_{\mathbf{w}:\|\mathbf{w}\|_1=1} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}}, \boldsymbol{\ell}^t)\right].$$

We wish to formalize the claim, made in Section 6, that Theorem 5.3 holds under the stronger benchmark. We do this by showing that, indeed, for distributions over unilaterals, the best-weights-in-hindsight benchmark is equivalent to the best voter in hindsight.

Theorem D.1. *For any voting rule that is a distribution over unilaterals, there exist deterministic weighting schemes with regret of $O(\sqrt{T\ln(n)})$ and $O(\sqrt{Tn\ln(n)})$ with respect to the best weight in hindsight benchmark, in the full-information and partial-information settings, respectively.*

Proof. It sufficies to show that

$$\min_{\mathbf{w}:\|\mathbf{w}\|_1=1} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{w}}, \boldsymbol{\ell}^t) = \min_{i\in[n]} \sum_{t=1}^{T} L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma}^t,\mathbf{e}_i}, \boldsymbol{\ell}^t), \tag{13}$$

as then the theorem follows from Theorem 5.3. In turn, to prove Equation (13) it is sufficient to show that $L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}, \boldsymbol{\ell})$ is a linear function in $\mathbf{w}$, because any linear function is optimized at an extreme point of the convex set $\{\mathbf{w} \mid \|\mathbf{w}\|_1 = 1\}$.

Let f be a distribution over unilaterals $g_1, \ldots, g_k$ with corresponding probabilities $q_1, \ldots, q_k$. Also, let $h_j : \mathcal{L}(A) \to A$ denote the function corresponding to g_j, for $j \in [k]$. Given a weight vector $\mathbf{w}$ such that $\|\mathbf{w}\|_1 = 1$ and $\boldsymbol{\sigma}$, let $\boldsymbol{\pi} = \boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}$. It holds that

$$f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}) = \sum_{j=1}^{k} q_j \sum_{\tau \in \mathcal{L}(A)} \pi_\tau \mathbf{e}_{h_j(\tau)} = \sum_{j=1}^{k} q_j \sum_{\tau \in \mathcal{L}(A)} \mathbf{e}_{h_j(\tau)} \sum_{i=1}^{n} w_i \mathbb{1}_{(\sigma_i = \tau)} = \sum_{i=1}^{n} w_i \sum_{j=1}^{k} q_j \mathbf{e}_{h_j(\sigma_i)}.$$

Therefore,

$$L_f(\boldsymbol{\pi}_{\boldsymbol{\sigma},\mathbf{w}}, \boldsymbol{\ell}) = \sum_{i=1}^{n} w_i \sum_{j=1}^{k} q_j \left(\mathbf{e}_{h_j(\sigma_i)} \cdot \boldsymbol{\ell}\right) = \sum_{i=1}^{n} w_i \sum_{j=1}^{k} q_j \ell_{h_j(\sigma_i)};$$

the right hand side is clearly linear in $\mathbf{w}$. $\square$