Optimal Rates and Efficient Algorithms for Online Bayesian Persuasion

Martino Bernasconi [1] Matteo Castiglioni [1] Andrea Celli [2] Alberto Marchesi [1] Francesco Trovò [1] Nicola Gatti [1]

Abstract

Bayesian persuasion studies how an informed sender should influence beliefs of rational receivers who take decisions through Bayesian updating of a common prior. We focus on the *online Bayesian persuasion* framework, in which the sender repeatedly faces one or more receivers with unknown and adversarially selected types. First, we show how to obtain a tight $\tilde{O}(T^{1/2})$ regret bound in the case in which the sender faces a single receiver and has partial feedback, improving over the best previously-known bound of $\tilde{O}(T^{4/5})$. Then, we provide the first no-regret guarantees for the multi-receiver setting under partial feedback. Finally, we show how to design no-regret algorithms with polynomial per-iteration running time by exploiting *type reporting*, thereby circumventing known intractability results on online Bayesian persuasion. We provide efficient algorithms guaranteeing a $O(T^{1/2})$ regret upper bound both in the single- and the multi-receiver scenario when type reporting is allowed.

1. Introduction

The *Bayesian persuasion* framework, introduced by Kamenica & Gentzkow (2011), is an economic model which helps to explain how individuals make decisions based on the information they receive from others, and how this information can be used to influence their behavior. This model is particularly useful for understanding strategic interactions in situations where individuals have different levels of information or expertise. The framework already found application in domains such as advertising (Bro Miltersen & Sheffet, 2012; Emek et al., 2014; Badanidiyuru et al., 2018; Castiglioni et al., 2022c; Bacchiocchi et al., 2022), voting (Cheng et al., 2015; Alonso & Câmara, 2016; Castiglioni et al., 2020a; Castiglioni & Gatti, 2021), routing (Vasserman et al., 2015; Bhaskar et al., 2016; Castiglioni et al., 2021a), security (Rabinovich et al., 2015; Xu et al., 2016), sequential decision making (Wu et al., 2022; Gan et al., 2022; Bernasconi et al., 2022), and in incentivized exploration in multi-armed bandits (Kremer et al., 2014; Cohen & Mansour, 2019; Mansour et al., 2016; Sellke & Slivkins, 2021; Mansour et al., 2022).

In the simplest instantiation of the model, there are a sender and a receiver with a common prior over a finite set of states of nature. The sender publicly commits to a *signaling scheme*, which is a randomized mapping from states of nature to signals being sent to the receiver. Then, the sender observes the realized state of nature, and they send a signal to the receiver following the signaling scheme. The receiver observes the signal, computes their posterior distribution over states, and selects an action maximizing their expected utility. The sender and the receiver obtain a payoff which is a function of the receiver's action, and of the realized state of nature. An optimal signaling scheme for the sender is one maximizing their expected utility.

The study of Bayesian persuasion from a computational perspective was initiated by Dughmi & Xu (2016), and the original model was later extended to more complex settings such as games with multiple receivers (see, *e.g.*, (Dughmi & Xu, 2017; Bhaskar et al., 2016; Xu, 2020)). A key question that has emerged is whether computational techniques can be used to ease some of the assumptions made in the original model by Kamenica & Gentzkow (2011). Two main lines of research have emerged from this question: one is aimed at developing robust algorithms that can bypass the common-prior assumption (Zu et al., 2021; Camara et al., 2020), and the other is focused on the robustness of persuasion when the sender is unaware of the receiver's goals (Castiglioni et al., 2020b; 2021b; Babichenko et al., 2021).

This work follows the second perspective, and studies the *online Bayesian persuasion* framework introduced by Castiglioni et al. (2020b; 2023). In this framework, the sender repeatedly faces a receiver whose type is unknown and chosen adversarially at each round from a finite set of possible types. This framework encompasses the problem of learning in repeated Stackelberg games (Letchford et al., 2009; Blum et al., 2014; Marecki et al., 2012; Balcan et al., 2015).

[1] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy [2] Department of Computing Sciences, Università Bocconi, Milan, Italy. Correspondence to: Martino Bernasconi <martino.bernasconideluca@polimi.it>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

Contributions We start by describing a general no-regret algorithm for online learning against an oblivious adversary with a *finite* number of possible loss functions. We use this algorithm to provide a tight $\tilde{O}(T^{1/2})$ regret upper bound in the setting with one receiver and partial feedback, improving over the $\tilde{O}(T^{4/5})$ rate by Castiglioni et al. (2020b). This result also improves the best known bound of $\tilde{O}(T^{2/3})$ for online learning in repeated Stackelberg games provided by Balcan et al. (2015). Then, we show that our general framework can be applied to obtain the first no-regret guarantees under partial feedback in the multi-receiver setting introduced by Castiglioni et al. (2021b). In particular, we provide a tight $\tilde{O}(T^{1/2})$ regret bound under the assumption the set of possible type profiles of the receivers is known beforehand by the sender. In each of these settings, our no-regret algorithms may suffer from exponential per-iteration running time, as expected from known hardness results for the online Bayesian persuasion settings (Castiglioni et al., 2020b). In the last part of the paper, we provide the first no-regret algorithms for online Bayesian persuasion with guaranteed polynomial per-iteration running time. We do that by considering the *type reporting* framework introduced by Castiglioni et al. (2022a), where the sender can commit to a *menu* of signaling schemes, and then let the receivers choose their preferred signaling scheme depending on their private types. In such a setting, we provide a $O(T^{1/2})$ regret upper bound for the single-receiver setting. Moreover, by designing a general procedure based on the follow the regularized leader algorithm, we show that it is possible to achieve the same rate of convergence with polynomial-time per-iteration time complexity also in the multi-receiver setting, when receivers have binary actions and the utility of the sender is specified by a function of receivers' actions that is either supermodular or anonymous.

2. Preliminaries

Vectors are denoted by bold symbols. Given a vector $\boldsymbol{x}$, we let x_i be its i-th component. The set $\{1, 2, \dots, n\}$ of the first n natural numbers is compactly denoted as $[n]$. Moreover, given a discrete set $\mathcal{X}$, we denote by $\Delta_{\mathcal{X}}$ the $|\mathcal{X}|$-simplex, while, given a set $\mathcal{Y}$, $\text{int}(\mathcal{Y})$ is the *interior* of $\mathcal{Y}$.

In the rest of this section, we formally describe the *online Bayesian persuasion* framework introduced by Castiglioni et al. (2020b; 2023). In particular, we focus on a (more general) version of such a framework introduced in a follow-up work by Castiglioni et al. (2021b). This models a repeated interaction between a *sender* and multiple *receivers*.

We denote by $\mathcal{R} := [n]$ a finite set of n receivers. Each receiver $r \in \mathcal{R}$ has a finite set $\mathcal{K}_r$ of m different types, and a finite set $\mathcal{A}_r$ of available *actions*. We let $\mathcal{K} := \bigtimes_{r\in\mathcal{R}} \mathcal{K}_r$ be the set of *type profiles*, *i.e.,* vectors $\boldsymbol{k} \in \mathcal{K}$ defining a type $k_r \in \mathcal{K}_r$ for each receiver $r \in \mathcal{R}$. Similarly, we let $\mathcal{A} := A^n$ be the set of *action profiles* $\boldsymbol{a} \in \mathcal{A}$ specifying an action $a_r \in A$ for each receiver $r \in \mathcal{R}$. Notice that, for ease of notation, we assume that all the receivers have the same action set, *i.e.*, $\mathcal{A}_r = A$ for all $r \in \mathcal{R}$. This comes w.l.o.g. as it is always possible to assign fictitious actions to the receivers whenever the assumption does *not* hold.

The payoffs of both the sender and the receivers depend on a random *state of nature*, which is drawn from a finite set Θ of d possible states according to a commonly-known *prior* probability distribution $\boldsymbol{\mu} \in \text{int}(\Delta_\Theta)$. The sender's payoffs also depend on the actions selected by the receivers, as defined by the function $u^{\mathsf{s}} : \mathcal{A} \times \Theta \to [0, 1]$. Moreover, as it is customary in the literature (see, *e.g.,* (Dughmi & Xu, 2017)), we assume that there are *no inter-agent externalities*, which means that the payoffs of a receiver only depend on the action played by them, and *not* on those played by other receivers. Formally, a receiver $r \in \mathcal{R}$ of type $k \in \mathcal{K}_r$ is characterized by a payoff function $u^r_k : A \times \Theta \to [0, 1]$.

As in the classical Bayesian persuasion framework by Kamenica & Gentzkow (2011), the sender gets to know the realized state of nature $\theta \sim \boldsymbol{\mu}$, and they have the ability to strategically disclose (part of) such information to the receivers, in order to maximize their own utility. This is achieved by committing beforehand to a *signaling scheme*, which is a randomized mapping from states of nature to signals being sent to the receivers. Formally, let $\mathcal{S} := \bigtimes_{r\in\mathcal{R}} \mathcal{S}_r$ be the finite set of *signal profiles*, *i.e.,* the set of vectors $\boldsymbol{s} \in \mathcal{S}$ defining a signal $s_r \in \mathcal{S}_r$ for each receiver $r \in \mathcal{R}$.[1] Then, a signaling scheme is a mapping $\phi : \Theta \to \Delta_{\mathcal{S}}$. We denote by $\phi_\theta(\boldsymbol{s})$ the probability of sending the signals in $\boldsymbol{s} \in \mathcal{S}$ when the state of nature is $\theta \in \Theta$. Moreover, given a signaling scheme ϕ, we define the resulting *marginal* signaling scheme for a receiver $r \in \mathcal{R}$ as $\phi^r : \Theta \to \Delta_{\mathcal{S}_r}$. Formally, for every $\theta \in \Theta$, the marginal signaling scheme ϕ^r defines the distribution over receiver r's signals that is induced by ϕ, which assigns probability

$$\phi^r_\theta(s') := \sum_{\boldsymbol{s}\in\mathcal{S}: s_r = s'} \phi_\theta(\boldsymbol{s}) \text{ to each } s' \in \mathcal{S}_r. \qquad (1)$$

The repeated interaction between the sender and the receivers goes on as follows. At each round $t \in [T]$, the sender commits to a signaling scheme ϕ_t (*i.e.,* ϕ_t is publicly known), and, subsequently, they observe the realized state of nature $\theta \sim \boldsymbol{\mu}$. Then, the sender draws a signal profile $\boldsymbol{s} \sim \phi_{t,\theta}$ and communicates to each receiver $r \in \mathcal{R}$ (whose type is unknown to the sender) their own private signal s_r. After observing the signal, each receiver $r \in \mathcal{R}$ updates their prior belief $\boldsymbol{\mu}$ according to Bayes rule, and, then, they select an action maximizing their expected utility.

The *posterior* $\boldsymbol{\xi}^{s_r} \in \Delta_\Theta$ computed by a receiver $r \in \mathcal{R}$

[1] In this work, we focus on *private* signaling, where the sender has the ability to privately communicate a signal to each receiver.

after observing a signal $s_r \in \mathcal{S}_r$ under signaling scheme ϕ is a probability distribution over states such that

$$\xi_\theta^{s_r} := \frac{\mu_\theta \, \phi_\theta^r(s_r)}{\sum_{\theta' \in \Theta} \mu_{\theta'} \phi_{\theta'}^r(s_r)} \qquad \text{for every } \theta \in \Theta.[2]$$

Given a posterior $\boldsymbol{\xi} \in \Delta_\Theta$, the set of *best-response actions* of a receiver $r \in \mathcal{R}$ of type $k \in \mathcal{K}_r$ is defined as follows:

$$\mathcal{B}_{\boldsymbol{\xi}}^{r,k} := \arg\max_{a \in A} \sum_{\theta \in \Theta} \xi_\theta \, u_k^r(a, \theta).$$

Moreover, assuming receivers break ties in favor of the sender, the sender's expected utility for selecting a signaling scheme ϕ given a receivers' type profile $\boldsymbol{k} \in \mathcal{K}$ is

$$u^{\mathsf{s}}(\phi, \boldsymbol{k}) := \sum_{\boldsymbol{s} \in \mathcal{S}} \left(\arg\max_{\boldsymbol{a} \in \bigtimes_{r \in \mathcal{R}} \mathcal{B}_{\boldsymbol{\xi}^{s_r}}^{r,k_r}} \sum_{\theta \in \Theta} \mu_\theta \phi_\theta(\boldsymbol{s}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \right).$$

We focus on the problem of computing a sequence $\{\phi_t\}_{t \in [T]}$ of signaling schemes which can be employed by the sender so as to maximize their utility. We assume that the sequence of receivers' type profiles $\{\boldsymbol{k}_t\}_{t \in [T]}$, with $\boldsymbol{k}_t \in \mathcal{K}$, is selected by an oblivious adversary. At each round $t \in [T]$ of the repeated interaction, the sender gets a payoff $u^{\mathsf{s}}(\phi_t, \boldsymbol{k}_t)$ and receives some feedbacks about receivers' types. In the *full feedback* setting, the sender gets to know the receivers' type profile $\boldsymbol{k}_t$, while in the *partial feedback* setting the sender only observes the action profile $\boldsymbol{a}_t \in \mathcal{A}$ played by the receivers at round t. In the partial feedback setting, as in the original online Bayesian persuasion model by Castiglioni et al. (2020b), we assume that the sender observes receivers' actions at each round. However, our algorithms only need that the sender observes the utility $u^{\mathsf{s}}(\phi_t, \boldsymbol{k}_t)$ achieved at round t, which is slightly more general.

We measure the performance of the sender by using the regret up to round T with respect to the best fixed signaling scheme in hindsight:

$$R_T := \max_\phi \sum_{t=1}^T u^{\mathsf{s}}(\phi, \boldsymbol{k}_t) - \sum_{t=1}^T \mathbb{E}[u^{\mathsf{s}}(\phi_t, \boldsymbol{k}_t)],$$

where the expectation is on the possible randomness of the algorithm.[3] Ideally, we would like an algorithm that generates a sequence $\{\phi_t\}_{t \in [T]}$ with the following properties: (i) the regret is polynomial in the size of the problem instance, *i.e.,* it is $\text{poly}(n, m, d, |A|)$, and goes to zero as $T \to \infty$; and (ii) the per-round running time is $\text{poly}(t, n, m, d, |A|)$.

3. Online Learning Against Adversaries with a Finite Number of Losses

We start by introducing a general framework that will be crucial in proving some of our main results in the rest of the paper. In particular, we propose a no-regret algorithm for a general online learning problem in which the agent's decisions are only evaluated in terms of D possible adversarially-selected loss functions. The algorithm that we propose attains a $\tilde{O}(\sqrt{T})$ regret bound, which is independent of the size of the decision space of the agent and only depends polynomially on the number of possible losses D.

In the online learning problem that we consider in this section, at each round $t \in [T]$, an agent takes a decision $\boldsymbol{x}_t$ from a set $\mathcal{X} \subseteq \mathbb{R}^M$, and, then, an adversary selects an element d_t from a finite set $\mathcal{D}$ of $D := |\mathcal{D}|$ elements. Then, the loss suffered by the agent is $L_{d_t}(\boldsymbol{x}_t)$, where functions $L_d : \mathcal{X} \to [0, 1]$ are loss functions indexed by the elements $d \in \mathcal{D}$. Thus, the performance of the agent over the T rounds is evaluated by means of the regret:

$$R_T := \sum_{t=1}^T \mathbb{E}[L_{d_t}(\boldsymbol{x}_t)] - \min_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^T L_{d_t}(\boldsymbol{x}),$$

where the expectation is with respect to the (possible) randomization that the agent adopts in choosing $\boldsymbol{x}_t$.

Next, we introduce a general no-regret algorithm that works by exploiting the linear structure of the online learning problem described above. In order to do so, we introduce a vector-valued function $\boldsymbol{\nu} : \mathcal{X} \to \mathbb{R}^D$ defined as $\boldsymbol{\nu}(\boldsymbol{x}) := [L_d(\boldsymbol{x})]_{d \in \mathcal{D}}$ for all $\boldsymbol{x} \in \mathcal{X}$. By observing that $L_d(\boldsymbol{x}) = \boldsymbol{\nu}(\boldsymbol{x})^\top \mathbf{1}_d$, where $\mathbf{1}_d \in \{0,1\}^D$ is a vector whose d-th component is the only one that is different from zero, we can cast the online learning problem as a new one with linear losses defined over the decision space $\boldsymbol{\nu}(\mathcal{X})$. Since $\boldsymbol{\nu}(\mathcal{X})$ may *not* be convex, the algorithm employs a regret minimizer $\mathfrak{R}$ working on the convex hull $\operatorname{co} \boldsymbol{\nu}(\mathcal{X})$.[4] This is possible since, instead of playing a $\boldsymbol{z} \in \operatorname{co} \boldsymbol{\nu}(\mathcal{X})$, the algorithm can replace it by a suitable randomization of $D + 1$ points in $\boldsymbol{\nu}(\mathcal{X})$, which is guaranteed to exist by the Carathéodory's theorem. See Algorithm 1 for the detailed procedure, where we denote by $\boldsymbol{\nu}^\dagger$ the inverse map of $\boldsymbol{\nu}$. Notice that, provided that a suitable regret minimizer $\mathfrak{R}$ is instantiated, the algorithm works both in the full feedback setting, where the agent observes d_t, and in the bandit feedback one, in which they only observe $L_{d_t}(\boldsymbol{x}_t)$.

The following theorem bounds the regret of Algorithm 1:

Theorem 3.1. *Algorithm 1 guarantees a cumulative regret* $R_T \leq R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X}))$*, where* $R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X}))$ *is the regret*

[2] For ease of notation, we omit the dependence of the posterior distribution $\boldsymbol{\xi}^{s_r}$ on the signaling scheme ϕ and the receiver r, as these will be clear from context.

[3] This notion of regret is also known as *Stackelberg regret* (Balcan et al., 2015; Chen et al., 2020).

[4] To see that the convex hull is necessary, let $\mathcal{X}$ be the unit sphere in $\mathbb{R}^2$ and $L_d(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^{2d}$ for $d \in \mathcal{D} = \{0.5, 1\}$. It is easy to verify that $\boldsymbol{\nu}(\mathcal{X}) = \{(x, \sqrt{x}) : x \in [0, 1]\}$, which is *not* convex.

```
Algorithm 1 No-regret Algorithm
Require: Regret minimizer $\mathfrak{R}$ for the set $\operatorname{co}\boldsymbol{\nu}(\mathcal{X})$ and linear losses; Inverse mapping $\boldsymbol{\nu}^\dagger$
Initialize regret minimizer $\mathfrak{R}$
for $t = 1, \ldots, T$ do
    $\operatorname{co}\boldsymbol{\nu}(\mathcal{X}) \ni \boldsymbol{z}_t \leftarrow \mathfrak{R}.\textsc{Reccomend}()$
    $\left\{(\boldsymbol{z}_t^i, \lambda_t^i)\right\}_{i\in[D+1]} \leftarrow \textsc{Carathéodory}(\boldsymbol{z}_t, \boldsymbol{\nu}(\mathcal{X}))$
    Draw $j \in [D+1]$ with probabilities $\lambda_t^1, \ldots, \lambda_t^{D+1}$
    Play $\boldsymbol{x}_t \leftarrow \boldsymbol{\nu}^\dagger(\boldsymbol{z}_t^j)$
    Observe $d_t \in \mathcal{D}$                              ▷ Full feedback
    Observe $L_{d_t}(\boldsymbol{x}_t)$                        ▷ Bandit feedback
    $\mathfrak{R}.\textsc{ObserveLoss}(L_{d_t})$                ▷ Full feedback
    $\mathfrak{R}.\textsc{ObserveLoss}(L_{d_t}(\boldsymbol{x}_t))$      ▷ Bandit feedback
end for
```

bound of a suitable regret minimizer $\mathfrak{R}$ for the set $\operatorname{co}\boldsymbol{\nu}(\mathcal{X})$.

In order to run Algorithm 1, one has to implement the Carathéodory oracle and the inverse map $\boldsymbol{\nu}^\dagger$. The following result shows that these can be implemented in "linear problems". These include as special cases many interesting settings, such as most of the online Bayesian persuasion problems studied in this paper.

Theorem 3.2. *If $\mathcal{X}$ is a polytope and $\boldsymbol{\nu}$ is a linear map, i.e., there exists $\mathbf{M} \in \mathbb{R}^{D\times M}$ such that $\nu(\boldsymbol{x}) = \mathbf{M}\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{X}$, then there exist algorithms implementing the* Carathéodory *oracle and the inverse map $\boldsymbol{\nu}^\dagger$.*

Moreover, in the case of "linear problems" as in Theorem 3.2, we can instantiate Algorithm 1 with specific regret minimizes $\mathfrak{R}$ for both the full and the bandit feedback scenarios, so as to obtain the following guarantees.

Corollary 3.3. *Under the assumptions of Theorem 3.2, with full feedback, there exists a regret minimizer $\mathfrak{R}$ such that Algorithm 1 guarantees cumulative regret*

$$R_T \leq \sqrt{DT}.$$

Corollary 3.4. *Under the assumptions of Theorem 3.2, with bandit feedback, there exists a regret minimizer $\mathfrak{R}$ such that Algorithm 1 guarantees cumulative regret*

$$R_T \leq 16D^{3/2}\sqrt{T\log T}.$$

4. Optimal Regret Bounds for Online Bayesian Persuasion with Partial Feedback

Next, we show that our general online learning framework introduced in Section 3 can be applied to the setting of online Bayesian persuasion with partial feedback, enabling the derivation of novel state-of-the-art results.

A standard revelation-principle-style argument shows that we can focus w.l.o.g. on signaling schemes that are direct and persuasive (see, *e.g.,* (Arieli & Babichenko, 2019)). In particular, a signaling scheme is *direct* if signals correspond to action recommendations. Formally, the set of signals of a receiver $r \in \mathcal{R}$ is $\mathcal{S}_r = A^m$, with each signal defining an action recommendation for each possible receiver r's type. Moreover, a direct signaling scheme is *persuasive* if each receiver's type is incentivized to follow the action recommendations issued by the sender. Formally, the set of direct and persuasive signaling schemes $\mathcal{P}$ is the set of all $\phi : \Theta \to \Delta_{A^{mn}}$ such that, for every receiver $r \in \mathcal{R}$, receiver r's type $k \in \mathcal{K}_r$, and action $a \in A$, it holds

$$\sum_{\theta\in\Theta}\sum_{\boldsymbol{a}\in A^{mn}} \mu_\theta \phi_\theta^r(\boldsymbol{a})(u_k^r(a_k^r,\theta) - u_k^r(a,\theta)) \geq 0, \tag{2}$$

where, by slightly abusing notation, we denote as A^{mn} the set $\mathcal{S}$ with direct signals, while, given $\boldsymbol{a} \in A^{mn}$, we let a_k^r be the action in $\boldsymbol{a}$ corresponding to type $k \in \mathcal{K}_r$ of receiver $r \in \mathcal{R}$. Intuitively, the inequality requires that, for a receiver r of type k, the utility obtained by following recommendations given by ϕ is greater than or equal to that achieved by deviating to any another action a. Notice that the set $\mathcal{P}$ can be encoded as a polytope, by adding to the persuasiveness constraints those ensuring that ϕ is well defined, namely $\sum_{\boldsymbol{a}\in A^{mn}} \phi_\theta^k(\boldsymbol{a}) = 1$ for all $\theta \in \Theta$.

Given any direct and persuasive signaling scheme $\phi \in \mathcal{P}$, the sender's utility under type profile $\boldsymbol{k} \in \mathcal{K}$ is

$$u^{\mathsf{s}}(\phi, \boldsymbol{k}) := \sum_{\theta\in\Theta}\sum_{\boldsymbol{a}\in A^{mn}} \mu_\theta \phi_\theta(\boldsymbol{a}) u^{\mathsf{s}}((a_{k_1}^1, \ldots, a_{k_n}^n), \theta),$$

where we remark that $a_{k_r}^r$ is the action recommendation specified by $\boldsymbol{a}$ for a receiver r whose realized type is k_r. Moreover, let us observe that $u^{\mathsf{s}}(\phi, \boldsymbol{k})$ is a linear function in the signaling scheme ϕ.

As it is well known, finding an optimal direct and persuasive signaling scheme is NP-hard, even when there is only one receiver and the distribution over receiver's types is known (Castiglioni et al., 2020b, Theorem 2). This implies that the polytope $\mathcal{P}$ has exponential size, since the sender's utility can be represented as a linear function of direct and persuasive signaling schemes. Moreover, classical reductions from offline to online optimization problems also show that there cannot be an efficient (*i.e.*, with polynomial per-iteration running time) algorithm that achieves no-regret in this setting (Roughgarden & Wang, 2019; Castiglioni et al., 2020b; Daskalakis & Syrgkanis, 2022).

A natural question is whether it is possible to design no-regret algorithm by relaxing the efficiency requirement on the per-iteration running time. This question has already been answered affirmatively by Castiglioni et al. (2020b) in single-receiver settings. In the following, we show that our online learning framework allows us to improve the regret bound in (Castiglioni et al., 2020b) to optimality, by matching known lower bounds, and, additionally, it also allows us to extend the result to multi-receiver settings.

4.1. Single-Receiver Setting under Partial Feedback

Next, we consider the case in which there is a single receiver, meaning that $n = 1$.[5] In such a setting, the sender can observe a different loss for each of the m different receiver's types. Formally, the map $\boldsymbol{\nu} : \mathcal{P} \to \mathbb{R}^m$ is defined by letting, for every signaling scheme $\phi \in \mathcal{P}$:

$$\boldsymbol{\nu}(\phi) := [-u^{\mathsf{s}}(\phi, k)]_{k\in\mathcal{K}}.$$

Then, we can apply Corollary 3.4 to obtain the following regret upper bound under partial feedback.

Theorem 4.1. *The single-receiver online Bayesian persuasion problem under partial feedback admits an algorithm which guarantees the following regret bound*

$$R_T = O(m^{2/3}\sqrt{T \log T}).$$

This result improves over the best known upper bound for the partial feedback case, which is of the order of $\tilde{O}(T^{4/5})$ and it is derived in the original paper introducing online Bayesian persuasion (Castiglioni et al., 2020b, Theorem 4).

4.2. Multi-Receiver Setting under Partial Feedback

Castiglioni et al. (2021b) introduce the online Bayesian persuasion problem with multiple receivers and adversarially-selected types. They provide an algorithm that, under full feedback and some technical assumptions, guarantees sublinear regret. In particular, their regret bound depends polynomially in the size of the problem instance when assuming that the number of possible receivers' type profiles is fixed. This is a reasonable assumption given that the total number of type profiles is $|\mathcal{K}| = m^n$, which is exponential in the number of receivers n. Under the same assumption, we provide the first no-regret algorithm under partial feedback.

Formally, we let $\overline{\mathcal{K}} \subseteq \mathcal{K}$ be the set of possible type profiles, so that, at each round $t \in [T]$, the receivers' type profile $\boldsymbol{k}_t$ belongs to $\overline{\mathcal{K}}$. We provide regret bounds which depend polynomially on the number of possible type profiles $|\overline{\mathcal{K}}|$. However, differently from Castiglioni et al. (2021b), in our algorithm working with partial feedback we assume that the set $\overline{\mathcal{K}}$ is known beforehand. Indeed, an "on the fly" construction of $\overline{\mathcal{K}}$ as in Castiglioni et al. (2021b) seems unfeasible under partial feedback, where, by definition, the sender does *not* observe $\boldsymbol{k}_t$.

For every type profile $\boldsymbol{k} \in \overline{\mathcal{K}}$, the sender gets utility $u^{\mathsf{s}}(\phi, \boldsymbol{k})$ by playing a signaling scheme ϕ. Then, we can define the map $\boldsymbol{\nu} : \mathcal{P} \to \mathbb{R}^{|\overline{\mathcal{K}}|}$ so that, for every signaling scheme $\phi \in \mathcal{P}$, it holds $\boldsymbol{\nu}(\phi) := [-u^{\mathsf{s}}(\phi, \boldsymbol{k})]_{\boldsymbol{k}\in\overline{\mathcal{K}}}$. Notice that $\boldsymbol{\nu}$ is a linear map from $\mathcal{P}$ to $\mathbb{R}^{|\overline{\mathcal{K}}|}$. Thus, by Corollary 3.4, Algorithm 1 gives the following regret bound.

Theorem 4.2. *The multi-receiver online Bayesian persuasion problem under partial feedback admits an algorithm which guarantees the following regret bound*

$$R_T = O\left(\left|\overline{\mathcal{K}}\right|^{2/3}\sqrt{T \log T}\right).$$

5. Polynomial-Time Per-Iteration Running Time through Type Reporting

In this section, we show that it is possible to circumvent the negative results which rule out the existence of a no-regret algorithm for online Bayesian persuasion with polynomial per-iteration running time. We do that by enriching the decision space of the sender. In particular, we consider the framework of Bayesian persuasion *with type reporting* introduced by Castiglioni et al. (2022a) for offline settings, where the sender has the ability to commit to a *menu* of signaling schemes, and then let the receivers choose their preferred signaling scheme depending on their private types.

5.1. Online Type Reporting

In the type-reporting model, at each round $t \in [T]$ of the repeated interaction, the sender proposes a *menu* of marginal signaling schemes to each receiver. We collectively denote them by $\varphi_t := \{\varphi_t^{r,k}\}_{r\in\mathcal{R},k\in\mathcal{K}_r}$, so that the menu proposed to receiver $r \in \mathcal{R}$ consists of a set of marginal distributions $\varphi_t^{r,k} : \Theta \to \Delta_{\mathcal{S}_r}$, one for each receiver's type $k \in \mathcal{K}_r$. Then, each receiver $r \in \mathcal{R}$ reports a type $k_r \in \mathcal{K}_r$ (possibly different from their true type) to the sender. The reported type k_r is such that the signaling scheme φ_t^{r,k_r} is the one guaranteeing to the receiver the highest expected utility among those in the menu.[6] Finally, the sender computes and commits to the signaling scheme $\phi_t : \Theta \to \Delta_{\mathcal{S}}$ which maximizes the sender's expected utility among the signaling schemes whose marginals are equal to the marginal signaling schemes φ_t^{r,k_r} corresponding to the types k_r reported by the receivers, *i.e.*, $\phi_t^r = \varphi^{r,k_r}$ for every $r \in \mathcal{R}$. From this point on, the interaction goes on as in the case without type reporting.

Notice that, in the type-reporting setting, the sender observes the types of the receivers at each round $t \in [T]$. Thus, in the type-reporting model, the sender always has full feedback.

Let us also remark that the assumption that the sender can only propose marginal signaling schemes to the receivers

[5] In the single-receiver setting, we omit the dependence on r from sets and other elements.

[6] Such a step can be equivalently implemented by extending the interaction between the sender and the receiver: the sender can ask each receiver $r \in \mathcal{R}$ to directly select a marginal signaling scheme $\varphi_t^{r,k}$ from the menu, and the receiver will be incentivized to select the one corresponding to their own true type.

is w.l.o.g., since the expected utility of each receiver only depends on their marginal signaling scheme, and *not* on those of the others (see Section 2). Therefore, the sender can delay the choice of the joint signaling scheme ϕ_t until after all the receivers reported their types.

By a revelation-principle-style argument (Castiglioni et al., 2022a), it is always possible to focus w.l.o.g. on *incentive compatible* (IC) menus $\varphi = \{\varphi^{r,k}\}_{r \in \mathcal{R}, k \in \mathcal{K}_r}$, which are those such that each receiver $r \in \mathcal{R}$ is incentivized to report their true type, say $k_r \in \mathcal{K}_r$. Formally, for all $k \neq k_r \in \mathcal{K}_r$,

$$\sum_{s_r \in \mathcal{S}_r} \max_{a \in A} \sum_{\theta \in \Theta} \mu_\theta \, \varphi_\theta^{r,k_r}(s_r) \, u_{k_r}^r(a, \theta) \geq \sum_{s_r \in \mathcal{S}_r} \max_{a \in A} \sum_{\theta \in \Theta} \mu_\theta \, \varphi_\theta^{r,k}(s_r) \, u_{k_r}^r(a, \theta), \quad (3)$$

where the $\max$ operators account for the fact that the receiver plays a best-response action after receiving a signal.

W.l.o.g., we can focus on menus that are *direct*, namely $\mathcal{S}_r = A$ for every $r \in \mathcal{R}$, and *persuasive*. We say that a direct menu $\varphi = \{\varphi^{r,k}\}_{r \in \mathcal{R}, k \in \mathcal{K}_r}$ is persuasive if the marginal signaling schemes $\varphi^{r,k}$ satisfy persuasiveness constraints similar to those of Equation (2) for every receiver $r \in \mathcal{R}$ and type $k \in \mathcal{K}_r$. Then, we define Λ as the set of menus which are IC, direct, and persuasive.

The sender's goal is to compute a sequence of IC menus $\{\varphi_t\}_{t \in [T]}$ and a sequence of signaling schemes $\{\phi_t\}_{t \in [T]}$ which are consistent with the menus, whose performance over the T rounds is measured in terms of the following notion of regret:

$$R_T := \max_{\varphi} \sum_{t=1}^{T} u^{\mathsf{s}}(\varphi, \boldsymbol{k}_t) - \sum_{t=1}^{T} \mathbb{E}\left[u^{\mathsf{s}}(\phi_t, \boldsymbol{k}_t)\right],$$

where, by overloading notation, we denoted with

$$u^{\mathsf{s}}(\varphi, \boldsymbol{k}) := \max_{\phi : \phi^r = \varphi^{r,k_r}} u^{\mathsf{s}}(\phi, \boldsymbol{k}) \quad (4)$$

the maximum utility of the sender when the receivers' type profile is $\boldsymbol{k} \in \mathcal{K}$. We remark that the above formulation of regret is stronger than the classical one in which a best-in-hindsight decision is fixed for all the rounds. Indeed, although the best menu φ is fixed for all $t \in [T]$, we allow the signaling scheme $\phi_t^\star \in \arg\max_{\phi : \phi^r = \varphi^{r,k_{t,r}}} u^{\mathsf{s}}(\phi, \boldsymbol{k}_t)$ to depend on the round t, as long as $\phi_t^\star$ has fixed marginals that are compatible with the best menu φ.

5.2. Single-Receiver Setting with Type Reporting

We start by studying the single-receiver setting (*i.e.,* $n = 1$).

In the type-reporting setting, it is not clear whether there exists a succinct representation of the polytope of persuasive menus or not. The reason for this is that encoding the inner maximizations of Equation (3) as a set of linear inequalities would require exponentially-many constraints. However, this does *not* rule out the existence of a succinct representation. Indeed, even if Λ has an exponential description, it is possible to show that it has polynomial *extension complexity* (Fiorini et al., 2012). In particular, we can show that there exists a succinct representation of Λ in a suitable higher dimensional space. This was already implicitly shown by Castiglioni et al. (2022b). Here, we provide a formal characterization for completeness.

Intuitively, the construction works as follows: we introduce extra variables l, called extension variables such that the extended polytope $\mathcal{L}$ is defined by variables $\ell \equiv (\varphi, l)$, where $\varphi_\theta^k \in \mathbb{R}_+^{|\mathcal{A}|}$ for each $\theta \in \Theta, k \in \mathcal{K}$ encode marginal signaling schemes, and we have one additional variable $l_a^{k,k'} \in \mathbb{R}$ for each $a \in A, k, k' \in \mathcal{K}$. The polytope $\mathcal{L}$ can be described by a polynomial number of constraints. This fact, together with the linear projection map $\pi : \mathcal{L} \to \Lambda$ defined as $\pi(\varphi, l) = \varphi$, proves the polynomial extension complexity of Λ. Formally, the extended polytope $\mathcal{L}$ can be described by the following inequalities:

$$\sum_{\theta \in \Theta} \sum_{a \in A} \mu_\theta \varphi_\theta^k(a) u_k(a, \theta) \geq \sum_{a \in A} l_a^{k,k'} \quad \forall k, k' \in \mathcal{K}_r \quad (5a)$$

$$l_a^{k,k'} \geq \sum_{\theta \in \Theta} \mu_\theta \varphi_\theta^{k'}(a) u_k(a', \theta) \quad \forall k, k' \in \mathcal{K}_r, a, a' \in A \quad (5b)$$

$$\sum_{a \in A} \varphi_\theta^k(a) = 1 \quad \forall k \in \mathcal{K}_r, \theta \in \Theta, \quad (5c)$$

where $l_a^{k,k'}$ represents the maximum utility obtained by a receiver of type k who reports type k', when type k' is recommended action a.

Then, we instantiate Algorithm 1 by taking the set $\mathcal{L}$ as the polytope $\mathcal{X}$, where we have one loss for each of the m types that can be reported by the receiver. We define $\boldsymbol{\nu} : \mathcal{L} \to \mathbb{R}^m$ as the vector-valued map associating each feasible point $\ell = (\varphi, l)$ with the m-dimensional vector of losses $\boldsymbol{\nu}(\ell) := [-u^{\mathsf{s}}(\varphi^k, k)]_{k \in \mathcal{K}}$, where the value of a menu φ for the sender against a receiver's type $k \in \mathcal{K}$ is $u^{\mathsf{s}}(\varphi^k, k)$ as the overall signaling scheme ϕ coincides with the signaling scheme φ^k, when there is a single receiver. Then, Corollary 3.3 yields the following result.

Theorem 5.1. *The single-receiver online Bayesian persuasion problem with type reporting admits an algorithm which guarantees regret $R_T \leq \sqrt{mT}$ and polynomial per-iteration running time.*

5.3. Multi-Receiver Setting with Type Reporting

In this section, we focus on the problem of designing a no-regret algorithm for the multi-receiver setting with type reporting. The method employed in the case of a single receiver is not applicable here, as the number of possible type profiles becomes exponentially large, resulting in exponentially-many possible loss functions. Moreover, it is not possible to directly design efficient algorithms working on the joint action space, since it has exponential size. In order to build a no-regret algorithm for this setting, the idea is to cast the learning problem into a decision space which is small enough to be manageable. In particular, we observe that the sender must commit only to the marginal signaling schemes $\{\varphi_t^{r,k_r}\}_{r\in\mathcal{R},k_r\in\mathcal{K}_r}$ before observing the receivers' types. Then, at each round t, the sender observed the type $k_{r,t}$ reported by each receiver $r \in \mathcal{R}$, and solves an offline optimization problem to compute an optimal joint signaling schemes ϕ_t whose marginal signaling schemes are $\{\varphi_t^{r,k_{r,t}}\}_{r\in\mathcal{R}}$. By exploiting this observation, we develop a no-regret algorithm that operates within the smaller decision space of marginal signaling schemes.

Let Λ_r be the set of IC, direct, and persuasive menus of marginal signaling schemes for receiver $r \in \mathcal{R}$. Formally, Λ_r is defined as the set of $\varphi^{r,k}$ that satisfy the constraints in Equations (5a)–(5c) for every $r \in \mathcal{R}$ and type $k \in \mathcal{K}_r$. Moreover, let $\Lambda := \bigtimes_{r\in\mathcal{R}} \Lambda_r$. Intuitively, an element of Λ includes a menu of marginal signaling schemes φ^r for each receiver $r \in \mathcal{R}$. Then, the action space of the learner is given by the set of IC and persuasive marginal signaling schemes Λ. The sender's utility when the agents are of type $\boldsymbol{k} \in \mathcal{K}$ is defined by a function $g^{\boldsymbol{k}} : \Lambda \to [0,1]$, where $g^{\boldsymbol{k}}(\varphi)$ is the value obtained by the following linear program which is an expansion of the maximization in Equation (4):

$$\max_{\phi\geq 0} \quad \sum_{\theta\in\Theta}\sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi_\theta(\boldsymbol{a}) u^{\mathsf{s}}_\theta(\boldsymbol{a}) \quad \text{s.t.} \tag{6a}$$

$$\sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_i=a'}} \phi_\theta(\boldsymbol{a}) = \varphi_\theta^{r,k_r}(a') \qquad \forall r\in\mathcal{R}, a'\in\mathcal{A}_r, \theta\in\Theta, \tag{6b}$$

where Equation (6a) is the utility of a signaling scheme ϕ and Equation (6b) encodes the constraints on the signaling scheme ϕ to have marginals $\{\varphi^{r,k_r}\}_{r\in\mathcal{R}}$. The function $g^{\boldsymbol{k}}(\varphi)$ is the solution to a parametric (in φ) linear program. If we want to solve an online problem involving $g^{\boldsymbol{k}}$, we first have to show that the offline problem $\max_{\varphi\in\Lambda} g^{\boldsymbol{k}}(\varphi)$ is in some sense computationally tractable. More precisely, we show that for any $\boldsymbol{k} \in \mathcal{K}$ the function $g^{\boldsymbol{k}}$ is concave.

Lemma 5.1. *The function $g^{\boldsymbol{k}}(\varphi)$ is concave in φ on Λ for each type profile $\boldsymbol{k} \in \mathcal{K}$.*

Moreover, we show that the function is particularly well behaved. In particular, we prove that it is Lipschitz-continuous with respect to the ℓ_2 norm. This will be useful to upper-bound the norm of gradients of the function $g^{\boldsymbol{k}}$.

Algorithm 2 No-Regret Algorithm Type-Reporting

```
Require: Any set of marginal signaling schemes $\varphi_1 \in \Lambda$,
    Learning rate $\alpha$
1: for $t = 1$ to $T$ do
2:     Propose the set of marginal signaling schemes $\varphi_t$
3:     Observe the receivers reported types $\boldsymbol{k}_t$
4:     $\phi_t \leftarrow$ a solution to LP (6) for $\varphi_t$ with value $g^{\boldsymbol{k}_t}(\varphi_t)$
5:     $\varphi_{t+1} \leftarrow \arg\max_{\varphi\in\Lambda} \sum_{\tau\leq t} g^{\boldsymbol{k}_\tau}(\varphi) - \frac{1}{2\alpha}\|\varphi\|_2^2$
6: end for
```

Lemma 5.2. *For each $\boldsymbol{k} \in \mathcal{K}$, the function $g^{\boldsymbol{k}}(\varphi)$ is $\sqrt{nd|A|}$-Lipschitz-continuous in φ with respect to $\|\cdot\|_2$.*

Since we have no access to the gradient of the functions $g^{\boldsymbol{k}}$, a natural choice to implement a no-regret algorithm is to apply *follow the regularized leader* (FTRL) (Abernethy et al., 2008; Hazan & Kale, 2010). Algorithm 2 describes the specific implementation of the FTRL-type algorithm. At each iteration the algorithm proposes a set of IC menus of marginal signaling schemes $\varphi_t \in \Lambda$. Then, the algorithm observes the reported types $\boldsymbol{k}_t$ (notice that the receivers report their true types since the menu is IC). The algorithm computes a signaling scheme ϕ solving LP (6) for the types $\boldsymbol{k}_t$, returning a signaling scheme with value $g^{\boldsymbol{k}}(\varphi_t)$. Finally, the algorithm updates the set of menus of signaling schemes by computing:

$$\varphi_{t+1} = \arg\max_{\varphi\in\Lambda} \sum_{\tau\in[t]} g^{\boldsymbol{k}_\tau}(\varphi) - \frac{1}{2\alpha}\|\varphi\|_2^2. \tag{7}$$

Following the standard FTRL analysis we can provide an upper bound on the regret for Algorithm 2.

Theorem 5.2. *Let $\alpha := \sqrt{m/T}$. Algorithm 2 guarantees a cumulative regret $R_T \leq nd|A|\sqrt{mT}$.*

5.4. An efficient Implementation for Multi-Receiver Online Bayesian Persuasion with Type Reporting

In the previous section, we provided a no-regret algorithm for the multi-receiver problem. However, we did not address the question of whether Algorithm 2 can be implemented efficiently. Specifically, determining ϕ_t and φ_{t+1} (Line 4 and 5, respectively) is not straightforward. In general, the sender's utility function cannot be represented in space polynomial in the number of players. For this reason, computational works on multi-receiver Bayesian persuasion focus on succinctly representable utility functions (see, *e.g.,* (Dughmi, 2017; Babichenko & Barman, 2017; Castiglioni et al., 2021b)). In particular, each receiver's action set A is binary, and the two actions are denoted by a_1 and a_0. Then, the sender's utility function can be compactly represented as a collection of functions $f^{\mathsf{s}}_\theta : 2^{\mathcal{R}} \to [0,1]$, where $f^{\mathsf{s}}_\theta(R)$ denotes the sender's utility when the state of nature is $\theta \in \Theta$ and $R \subseteq \mathcal{R}$ is the set of receivers playing a_1. The literature usually examines three common types of utility functions: *supermodular*, *submodular*, and *anonymous*. For the case of submodular functions, it is well known that even in the offline setting without types, the problem is NP-hard to approximate up to within any factor better than $(1 - 1/e)$ (Babichenko & Barman, 2017). Therefore, in this section, we show that Algorithm 2 can be implemented efficiently when the sender's utility function is monotone supermodular or monotone anonymous.

Definition 5.3. The function f^{s}_θ is *supermodular* if, for every $R, R' \subseteq \mathcal{R}$, it holds

$$f^{\mathsf{s}}_\theta(R \cap R') + f^{\mathsf{s}}_\theta(R \cup R') \geq f^{\mathsf{s}}_\theta(R) + f^{\mathsf{s}}_\theta(R').$$

Finally, the function f^{s}_θ is *anonymous* if $f^{\mathsf{s}}_\theta(R) = f^{\mathsf{s}}_\theta(R')$ for all $R, R' \subseteq \mathcal{R}$ such that $|R| = |R'|$.

We show that we can efficiently solve LP (6) and the concave program of Equation (7) (which both have an exponential number of variables, but polynomially many constraints) by writing their dual formulation, and then using the ellipsoid method with a suitable efficient separation oracle.

As a separation oracle, we use the following general optimization oracle.

Definition 5.4 (Optimization Oracle)**.** Given as inputs a function f^{s}_θ and a vector of weights $w \in \mathbb{R}^n$, with w_r denoting the component corresponding to receiver r, an *optimization oracle* $\mathcal{O}$ returns a subset of receivers such that

$$\mathcal{O}(f^{\mathsf{s}}_\theta, w) \in \arg\max_{R \subseteq \mathcal{R}} \left\{ f^{\mathsf{s}}_\theta(R) + \sum_{r \in R} w_r \right\}.$$

Moreover, will will use the following known result.

Lemma 5.3 (Babichenko & Barman (2017) and Dughmi & Xu (2017))**.** *The optimization oracle $\mathcal{O}(f^{\mathsf{s}}_\theta, w)$ can be implemented in polynomial time when f^{s}_θ is a supermodular or anonymous monotone utility function.*

In the following, we show that when we have access to the separation oracle $\mathcal{O}$, both the optimization problems in Line 4 and Line 5 can be solved in polynomial time using the ellipsoid method. We start by providing a polynomial-time algorithm for LP (6). Intuitively, the problem is equivalent to that of finding an optimal signaling scheme in a problem with fixed marginal signaling schemes. In particular, by rewriting LP (6) for the specific case of a binary action space and by taking its dual, we obtain

$$\min_x \sum_{r \in \mathcal{R}, \theta \in \Theta} \varphi^{r,k_r}_\theta(a_1) x_{r,\theta} \quad \text{s.t.}$$
$$\sum_{r \in R} x_{r,\theta} \geq \mu_\theta f^{\mathsf{s}}_\theta(R) \quad \forall R \subseteq \mathcal{R}, \theta \in \Theta,$$

where the dual variables are $\{x_{r,\theta}\}_{r \in \mathcal{R}, \theta \in \Theta}$ (more details on the derivation are provided in Appendix D). A separation oracle for the dual problem above can be implemented applying the optimization oracle $\mathcal{O}(f^{\mathsf{s}}_\theta, -x_\theta/\mu_\theta)$ for each state of nature $\theta \in \Theta$. Let $R^*_\theta \coloneqq \mathcal{O}(f^{\mathsf{s}}_\theta, -x_\theta/\mu_\theta)$. If there exists $\theta \in \Theta$ such that

$$f^{\mathsf{s}}_\theta(R^*_\theta) - \sum_{r \in R^*_\theta} \frac{x_{r,\theta}}{\mu_\theta} \geq 0,$$

then we can use the violated constraint (θ, R^*_θ) as a separating hyperplane. Then, we can run the ellipsoid method equipped with such a separation oracle on the dual of LP (6). This procedure, together with known properties of the ellipsoid method (see, *e.g.,* (Khachiyan, 1980; Grötschel et al., 2012)), yields the following result.

Lemma 5.4. *Given access to an optimization oracle $\mathcal{O}$, there exists a polynomial-algorithm that solves LP (6).*

Next, we prove that the concave program of Equation (7) can be solved efficiently when having access to the optimization oracle $\mathcal{O}$. In order to solve the concave program of Equation (7), we start by rewriting the problem on the space of joint signaling schemes ϕ. To do that, we need to introduce constraints that ensure that the joint signaling scheme ϕ is well defined with respect to marginals φ (see Equation (9) in Appendix D). Then, we compute the Lagrangian relaxation of the resulting problem. By noticing that the problem is concave, and that Slater's condition holds, we recover strong duality. Finally, we use KKT conditions to remove the exponentially-many variables ϕ, and thereby obtaining a concave optimization problem with polynomially-many variables and exponentially-many constraints. By applying a similar procedure to the one we used for Lemma 5.4, we can solve such a problem via the ellipsoid algorithm by using the oracle $\mathcal{O}$ of Definition 5.4 as a separation oracle.

Lemma 5.5. *Given access to an optimization oracle $\mathcal{O}$, there exists a polynomial-time algorithm that solves the problem of Equation (7).*

By applying Lemma 5.3, Lemma 5.4, and Lemma 5.5, we can conclude the following:

Theorem 5.5. *In settings in which receivers have binary actions, and the sender has a monotone supermodular or a monotone anonymous utility function, Algorithm 2 has polynomial per-iteration running time and guarantees*

$$R_T \leq nd|A|\sqrt{mT}.$$

6. Further Applications

The main motivation for introducing the reduction from online problems with finite number of losses to online linear optimization of Section 3 was to solve online Bayesian persuasion problems. In this section, we highlight two further applications of our framework beyond Bayesian persuasion.

Learning in Security Games Balcan et al. (2015) extended classic (one-shot) security games (see, *e.g.,* (Tambe, 2011)) by introducing the problem of learning a no-regret strategy for the defender against a sequence of attackers that is adversarially selected. In their model, at each round t, the defender chooses a strategy $\boldsymbol{x}_t$, which is a distribution over N targets. Then, an attacker of type $d_t \in D$, best responds to such a strategy and the defender experiences a loss of $L_{d_t}(\boldsymbol{x}_t)$. Our reduction yields a $\tilde{O}(\text{poly}(D)\sqrt{T})$ regret bound under partial feedback, which improves the previously-known regret bound given by Balcan et al. (2015), which is of order $O(\text{poly}(ND)T^{2/3})$.

Online Bidding in Combinatorial Auction Daskalakis & Syrgkanis (2022) studied online learning in repeated combinatorial auctions. In these auctions the action space is combinatorial and, therefore, exponentially large. However, Daskalakis & Syrgkanis (2022) show that whenever the different number of bid profiles of the other bidders is finite and small (of size D), it is possible to design $O(\sqrt{DT})$ regret algorithms under *full feedback*. Our reduction to online linear optimization allows us to match their bound with full-information feedback, and also gives a $\tilde{O}(\text{poly}(D)\sqrt{T})$ bound for the more realistic case of *partial feedback*, *i.e.,* each player only observes their own utility.

Acknowledgements

This paper is supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence).

References

Abernethy, J., Hazan, E. E., and Rakhlin, A. Competing in the dark: An efficient algorithm for bandit linear optimization. In *21st Annual Conference on Learning Theory, COLT 2008*, pp. 263–273, 2008.

Alonso, R. and Câmara, O. Persuading voters. *American Economic Review*, 106(11):3590–3605, 2016.

Arieli, I. and Babichenko, Y. Private bayesian persuasion. *Journal of Economic Theory*, 182:185–217, 2019.

Babichenko, Y. and Barman, S. Algorithmic Aspects of Private Bayesian Persuasion. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67, pp. 34:1–34:16, 2017.

Babichenko, Y., Talgam-Cohen, I., Xu, H., and Zabarnyi, K. Regret-minimizing Bayesian persuasion. *arXiv preprint arXiv:2105.13870*, 2021.

Bacchiocchi, F., Castiglioni, M., Marchesi, A., Romano, G., and Gatti, N. Public signaling in bayesian ad auctions. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pp. 39–45. ijcai.org, 2022.

Badanidiyuru, A., Bhawalkar, K., and Xu, H. Targeting and signaling in ad auctions. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2545–2563, 2018.

Balcan, M.-F., Blum, A., Haghtalab, N., and Procaccia, A. D. Commitment without regrets: Online learning in Stackelberg security games. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 61–78, 2015.

Bernasconi, M., Castiglioni, M., Marchesi, A., Gatti, N., and Trovò, F. Sequential information design: Learning to persuade in the dark. In *NeurIPS*, 2022.

Bertsimas, D. and Tsitsiklis, J. N. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

Bhaskar, U., Cheng, Y., Ko, Y. K., and Swamy, C. Hardness results for signaling in Bayesian zero-sum and network routing games. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 479–496, 2016.

Blum, A., Haghtalab, N., and Procaccia, A. D. Learning optimal commitment to overcome insecurity. In *Advances in Neural Information Processing Systems*, pp. 1826–1834. 2014.

Bro Miltersen, P. and Sheffet, O. Send mixed signals: earn more, work less. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 234–247, 2012.

Camara, M. K., Hartline, J. D., and Johnsen, A. Mechanisms for a no-regret agent: Beyond the common prior. In *2020 ieee 61st annual symposium on foundations of computer science (focs)*, pp. 259–270. IEEE, 2020.

Castiglioni, M. and Gatti, N. Persuading voters in district-based elections. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5244–5251, 2021.

Castiglioni, M., Celli, A., and Gatti, N. Persuading voters: It's easy to whisper, it's hard to speak loud. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pp. 1870–1877, 2020a.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Online Bayesian persuasion. *Advances in Neural Information Processing Systems*, 33:16188–16198, 2020b.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Signaling in bayesian network congestion games: the subtle power of symmetry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5252–5259, 2021a.

Castiglioni, M., Marchesi, A., Celli, A., and Gatti, N. Multi-receiver online bayesian persuasion. In *International Conference on Machine Learning*, pp. 1314–1323. PMLR, 2021b.

Castiglioni, M., Marchesi, A., and Gatti, N. Bayesian persuasion meets mechanism design: Going beyond intractability with type reporting. pp. 226–234, 2022a.

Castiglioni, M., Marchesi, A., and Gatti, N. Bayesian persuasion meets mechanism design: Going beyond intractability with type reporting. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pp. 226–234, 2022b.

Castiglioni, M., Romano, G., Marchesi, A., and Gatti, N. Signaling in posted price auctions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(5):4941–4948, Jun. 2022c.

Castiglioni, M., Celli, A., Marchesi, A., and Gatti, N. Regret minimization in online bayesian persuasion: Handling adversarial receiver's types under full and partial feedback models. *Artificial Intelligence*, 314:103821, 2023.

Chen, Y., Liu, Y., and Podimata, C. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020.

Cheng, Y., Cheung, H. Y., Dughmi, S., Emamjomeh-Zadeh, E., Han, L., and Teng, S.-H. Mixture selection, mechanism design, and signaling. In *56th Annual Symposium on Foundations of Computer Science*, pp. 1426–1445, 2015.

Cohen, L. and Mansour, Y. Optimal algorithm for bayesian incentive-compatible exploration. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 135–151, 2019.

Daskalakis, C. and Syrgkanis, V. Learning in auctions: Regret is hard, envy is easy. *Games and Economic Behavior*, 2022.

Dughmi, S. Algorithmic information structure design: a survey. *ACM SIGecom Exchanges*, 15(2):2–24, 2017.

Dughmi, S. and Xu, H. Algorithmic Bayesian persuasion. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 412–425, 2016.

Dughmi, S. and Xu, H. Algorithmic persuasion with no externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 351–368, 2017.

Emek, Y., Feldman, M., Gamzu, I., PaesLeme, R., and Tennenholtz, M. Signaling schemes for revenue maximization. *ACM Transactions on Economics and Computation*, 2(2):1–19, 2014.

Fiorini, S., Rothvoß, T., and Tiwary, H. R. Extended formulations for polygons. *Discrete & computational geometry*, 48(3):658–668, 2012.

Gan, J., Majumdar, R., Radanovic, G., and Singla, A. Bayesian persuasion in sequential decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 5025–5033, 2022.

Grötschel, M., Lovász, L., and Schrijver, A. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

Hazan, E. and Kale, S. Extracting certainty from uncertainty: Regret bounded by variation in costs. *Machine learning*, 80(2):165–188, 2010.

Kamenica, E. and Gentzkow, M. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Khachiyan, L. G. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.

Kremer, I., Mansour, Y., and Perry, M. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

Letchford, J., Conitzer, V., and Munagala, K. Learning and approximating the optimal strategy to commit to. In *International Symposium on Algorithmic Game Theory*, pp. 250–262, 2009.

Mansour, Y., Slivkins, A., Syrgkanis, V., and Wu, Z. S. Bayesian exploration: Incentivizing exploration in Bayesian games. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 661–661, 2016.

Mansour, Y., Slivkins, A., Syrgkanis, V., and Wu, Z. S. Bayesian exploration: Incentivizing exploration in Bayesian games. *Operations Research*, 70(2):1105–1127, 2022.

Marecki, J., Tesauro, G., and Segal, R. Playing repeated Stackelberg games with unknown opponents. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pp. 821–828, 2012.

Nesterov, Y. and Nemirovskii, A. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Orabona, F. A modern introduction to online learning, 2019. URL https://arxiv.org/abs/1912.13213.

Rabinovich, Z., Jiang, A. X., Jain, M., and Xu, H. Information disclosure as a means to security. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp. 645–653, 2015.

Roughgarden, T. and Wang, J. R. Minimizing regret with multiple reserves. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–18, 2019.

Sellke, M. and Slivkins, A. The price of incentivizing exploration: A characterization via thompson sampling and sample complexity. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 795–796, 2021.

Shalev-Shwartz, S. et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Tambe, M. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge university press, 2011.

Vasserman, S., Feldman, M., and Hassidim, A. Implementing the wisdom of waze. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 660–666, 2015.

Wu, J., Zhang, Z., Feng, Z., Wang, Z., Yang, Z., Jordan, M. I., and Xu, H. Sequential information design: Markov persuasion process and its efficient reinforcement learning. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 471–472, 2022.

Xu, H. On the tractability of public persuasion with no externalities. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2708–2727. SIAM, 2020.

Xu, H., Freeman, R., Conitzer, V., Dughmi, S., and Tambe, M. Signaling in Bayesian Stackelberg games. In *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*, pp. 150–158, 2016.

Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 928–936, 2003.

Zu, Y., Iyer, K., and Xu, H. Learning to persuade on the fly: Robustness against ignorance. *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 927–928, 2021.

A. Proofs Omitted from Section 3

Theorem 3.1. *Algorithm 1 guarantees a cumulative regret $R_T \leq R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X}))$, where $R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X}))$ is the regret bound of a suitable regret minimizer $\mathfrak{R}$ for the set $\operatorname{co} \boldsymbol{\nu}(\mathcal{X})$.*

Proof. First, notice that, given any $\boldsymbol{z}_t \in \operatorname{co} \boldsymbol{\nu}(\mathcal{X})$, thanks to Carathéodory's theorem there always exist $D+1$ points $\{\boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^{D+1}\} \subset \boldsymbol{\nu}(\mathcal{X})$ and a corresponding probability distribution $\boldsymbol{\lambda} = (\lambda_t^1, \ldots, \lambda_t^{D+1}) \in \Delta^{D+1}$ such that $\boldsymbol{z}_t = \sum_{i=1}^{D+1} \lambda_t^i \, \boldsymbol{z}_t^i$. Such points $\boldsymbol{z}_t^i$ with their corresponding probabilities λ_t^i are those returned by the procedure $\textsc{Carathéodory}(\boldsymbol{z}_t, \boldsymbol{\nu}(\mathcal{X}))$ called by Algorithm 1. Thus, given how the algorithm selects the $\boldsymbol{x}_t \in \mathcal{X}$ to be played at each $t \in [T]$, it holds $\mathbb{E}\left[L_{d_t}(\boldsymbol{x}_t)\right] = \boldsymbol{\nu}(\boldsymbol{x}_t)^\top \mathbf{1}_{d_t} = \boldsymbol{z}_t^\top \mathbf{1}_{d_t}$.

Second, by using the no-regret property of the regret minimizer $\mathfrak{R}$, the following holds:

$$\begin{aligned} R_T &= \sum_{t=1}^T \mathbb{E}[L_{d_t}(\boldsymbol{x}_t)] - \min_{\boldsymbol{x} \in \mathcal{X}} \sum_{t=1}^T L_{d_t}(\boldsymbol{x}) \\ &= \sum_{t=1}^T \boldsymbol{z}_t^\top \mathbf{1}_{d_t} - \min_{\boldsymbol{z} \in \boldsymbol{\nu}(\mathcal{X})} \sum_{t=1}^T \boldsymbol{z}^\top \mathbf{1}_{d_t} \\ &\leq \sum_{t=1}^T \boldsymbol{z}_t^\top \mathbf{1}_{d_t} - \min_{\boldsymbol{z} \in \operatorname{co} \boldsymbol{\nu}(\mathcal{X})} \sum_{t=1}^T \boldsymbol{z}^\top \mathbf{1}_{d_t} \\ &\leq R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X})) \end{aligned}$$

where the first inequality holds since $\boldsymbol{\nu}(\mathcal{X}) \subseteq \operatorname{co} \boldsymbol{\nu}(\mathcal{X})$. □

Theorem 3.2. *If $\mathcal{X}$ is a polytope and $\boldsymbol{\nu}$ is a linear map, i.e., there exists $\mathbf{M} \in \mathbb{R}^{D \times M}$ such that $\nu(\boldsymbol{x}) = \mathbf{M}\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{X}$, then there exist algorithms implementing the* Carathéodory *oracle and the inverse map $\boldsymbol{\nu}^\dagger$.*

Proof. If $\mathcal{X}$ is a polytope and $\boldsymbol{\nu}$ is a linear map then $\boldsymbol{\nu}(\mathcal{X})$ is a polytope, and thus elements of $\operatorname{co} \boldsymbol{\nu}(\mathcal{X})$ correspond to elements of $\boldsymbol{\nu}(\mathcal{X})$. Therefore, the Carathéodory oracle can be implemented as just returning the one point density at $\boldsymbol{z}$ for every $\boldsymbol{z} \in \operatorname{co} \boldsymbol{\nu}(\mathcal{X})$.

Moreover, since $\boldsymbol{\nu}$ is linear we can implement $\boldsymbol{\nu}^\dagger$ by computing a generalized inverse of its matrix representation $\mathbf{M}$, and produce $\boldsymbol{\nu}^\dagger(\boldsymbol{z}) = \mathbf{M}^\dagger \boldsymbol{z} \in \mathcal{X}$. By definition of generalized inverse that holds for all $\boldsymbol{z} \in \boldsymbol{\nu}(\mathcal{X})$, *i.e.,* there exists an $\boldsymbol{x}$ such that $\mathbf{M}\boldsymbol{x} = \boldsymbol{z}$, we have that

$$\boldsymbol{\nu}(\boldsymbol{\nu}^\dagger(\boldsymbol{z})) = \mathbf{M}\mathbf{M}^\dagger \boldsymbol{z} = \mathbf{M}\mathbf{M}^\dagger \mathbf{M}\boldsymbol{x} = \mathbf{M}\boldsymbol{x} = \boldsymbol{z},$$

which concludes the proof. □

Corollary 3.3. *Under the assumptions of Theorem 3.2, with full feedback, there exists a regret minimizer $\mathfrak{R}$ such that Algorithm 1 guarantees cumulative regret*

$$R_T \leq \sqrt{DT}.$$

Proof. We can set $\mathfrak{R}$ to be Online Gradient Descent (OGD) (Zinkevich, 2003). Indeed, we have that the gradient of the losses in $\operatorname{co} \boldsymbol{\nu}(\mathcal{X})$ is bounded by 1 in the ℓ_2-norm, and that $\operatorname{co} \boldsymbol{\nu}(\mathcal{X}) \subset [0,1]^D$, which gives a D bound on the diameter w.r.t. the the ℓ_2-norm. Thus, by setting the learning rate of OGD as $\sqrt{D/T}$ we obtain a regret bound of $R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X})) \leq \sqrt{DT}$ (Orabona, 2019). □

Corollary 3.4. *Under the assumptions of Theorem 3.2, with bandit feedback, there exists a regret minimizer $\mathfrak{R}$ such that Algorithm 1 guarantees cumulative regret*

$$R_T \leq 16 D^{3/2} \sqrt{T \log T}.$$

Proof. Under partial feedback, we obtain the regret bound above by equipping Algorithm 1 with a suitably-defined regret minimizer $\mathfrak{R}$. In particular, $\mathfrak{R}$ must work by observing only realizations of an unbiased estimator of $\boldsymbol{z}_t^\top \mathbf{1}_{d_t}$ instead of its actual value, since Algorithm 1 does *not* play $\boldsymbol{z}_t$, but it employs a sampling process that is equivalent to playing $\boldsymbol{z}_t$ in expectation. Such a regret minimizer $\mathfrak{R}$ can be implemented by the algorithm introduced by Abernethy et al. (2008), as any polytope in $\mathbb{R}^D$ has a D-self concordant barrier Nesterov & Nemirovskii (1994, Theorem 2.5.1). This yields $R_T^{\mathfrak{R}}(\operatorname{co} \boldsymbol{\nu}(\mathcal{X})) \leq 16 D^{3/2} (T \log T)^{1/2}$, which proves our statement. □

B. Proofs Omitted from Section 5.2

Theorem 5.1. *The single-receiver online Bayesian persuasion problem with type reporting admits an algorithm which guarantees regret $R_T \le \sqrt{mT}$ and polynomial per-iteration running time.*

Proof. By Corollary 3.3, Algorithm 1 produces a sequence $(\ell_t)_{t=1}^T, \ell_t \in \mathcal{L}$, such that

$$\sum_{t=1}^{T} \boldsymbol{\nu}(\ell_t)^\top \mathbf{1}_{k_t} - \min_{\ell \in \mathcal{L}} \sum_{t=1}^{T} \boldsymbol{\nu}(\ell)^\top \mathbf{1}_{k_t} \le \sqrt{mT}.$$

Then, the sender commits to the menu which is the projection of ℓ_t onto Λ, *i.e.*, $\varphi_t = \pi(\ell_t)$. Since $\boldsymbol{\nu}(\ell)$ is independent from the extension variables l we get that:

$$\boldsymbol{\nu}(\ell_t)^\top \mathbf{1}_{k_t} = -u^{\mathsf{s}}(\pi(\ell_t), k_t) = -u^{\mathsf{s}}(\varphi_t, k_t)$$

and similarly $\boldsymbol{\nu}(\ell)^\top \mathbf{1}_{k_t} = -u^{\mathsf{s}}(\varphi, k_t)$, which proves the statement. □

C. Proofs Omitted from Section 5.3

Lemma C.1. *For any $\varphi \in \Lambda$ we can write $g^{\boldsymbol{k}}(\varphi)$ as a solution of a standard-form linear program with $|\mathcal{A}| \cdot |\Theta|$ variables and constraints, and in such a standard-form linear program, the variables φ, are its* right-hand *side vector.*

Proof. We define a standard form linear program with n variables and n constraints if it is of the form:

$$\max_{\boldsymbol{x}} \boldsymbol{c}^\top \boldsymbol{x},\ s.t.$$
$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \ge 0,$$

where $\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$. We define two one-to-one mappings $\pi_1 : |\mathcal{A}| \times |\Theta| \to [|\mathcal{A}| \cdot |\Theta|]$ and $\pi_2 : |\mathcal{R}| \times |A| \times |\Theta| \to [|\mathcal{R}| \cdot |A| \cdot |\Theta|]$ such that $\pi_1(\cdot)$ associate every tuple of actions $\boldsymbol{a}$ and state of nature θ to the index $\pi_1(\boldsymbol{a}, \theta)$, while $\pi_2(\cdot)$ associate every receiver r, action $a \in A$ and state of nature θ to the index $\pi_2(r, a, \theta)$. Then we can define $i := \pi_1(\boldsymbol{a}, \theta)$ and $j := \pi_2(r, a, \theta')$ so that:

- $\boldsymbol{x}[i] := \phi_\theta(\boldsymbol{a})$
- $\boldsymbol{c}[i] := \mu_\theta \cdot u^{\mathsf{s}}(\boldsymbol{a}, \theta)$
- $\boldsymbol{b}[j] := \varphi_\theta^{r,k_r}(a)$
- $\mathbf{A}[j, i] := \mathbb{I}(a_r = a, \theta = \theta')$.

Then we can write LP 6 as $\max_{\boldsymbol{x}} \boldsymbol{c}^\top \boldsymbol{x}$ subject to $\boldsymbol{x} \ge 0$ and $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$. We note that the variables φ only appear in the right-hand side vector $\boldsymbol{b}$ in the standard-form linear program above. □

Lemma 5.1. *The function $g^{\boldsymbol{k}}(\varphi)$ is concave in φ on Λ for each type profile $\boldsymbol{k} \in \mathcal{K}$.*

Proof. Let $\boldsymbol{k} \in \mathcal{K}$ be a tuple of types. Lemma C.1 relates the solution $g^{\boldsymbol{k}}(\varphi)$ of LP 6 to the solution of a standard-form linear program in which φ is the right-hand side vector of an equality constraint. Thus, for every fixed $\boldsymbol{k}$, the function $g^{\boldsymbol{k}}(\varphi)$ is known to be concave in φ (Bertsimas & Tsitsiklis, 1997, Theorem 5.1). □

Lemma 5.2. *For each $\boldsymbol{k} \in \mathcal{K}$, the function $g^{\boldsymbol{k}}(\varphi)$ is $\sqrt{nd|A|}$-Lipschitz-continuous in φ with respect to $\|\cdot\|_2$.*

Proof. First we note that for any fixed tuple of types $\boldsymbol{k}$, the menus $\varphi_\theta^{r,k}$ for $k \neq k_r$ do not appear, thus, in this proof, we can ease the notion by dropping k_r from φ_θ^{r,k_r}, which will be denoted by just φ_θ^r.

Then, for ease of clarity, we define

$$o^{\boldsymbol{k}}(\phi) := \sum_{\theta \in \Theta} \sum_{\boldsymbol{a} \in \mathcal{A}} \mu_\theta \phi_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta),$$

and

$$\mathcal{M}^{\boldsymbol{k}}(\varphi) := \left\{ \phi \,\middle|\, \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r\in\hat{a}}} \phi_\theta(\boldsymbol{a}) = \varphi^r_\theta(\hat{a}),\, \forall r\in\mathcal{R}, \hat{a}\in\mathcal{A}_r, \theta\in\Theta \right\},$$

which are the objective function and the constraints polytope of LP 6, respectively. Formally, it holds that

$$g^{\boldsymbol{k}}(\varphi) = \max_{\phi\in\mathcal{M}^{\boldsymbol{k}}(\varphi)} o^{\boldsymbol{k}}(\phi).$$

We will also use the function $\pi_2 : \mathcal{R}\times\mathcal{A}\times\Theta \to [|\mathcal{R}|\cdot|\mathcal{A}|\cdot|\Theta|]$ introduced in Lemma C.1, that associate for every $(\hat{r},\hat{a},\hat{\theta})\in\mathcal{R}\times\mathcal{A}_r\times\Theta$ an index $i=\pi_2(\hat{r},\hat{a},\hat{\theta})$. We first prove the 1-Lipschitzness of $g^{\boldsymbol{k}}$ w.r.t. to $\|\cdot\|_1$. Consider any two $\varphi,\overline{\varphi}\in\Lambda$.

Let then $\phi\in\arg\max_{\phi'\in\mathcal{M}^{\boldsymbol{k}}(\varphi)} o^{\boldsymbol{k}}(\phi')$ and $\overline{\phi}\in\arg\max_{\phi'\in\mathcal{M}^{\boldsymbol{k}}(\overline{\varphi})} o^{\boldsymbol{k}}(\phi')$ the values of the solutions of LP 6 w.r.t. φ and $\overline{\varphi}$, respectively.

The idea of the proof is to construct a new variable $\varphi^\star$ and $\phi^\star$ that satisfies the following conditions:

1. $\phi^\star\in\mathcal{M}^{\boldsymbol{k}}(\phi^\star)$

2. $0\preceq\varphi^\star\preceq\overline{\varphi}$, which has to be interpreted element-wise.

3. $\|\varphi-\overline{\varphi}\|_1 + o^{\boldsymbol{k}}(\phi^\star)\geq o^{\boldsymbol{k}}(\phi)$.

Note that we do not require that $\varphi^\star\in\Lambda$. Assume that we can have such a $\varphi^\star$ and $\phi^\star$ then we can easily prove 1-Lipschitzness w.r.t. $\|\cdot\|_1$ as follows:

$$\begin{aligned} g^{\boldsymbol{k}}(\overline{\varphi}) &\geq o^{\boldsymbol{k}}\left(\overline{\phi}\right)\\ &\geq o^{\boldsymbol{k}}(\phi^\star)\\ &\geq o^{\boldsymbol{k}}(\phi) - \|\varphi-\overline{\varphi}\|_1\\ &= g^{\boldsymbol{k}}(\varphi) - \|\varphi-\overline{\varphi}\|_1, \end{aligned}$$

where the first inequality holds since $\overline{\varphi}-\varphi^\star\succeq 0$ by assumption and thus $\overline{\phi}\succeq\phi^\star$ which implies that $o^{\boldsymbol{k}}(\overline{\phi})\geq o^{\boldsymbol{k}}(\phi^\star)$, and the second inequality holds by assumption on $\varphi^\star$. This in turn implies that $|g^{\boldsymbol{k}}(\overline{\varphi})-g^{\boldsymbol{k}}(\varphi)|\leq\|\varphi-\overline{\varphi}\|_1$ since the construction is symmetric w.r.t. φ and $\overline{\varphi}$. After we prove that $|g^{\boldsymbol{k}}(\overline{\varphi})-g^{\boldsymbol{k}}(\varphi)|\leq\|\varphi-\overline{\varphi}\|_1$ we can easily conclude the proof by observing that $\|\varphi-\overline{\varphi}\|_1\leq\sqrt{nd|\mathcal{A}_r|}\cdot\|\varphi-\overline{\varphi}\|_2$.

Now we show the existence such a $\varphi^\star$ and the related $\phi^\star\in\mathcal{M}^{\boldsymbol{k}}(\varphi^\star)$ by explicitly building it iteratively as follows. The procedure above maintains variables (φ^t,ϕ^t) that is updated as detailed in Algorithm 3.

Algorithm 3

```
$\varphi^0 \leftarrow \varphi$
$\phi^0 \leftarrow \phi$
$T \leftarrow |\mathcal{R}| \cdot |\mathcal{A}_r| \cdot |\Theta|$
$\tilde{\varphi} \leftarrow \min(\overline{\varphi}, \varphi)$
for $t = 1$ to $T$ do
    $(\hat{r}, \hat{a}, \hat{\theta}) \leftarrow \pi_2^{-1}(t)$
    $\delta_t \leftarrow \varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a}) - \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})$
    $\varphi^t \leftarrow \varphi^{t-1}$
    $\phi^t \leftarrow \phi^{t-1}$
    if $\varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a}) \geq \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})$ then
        $\varphi_{\hat{\theta}}^{t,\hat{r}}(\hat{a}) \leftarrow \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})$
        $\varphi_{\hat{\theta}}^{t,r'}(a') \leftarrow \varphi_{\hat{\theta}}^{t-1,\hat{r}}(a') - \frac{\delta_t}{\varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a})} \sum_{\boldsymbol{a} \in \mathcal{A}: a_r = \hat{a}, a_{r'} = a'} \phi_{\hat{\theta}}^{t-1}(\boldsymbol{a})$, $\forall r' \neq \hat{r}, a' \in \mathcal{A}_{r'}$
        $\phi_{\hat{\theta}}^{t}(\boldsymbol{a}) \leftarrow \phi_{\hat{\theta}}^{t-1}(\boldsymbol{a}) \left(1 - \frac{\delta_t}{\varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a})}\right)$, $\forall \boldsymbol{a} : a_r = \hat{a}$
    end if
end for
return $\varphi^\star := \varphi^T, \phi^\star := \phi^T$
```

The idea of the procedure in Algorithm 3, is to maintain the constraints $\phi^t \in \mathcal{M}^{\boldsymbol{k}}(\varphi^t)$ valid trough tout the procedure, and to update ϕ^t as to guarantee that $o^{\boldsymbol{k}}(\phi^t) \geq o^{\boldsymbol{k}}(\phi^{t-1}) - \delta_t$.

Now we see that the constraints $\phi^t \in \mathcal{M}^{\boldsymbol{k}}(\varphi^t)$ are maintained at iteration t, assuming that are satisfied at time $t-1$.

Define $(\hat{r}, \hat{\theta}, \hat{a}) = \pi_2^{-1}(t)$ and consider the following two cases:

- If $\varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a}) \leq \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})$:

Then we trivially have that $\phi^t \in \mathcal{M}^{\boldsymbol{k}}(\varphi^t)$ as $\phi^t = \phi^{t-1}$ and $\varphi^t = \varphi^{t-1}$ and $\phi^{t-1} \in \mathcal{M}^{\boldsymbol{k}}(\varphi^{t-1})$ by assumption.

- If otherwise $\varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a}) \geq \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})$. We can divide the variables $(r, a, \theta) \in \mathcal{R} \times \mathcal{A}_r \times \Theta$ into three sets

a) $A_1 := \{(r, \theta, \hat{a})\}$

b) $A_2 := \{(r, \theta, a) : a \in \mathcal{A}_r, a \neq \hat{a}\}$

c) $A_3 := \{(r', a', \hat{\theta}) : r' \in \mathcal{R}/\{\hat{r}\}, a' \in \mathcal{A}_{r'}\}$

d) $A_4 := \{(r, a, \theta') : \theta' \in \Theta, \theta' \neq \hat{\theta}\}$

Notice that these sets are disjoint and their union is $\mathcal{R} \times \mathcal{A}_r \times \Theta$.

a) For any $(r, a, \theta) \in A_1$ we have:

$$\begin{aligned}
\sum_{\boldsymbol{a} \in \mathcal{A}: a_r = a} \phi_\theta^t(\boldsymbol{a}) &= \sum_{\boldsymbol{a} \in \mathcal{A}: a_r = a} \phi_\theta^{t-1}(\boldsymbol{a}) \left(1 - \frac{\delta_t}{\varphi_\theta^{t-1,r}(a)}\right) \\
&= \varphi_\theta^{t-1}(a) \left(1 - \frac{\delta_t}{\varphi_\theta^{t-1}(a)}\right) \\
&= \varphi_\theta^{t-1}(a) - \delta_t \\
&= \tilde{\varphi}_\theta^r(a).
\end{aligned}$$

b) For any $(r, a, \theta) \in A_2$ we have:

$$\sum_{\boldsymbol{a} \in \mathcal{A}: a_r = a'} \phi_\theta^t(\boldsymbol{a}) = \sum_{\boldsymbol{a} \in \mathcal{A}: a_r = a'} \phi_\theta^{t-1}(\boldsymbol{a}) = \varphi_\theta^{t-1,r}(a') = \varphi_\theta^{t,r}(a').$$

as the those variable are not updated at round t.

c) For any $(r, a, \theta) \in A_3$ we have:

$$
\begin{aligned}
\sum_{\boldsymbol{a}\in\mathcal{A}:a_r=a} \phi^t_\theta(\boldsymbol{a}) &= \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a\\ a_{\hat{r}}=\hat{a}}} \phi^t_\theta(\boldsymbol{a}) + \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a\\ a_{\hat{r}}\neq\hat{a}}} \phi^t_\theta(\boldsymbol{a}) \\
&= \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a\\ a_{\hat{r}}=\hat{a}}} \phi^{t-1}_\theta(\boldsymbol{a}) \left(1 - \frac{\delta_t}{\varphi^{t-1,\hat{r}}_\theta(\hat{a})}\right) + \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a\\ a_{\hat{r}}\neq\hat{a}}} \phi^{t-1}_\theta(\boldsymbol{a}) \\
&= \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a}} \phi^{t-1}_{\theta'}(\boldsymbol{a}) - \frac{\delta_t}{\varphi^{t-1,\hat{r}}_\theta(\hat{a})} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_r=a\\ a_{\hat{r}}=\hat{a}}} \phi^{t-1}_\theta(\boldsymbol{a}) \\
&= \varphi^{t,r}_\theta(a),
\end{aligned}
$$

where for the second equality we used the update of update of $\phi^{t-1}(\boldsymbol{a})$ in Line 13 of Algorithm 3. While the last equality follows from the update of Line 12.

d) For any $(r, a, \theta) \in A_4$ we have that none of the variable are updated an thus the statement holds by inductive assumption.

This proves that $\phi^\star \in \mathcal{M}^{\boldsymbol{k}}(\varphi^\star)$.

On the other hand it is evident that $\varphi^\star \preceq \overline{\varphi}$ thanks to update of Line 11 in Algorithm 3. In particular it also holds that $\varphi^{t,\hat{r}}_{\hat{\theta}}(\hat{a}) \leq \overline{\varphi}^{\hat{r}}_{\hat{\theta}}(\hat{a})$ for all $t = \pi_2(\hat{r}, \hat{a}, \hat{\theta})$.

We are left to show that $\|\varphi - \overline{\varphi}\|_1 + o^{\boldsymbol{k}}(\phi^\star) \geq o^{\boldsymbol{k}}(\phi)$. Consider the following inequalities:

$$
\begin{aligned}
o^{\boldsymbol{k}}(\phi^t) &:= \sum_{\theta\in\Theta}\sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^t_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \\
&= \mu_{\hat{\theta}} \sum_{\boldsymbol{a}\in\mathcal{A}} \phi^t_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \sum_{\theta\in\Theta/\{\hat{\theta}\}} \sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^t_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \\
&= \mu_{\hat{\theta}} \sum_{\boldsymbol{a}\in\mathcal{A}} \phi^t_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \sum_{\theta\in\Theta/\{\hat{\theta}\}} \sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^{t-1}_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \\
&= \mu_{\hat{\theta}} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}=\hat{a}}} \phi^t_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \mu_{\hat{\theta}} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}\neq\hat{a}}} \phi^t_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \sum_{\theta\in\Theta/\{\hat{\theta}\}} \sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^{t-1}_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \\
&= \mu_{\hat{\theta}} \left(1 - \frac{\delta_t}{\varphi^{t-1,\hat{r}}_{\hat{\theta}}(\hat{a})}\right) \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}=\hat{a}}} \phi^{t-1}_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \mu_{\hat{\theta}} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}\neq\hat{a}}} \phi^{t-1}_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) + \sum_{\theta\in\Theta/\{\hat{\theta}\}} \sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^{t-1}_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) \\
&= \sum_{\theta\in\Theta}\sum_{\boldsymbol{a}\in\mathcal{A}} \mu_\theta \phi^{t-1}_\theta(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \theta) - \frac{\delta_t}{\varphi^{t-1,\hat{r}}_{\hat{\theta}}(\hat{a})} \mu_{\hat{\theta}} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}=\hat{a}}} \phi^{t-1}_{\hat{\theta}}(\boldsymbol{a}) u^{\mathsf{s}}(\boldsymbol{a}, \hat{\theta}) \\
&\geq o^{\boldsymbol{k}}(\phi^{t-1}_\theta) - \frac{\delta_t}{\varphi^{t-1,\hat{r}}_{\hat{\theta}}(\hat{a})} \sum_{\substack{\boldsymbol{a}\in\mathcal{A}:\\ a_{\hat{r}}=\hat{a}}} \phi^{t-1}_{\hat{\theta}}(\boldsymbol{a}) \\
&= o^{\boldsymbol{k}}(\phi^{t-1}_\theta) - \delta_t.
\end{aligned}
$$

Then we can telescope the inequality to show that:

$$
o^{\boldsymbol{k}}(\phi^\star) \geq o^{\boldsymbol{k}}(\phi) - \sum_{t=1}^{T} \delta_t.
$$

Then it is easy to show that $\delta_t = \varphi_{\hat{\theta}}^{t-1,\hat{r}}(\hat{a}) - \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a}) \leq \overline{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a}) - \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a}) \leq |\overline{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a}) - \tilde{\varphi}_{\hat{\theta}}^{\hat{r}}(\hat{a})|$ and thus

$$o^{\boldsymbol{k}}(\phi^\star) \geq o^{\boldsymbol{k}}(\phi) - \|\varphi - \overline{\varphi}\|_1,$$

as wanted. □

Theorem 5.2. *Let $\alpha := \sqrt{m/T}$. Algorithm 2 guarantees a cumulative regret $R_T \leq nd|A|\sqrt{mT}$.*

Proof. First notice that:

$$\max_{\varphi \in \Lambda} \sum_{t=1}^{T} g^{\boldsymbol{k}_t}(\varphi) = \max_{\varphi \in \Lambda} \sum_{t=1}^{T} u^{\mathsf{s}}(\varphi, \boldsymbol{k}_t)$$

which follows from the definition of $u^{\mathsf{s}}(\varphi, \boldsymbol{k})$ given in Equation (4). On the other hand it is clear that $g^{\boldsymbol{k}_t}(\varphi_t) = u^{\mathsf{s}}(\phi_t, \boldsymbol{k}_t)$ thanks to the update of Line 4 of Algorithm 2.

Thus we can write the regret of Algorithm 2 as:

$$R_T = \max_{\varphi \in \Lambda} \sum_{t=1}^{T} g^{\boldsymbol{k}_t}(\varphi) - \sum_{t=1}^{T} g^{\boldsymbol{k}_t}(\varphi_t).$$

Then, by Lemma 5.1 we have that the reward functions $g^{\boldsymbol{k}}(\cdot)$ are concave for all $\boldsymbol{k} \in \mathcal{K}$.

Moreover we know that the by Lemma 5.2 that for all $\boldsymbol{k} \in \mathcal{K}$ the functions $g^{\boldsymbol{k}}(\cdot)$, are $\sqrt{nd|A|}$-Lipschitz w.r.t. $\|\cdot\|_2$ and thus, by Shalev-Shwartz et al. (2012, Lemma 2.6), we have that all the subgradients of $-g^{\boldsymbol{k}}(\cdot)$ have norm bounded by the Lipschitz constant. This clearly implies $G := \sup_{\varphi \in \Lambda} \|\partial g^{\boldsymbol{k}}(\varphi)\|_2 \leq \sqrt{nd|A|}$.

Moreover, the regularizer $\frac{1}{2}\|\cdot\|_2^2$ is trivially 1-strongly convex w.r.t. $\|\cdot\|_2$.

Finally we have that the diameter of the polytope Λ, induced by the regularizers is bounded by $\frac{1}{2}ndm|A|$, as Λ is a contained in the $ndm|A|$-dimensional hypercube. Formally $D := \sqrt{\max_{\phi \in \Lambda} \frac{1}{2}\|\phi\|_2^2 - \min_{\phi' \in \Lambda} \frac{1}{2}\|\phi'\|_2^2} \leq \sqrt{\frac{1}{2}ndm|A|}$.

A standard application of Orabona (2019, Corollary 7.9) gives a bound of:

$$R_T \leq \frac{D^2}{\alpha} + \frac{1}{2}\alpha G^2 T \leq \frac{1}{2\alpha} ndm|A| + \frac{1}{2}\alpha nd|A|T.$$

Setting $\alpha = \sqrt{m/T}$ gives the result. □

D. Proofs Omitted from Section 5.4

Lemma 5.4. *Given access to an optimization oracle $\mathcal{O}$, there exists a polynomial-algorithm that solves LP (6).*

Proof. We defining for any $R \subset \mathcal{R}$, $\boldsymbol{a}_R$ as the tuple in which action a_1 is recommended to all the receivers in R and a_0 to the others. Formally $a_r = a_1$ for all $r \in R$, and $a_r = a_0$ for all $r \in \mathcal{R}/R$. Then , rewriting LP 6 for the specific case of binary actions per receiver, we obtain:

$$\max_{\phi \geq 0} \sum_{\theta \in \Theta} \sum_{R \subseteq \mathcal{R}} \mu_\theta \phi_\theta(\boldsymbol{a}_R) f_\theta^{\mathsf{s}}(R) \quad \text{s.t.} \tag{8a}$$

$$\sum_{R \in \mathcal{R}: r \in R} \phi_\theta(\boldsymbol{a}_R) = \varphi_\theta^{r,k_r}(a_1), \quad \forall r \in \mathcal{R}, \forall \theta \in \Theta \tag{8b}$$

$$\sum_{R \in \mathcal{R}} \phi_\theta(\boldsymbol{a}_R) = 1, \quad \forall \theta \in \Theta \tag{8c}$$

The dual of such LP reads as follows:

$$\min_x \sum_{r \in \mathcal{R}, \theta \in \Theta} \varphi_\theta^{r,k_r}(a_1) x_{r,\theta} \quad \text{s.t.}$$

$$\sum_{r \in R} x_{r,\theta} \geq \mu_\theta f^{\mathsf{s}}_\theta(R), \quad \forall R \subseteq \mathcal{R}, \theta \in \Theta,$$

where the dual variables are $\{x_{r,\theta}\}_{r\in\mathcal{R},\theta\in\Theta}$. A separation oracle for dual problem can be implemented exploiting the optimization oracle $\mathcal{O}(f^{\mathsf{s}}_\theta, -x_\theta/\mu_\theta)$ for each θ. If,for at least one θ, the value of $\mathcal{O}(f^{\mathsf{s}}_\theta, -x_\theta/\mu_\theta)$ is larger that 0 then we can use the the violated constraint as a separating hyperplane. □

Lemma 5.5. *Given access to an optimization oracle $\mathcal{O}$, there exists a polynomial-time algorithm that solves the problem of Equation* (7).

Proof. We defining for any $R \subset \mathcal{R}$, $\boldsymbol{a}_R$ as the tuple in which action a_1 is recommended to all the receivers in R and a_0 to the others. Formally $a_r = a_1$ for all $r \in R$, and $a_r = a_0$ for all $r \in \mathcal{R}/R$. With this definition, for any sequence of type's tuples, the problem $\max_{\varphi\in\Lambda}\sum_{\tau\in[t]} g^{\boldsymbol{k}_\tau}(\varphi) - \frac{1}{2\alpha}\|\varphi\|_2^2$ can be rewritten as:

$$\max_{\phi\geq 0,\varphi\in\Lambda} \sum_{\substack{\tau\in[t]\\ \theta\in\Theta\\ R\subseteq\mathcal{R}}} \mu_\theta \phi_{\tau,\theta}(\boldsymbol{a}_R) f^{\mathsf{s}}_\theta(R) - \frac{1}{2\alpha} \sum_{\substack{r\in\mathcal{R},k\in\mathcal{K}_r,\\ \theta\in\Theta,a\in A}} \varphi^{r,k}_\theta(a)^2 \quad \text{s.t.} \tag{9a}$$

$$\sum_{\substack{R\subseteq\mathcal{R}:\\ r\in R}} \phi_{\tau,\theta}(\boldsymbol{a}_R) = \varphi^{r,k_{\tau,r}}_\theta(a_1), \quad \forall r\in\mathcal{R}, \theta\in\Theta, \tau\in[t] \tag{9b}$$

$$\sum_{R\subseteq\mathcal{R}} \phi_{\tau,\theta}(\boldsymbol{a}_R) = 1, \quad \forall\tau\in[t], \theta\in\Theta \tag{9c}$$

We Lagrangyfing Problem (9) by introducing the following dual variables

- $x_{r,\theta,\tau} \in \mathbb{R}$ for each $r \in \mathcal{R}, \theta \in \theta, \tau \in [t]$, which is the dual variable of the constrain 6b
- $y_{\tau,\theta} \in \mathbb{R}$ for each $\tau \in [t]$, $\theta \in \Theta$, which is the dual variable of the constrain 9c
- $z_{r,k,k'} \in \mathbb{R}_+$ for each $r \in \mathcal{R}, k, k' \in \mathcal{K}_r$, which is the dual variable of the constrain 5a
- $\alpha_{r,k,k',a,a'} \in \mathbb{R}_+$ for each $r \in \mathcal{R}, k, k' \in \mathcal{K}_r, a, a' \in A$, which is the dual variable of the constrain 5b
- $\beta_{r,k,\theta} \in \mathbb{R}$ for each $r \in \mathcal{R}, k \in \mathcal{K}_r, \theta \in \Theta$, which is the dual variable of the constrain 5c
- $\gamma_{\theta,R} \in \mathbb{R}_+$ for each $\theta \in \Theta$, $R \subseteq \mathcal{R}$, for the constraint $\phi \geq 0$
- $\eta_{r,k,\theta,a} \in \mathbb{R}_+$ for each $r \in \mathcal{R}$, $k \in \mathcal{K}_r$, $\theta \in \Theta$, and $a \in A$, for the constraint $\varphi \geq 0$

The the Lagrangian of Problem 9 reads:

$$\begin{aligned} L(\phi,\varphi,x,y,z,\alpha,\beta,\gamma,\eta) = & \sum_{\substack{\tau\in[t],\theta\in\Theta\\ R\subseteq\mathcal{R}}} \mu_\theta \phi_{\tau,\theta}(\boldsymbol{a}_R) f^{\mathsf{s}}_\theta(R) - \frac{1}{2\alpha} \sum_{\substack{r\in\mathcal{R},k\in\mathcal{K}_r,\\ \theta\in\Theta,a\in A}} \varphi^{r,k}_\theta(a)^2 \\ & + \sum_{\substack{\tau\in[t],\theta\in\Theta\\ r\in\mathcal{R}}} x_{r,\theta,\tau} \left(\sum_{\substack{R\subseteq\mathcal{R}:\\ r\in R}} \phi_{\tau,\theta}(\boldsymbol{a}_R) - \varphi^{r,k_{\tau,r}}_\theta(a_1) \right) \\ & + \sum_{\substack{\tau\in[t],\\ \theta\in\Theta}} y_{\tau,\theta} \left(\sum_{R\subseteq\mathcal{R}} \phi_{\tau,\theta}(\boldsymbol{a}_R) - 1 \right) \\ & + \sum_{\substack{r\in\mathcal{R},\\ k,k'\in\mathcal{K}_r}} z_{r,k,k'} \left(\sum_{a\in A}\sum_{\theta\in\Theta} \mu_\theta\, \varphi^{r,k}_\theta(a)\, u^r_k(a,\theta) - \sum_{a\in A} l^{r,k,k'}_a \right) \end{aligned}$$

$$+\sum_{\substack{r\in\mathcal{R},k,k'\in\mathcal{K}_r,\\ a,a'\in A}}\alpha_{r,k,k',a,a'}\left(l_a^{r,k,k'}-\sum_{\theta\in\Theta}\mu_\theta\,\varphi_\theta^{r,k'}(a)\,u_k^r(a',\theta)\right)$$

$$+\sum_{\substack{r\in\mathcal{R},\\ k\in\mathcal{K}_r,\theta\in\Theta}}\beta_{r,k,\theta}\left(\sum_{a\in A}\varphi_\theta^{r,k}(a)-1\right)+\sum_{\substack{\theta\in\Theta,\\ R\subseteq\mathcal{R}}}\gamma_{\theta,R}\phi_\theta(\boldsymbol{a}_R)+\sum_{\substack{r\in\mathcal{R},k\in\mathcal{K}_r,\\ \theta\in\Theta,a\in A}}\eta_{r,k,\theta,a}\varphi_\theta^{r,k}(a).$$

We observe that Slater's condition holds for Problem 9. This holds since all constraints are linear and there exists a feasible solution. This is easily seen as there exists a set of feasible menu of IC marginal signaling schemes. Moreover, given a set of menus and a vector of types, it is possible to design consistent signaling schemes by taking the product distribution of the marginal signaling schemes relative to the types. Therefore, by strong duality, the optimal primal and dual variables must satisfy the KKT conditions. In particular it must hold that $\mathbf{0}\in\partial_{\phi_{\tau,\theta}(\boldsymbol{a}_R)}(L)$ for each $\tau\in[t]$, $\theta\in\Theta$, $R\subseteq\mathcal{R}$. Formally, for each $\tau\in[t]$, $\theta\in\Theta$, and $R\subseteq\mathcal{R}$, we have:

$$\partial_{\phi_{\tau,\theta}(\boldsymbol{a}_R)}(L)=\mu_\theta f_\theta^{\mathsf{s}}(R)+\sum_{r\in R}x_{r,\theta,\tau}+y_{\tau,\theta}+\gamma_{\theta,R}=0. \tag{10}$$

Moreover, it must also hold that $\mathbf{0}\in\partial_{\varphi_\theta^{r,k}(a)}(L)$. Formally, for each $r\in\mathcal{R}$, $k\in\mathcal{K}_r$, $\theta\in\Theta$, and $a\in A$ it holds

$$-\frac{\varphi_\theta^{r,k}(a)}{\alpha}-\mathbb{I}_{a=a_1}\sum_{\substack{\tau\in[t]:\\ k=k_{\tau,r}}}x_{r,\theta,\tau}+\left(\sum_{k'\in\mathcal{K}_r}z_{r,k,k'}\right)\mu_\theta u_k^r(a,\theta)-\sum_{\substack{a'\in A,\\ k'\in\mathcal{K}_r}}\alpha_{r,k',k,a,a'}\mu_\theta u_{k'}^r(a',\theta)+\beta_{r,k,\theta}+\eta_{r,k,\theta,a}=0,$$

which implies that for each $r\in\mathcal{R}$, $k\in\mathcal{K}_r$, $\theta\in\Theta$, and $a\in A$:

$$\frac{\varphi_\theta^{r,k}(a)}{\alpha}=-\mathbb{I}_{a=a_1}\sum_{\substack{\tau\in[t]:\\ k=k_{\tau,r}}}x_{r,\theta,\tau}+\left(\sum_{k'\in\mathcal{K}_r}z_{r,k,k'}\right)\mu_\theta u_k^r(a,\theta)-\sum_{\substack{a'\in A,\\ k'\in\mathcal{K}_r}}\alpha_{r,k',k,a,a'}\mu_\theta u_{k'}^r(a',\theta)+\beta_{r,k,\theta}+\eta_{r,k,\theta,a}. \tag{11}$$

Similarly, it must hold that $\mathbf{0}\in\partial_{l_a^{r,k,k'}}(L)$. Formally, for each $r\in\mathcal{R}$, $k,k'\in\mathcal{K}_r$, and $a\in A$, it holds

$$\partial_{l_a^{r,k,k'}}(L)=-z_{r,k,k'}+\sum_{a'\in A}\alpha_{r,k,k',a,a'}=0 \tag{12}$$

Finally, plugging Equation (10), Equation (11) and Equation (12) back into the Lagrangian we get:

$$L(\phi,\varphi,x,y,z,\alpha,\beta,\gamma,\eta)=\frac{1}{2\alpha}\sum_{\substack{r\in\mathcal{R},k\in\mathcal{K}_r,\\ \theta\in\Theta,a\in A}}\varphi_\theta^{r,k}(a)^2-\sum_{\tau\in[t],\theta\in\Theta}y_{\tau,\theta}-\sum_{r\in\mathcal{R},k\in\mathcal{K}_r,\theta\in\Theta}\beta_{r,k,\theta}.$$

Finally the dual problem of Problem 9 can be written as follows:

$$\min_{\varphi,x,\beta\le 0}\left\{\frac{1}{2\alpha}\sum_{\substack{r\in\mathcal{R},k\in\mathcal{K}_r,\\ \theta\in\Theta,a\in A}}\varphi_\theta^{r,k}(a)^2-\sum_{\tau\in[t],\theta\in\Theta}y_{\tau,\theta}-\sum_{r\in\mathcal{R},k\in\mathcal{K}_r,\theta\in\Theta}\beta_{r,k,\theta}\right\}\quad\text{s.t.} \tag{13a}$$

$$\mu_\theta f_\theta^{\mathsf{s}}(R)+\sum_{r\in R}x_{r,\theta,\tau}+y_{\tau,\theta}\le 0,\quad\forall\tau\in[t],\theta\in\Theta,R\subseteq\mathcal{R} \tag{13b}$$

$$\frac{\varphi_\theta^{r,k}(a)}{\alpha}\le-\mathbb{I}_{a=a_1}\sum_{\substack{\tau\in[t]:\\ k=k_{\tau,r}}}x_{r,\theta,\tau}+\left(\sum_{k'\in\mathcal{K}_r}z_{r,k,k'}\right)\mu_\theta u_k^r(a,\theta)-\sum_{\substack{a'\in A,\\ k'\in\mathcal{K}_r}}\alpha_{r,k',k,a,a'}\mu_\theta u_{k'}^r(a',\theta)+\beta_{r,k,\theta,a},$$
$$\forall r\in\mathcal{R},k\in\mathcal{K}_r,\theta\in\Theta,a\in A \tag{13c}$$

$$- z_{r,k,k'} + \sum_{a' \in A} \alpha_{r,k,k',a,a'} = 0, \quad \forall r \in \mathcal{R}, k, k' \in \mathcal{K}_r, \forall a \in A \tag{13d}$$

where the constraint of Equation (10) becomes the constraint of Equation (13b) since the dual variable γ is positive. Similarly, the constraint of Equation (11) becomes the constraint of Equation (13c) as the dual variable η is positive.

We now remark that the above dual problem can be solve in polynomial time, when we have access to the optimization oracle $\mathcal{O}$.

Problem 13 is convex. Hence, we can solve it applying the ellipsoid method. The separation over Constraint (13c) can be done in polynomial-time since there are polynomially-many constraints. Moreover, the separation problem relative to the objective can be solved in polynomial time since there are polynomially-many variables and the objective is convex. Finally, the separation over the constraint of Equation (13b) must solve

$$\arg\max_{R} \left\{ \mu_\theta f^{\mathsf{s}}_\theta(R) + \sum_{r \in \mathcal{R}} x_{r,\theta,\tau} \right\},$$

for each possible $\tau \in [t]$ and $\theta \in \Theta$, which can be done by exploiting the optimization oracle $\mathcal{O}(f^{\mathsf{s}}_\theta, x_{\theta,\tau}/\mu_\theta)$ for all $\tau \in [t]$ and $\theta \in \Theta$.

If any of these solution are greater than $-y_{\tau,\theta}$, we return the relative constraint, otherwise all the constraints (13c) are satisfied. Hence, the ellipsoid method runs in polynomial-time and find an arbitrary good approximation. For the easy of exposition, we ignore the arbitrary small approximation error of the ellipsoid method. □

Theorem 5.5. *In settings in which receivers have binary actions, and the sender has a monotone supermodular or a monotone anonymous utility function, Algorithm 2 has polynomial per-iteration running time and guarantees*

$$R_T \leq nd|A|\sqrt{mT}.$$

Proof. Since, by Lemma 5.3, there exists a polynomial-time oracle $\mathcal{O}$, applying Lemma 5.4 and 5.5 we can compute Line 4 and 5 of Algorithm 2 in polynomial-time. Moreover, it is easy to see that all the other operations of the algorithm can be executed in polynomial time. □