A. Proofs Omitted from Section 4

In this section, we provide the complete proof of the hardness result in Theorem 1. This is based on a reduction from the promise-problem version of LABEL-COVER, which we define next.

The following is the formal definition of an instance of the LABEL-COVER problem.

Definition 3 (LABEL-COVER instance)**.** *An instance of* LABEL-COVER *consists of a tuple* (G, Σ, Π)*, where:*

- $G := (U, V, E)$ *is a* bipartite graph *defined by two disjoint sets of nodes* U *and* V*, connected by the edges in* $E \subseteq U \times V$*, which are such that all the nodes in* U *have the same degree;*
- Σ *is a finite set of* labels*; and*
- $\Pi := \{\Pi_e : \Sigma \to \Sigma \mid e \in E\}$ *is a finite set of* edge constraints.

Definition 4 (Labeling)**.** *Given an instance* (G, Σ, Π) *of* LABEL-COVER*, a* labeling *of the graph* G *is a mapping* $\pi : U \cup V \to \Sigma$ *that assigns a label to each vertex of* G *such that all the edge constraints are satisfied. Formally, a labeling* π *satisfies the constraint for an edge* $e = (u, v) \in E$ *if* $\pi(v) = \Pi_e(\pi(u))$.

The classical LABEL-COVER problem is the search problem of finding a valid labeling for a LABEL-COVER instance given as input. In the following, we consider a different version of the problem, which is the *promise problem* associated with LABEL-COVER instances, defined as follows.

Definition 5 (GAP-LABEL-COVER$_{c,b}$)**.** *For any pair of numbers* $0 < b < c < 1$*, we define* GAP-LABEL-COVER$_{c,b}$ *as the following promise problem.*

- `Input`: *An instance* (G, Σ, Π) *of* LABEL-COVER *such that either one of the following is true:*
  - *there exists a labeling* $\pi : U \cup V \to \Sigma$ *that satisfies at least a fraction* c *of the edge constraints in* Π*;*
  - *any labeling* $\pi : U \cup V \to \Sigma$ *satisfies less than a fraction* b *of the edge constraints in* Π.
- `Output`: *Determine which of the above two cases hold.*

In order to prove Theorem 1, we make use of the following result due to Raz (1998) and Arora et al. (1998).

Theorem 8 (Raz (1998); Arora et al. (1998))**.** *For any* $\epsilon > 0$*, there exists a constant* $k_\epsilon \in \mathbb{N}$ *that depends on* ϵ *such that the promise problem* GAP-LABEL-COVER$_{1,\epsilon}$ *restricted to inputs* (G, Σ, Π) *with* $|\Sigma| = k_\epsilon$ *is* NP*-hard.*

Next, we provide the complete proof of Theorem 1.

Theorem 1. *For every* $0 < \alpha \leq 1$*, it is* NP*-hard to compute an* α*-approximate solution to* BAYESIAN-OPT-SIGNAL*, even when the sender's utility is such that, for every* $\theta \in \Theta$*,* $f_\theta(R) = 1$ *iff* $|R| \geq 2$*, while* $f_\theta(R) = 0$ *otherwise.*

Proof. We provide a reduction from GAP-LABEL-COVER$_{1,\epsilon}$. Our reduction maps an instance (G, Σ, Π) of LABEL-COVER to an instance of BAYESIAN-OPT-SIGNAL with the following properties:

- (*completeness*) if the LABEL-COVER instance admits a labeling satisfying all the edge constraints (recall $c = 1$), then the BAYESIAN-OPT-SIGNAL instance has a signaling scheme with sender's expected utility $\geq \left(1 - \frac{\epsilon}{|\Sigma|}\right) \frac{1}{|\Sigma|} \geq \frac{1}{2|\Sigma|}$;
- (*soundness*) if the LABEL-COVER instance is such that any labeling satisfies at most a fraction ϵ of the edge constraints, then an optimal signaling scheme in the BAYESIAN-OPT-SIGNAL instance has sender's expected utility at most $\frac{2\epsilon}{|\Sigma|}$.

By Theorem 8, for any $\epsilon > 0$ there exists a constant $k_\epsilon \in \mathbb{N}$ that depends on ϵ such that GAP-LABEL-COVER$_{1,\epsilon}$ restricted to inputs (G, Σ, Π) with $|\Sigma| = k_\epsilon$ is NP-hard. Given $0 < \alpha \leq 1$, by setting $\epsilon = \frac{\alpha}{4}$ and noticing that $\frac{2\epsilon/|\Sigma|}{1/2|\Sigma|} = 4\epsilon = \alpha$, we can conclude that it is NP-hard to compute an α-approximate solution to BAYESIAN-OPT-SIGNAL.

Construction Given an instance (G, Σ, Π) of LABEL-COVER defined over a bipartite graph $G := (U, V, E)$, we build an instance of BAYESIAN-OPT-SIGNAL as follows.

- For each label $\sigma \in \Sigma$, there is a corresponding state of nature $\theta_\sigma \in \Theta$. Moreover, there is an additional state $\theta_0 \in \Theta$. Thus, the total number of possible states is $d = |\Sigma| + 1$.
- The prior distribution is $\boldsymbol{\mu} \in \text{int}(\Delta_\Theta)$ such that $\mu_{\theta_\sigma} = \frac{\epsilon}{|\Sigma|^2}$ for every $\theta_\sigma \in \Theta$ and $\mu_{\theta_0} = 1 - \frac{\epsilon}{|\Sigma|}$.
- For every vertex $v \in U \cup V$ of the graph G, there is a receiver $r_v \in \mathcal{R}$. Thus, $n = |U \cup V|$.
- Each receiver $r_v \in \mathcal{R}$ has $m_{r_v} = |\Sigma| + 1$ possible types. The set of types of receiver r_v is $\mathcal{K}_{r_v} = \{k_\sigma \mid \sigma \in \Sigma\} \cup \{k_0\}$.
- A receiver $r_v \in \mathcal{R}$ of type $k_\sigma \in \mathcal{K}_{r_v}$ has utility such that $u_{\theta_\sigma}^{r_v, k_\sigma} = \frac{1}{2}$ and $u_{\theta_{\sigma'}}^{r_v, k_\sigma} = -1$ for all $\theta_{\sigma'} \in \Theta : \theta_{\sigma'} \neq \theta_\sigma$, while $u_{\theta_0}^{r_v, k_\sigma} = -\frac{\epsilon}{2|\Sigma|^2}$. Moreover, a receiver $r_v \in \mathcal{R}$ of type k_0 has utility such that $u_\theta^{r_v, k_0} = -1$ for all $\theta \in \Theta$.
- The sender's utility is such that, for every $\theta \in \Theta$, the function $f_\theta : 2^{\mathcal{R}} \to [0, 1]$ satisfies $f_\theta(R) = 1$ if and only if $R \subseteq \mathcal{R} : |R| \geq 2$, while $f_\theta(R) = 0$ otherwise.
- The subset $K \subseteq \mathcal{K}$ of type profiles that can occur with positive probability is $K := \{\boldsymbol{k}^{uv,\sigma} \mid e = (u, v) \in E, \sigma \in \Sigma\}$, where, for every edge $e = (u, v) \in E$ and label $\sigma \in \Sigma$, the type profile $\boldsymbol{k}^{uv,\sigma} \in \mathcal{K}$ is such that $k_{r_u}^{uv,\sigma} = k_\sigma$, $k_{r_v}^{uv,\sigma} = k_{\sigma'}$ with $\sigma' = \Pi_e(u)$, and $k_{r_{v'}}^{uv,\sigma} = k_0$ for every $r_{v'} \in \mathcal{R} : r_{v'} \notin \{r_u, r_v\}$.
- The probability distribution $\boldsymbol{\lambda} \in \text{int}(\Delta_K)$ is such that $\lambda_{\boldsymbol{k}} = \frac{1}{|E||\Sigma|}$ for every $\boldsymbol{k} \in K$.

Notice that, in the BAYESIAN-OPT-SIGNAL instances used for the reduction, the sender's payoff is 1 if and only if at least two receivers play action a_1, while it is 0 otherwise. Let us also recall that direct signals for a receiver $r_v \in \mathcal{R}$ are defined by the set $\mathcal{S}_{r_v} := 2^{\mathcal{K}_{r_v}}$, with a signal being represented as the set of receiver's types that are recommended to play action a_1.

Completeness Let $\pi : U \cup V \to \Sigma$ be a labeling of the graph G that satisfies all the edge constraints. We define a corresponding direct signaling scheme $\phi : \Theta \to \Delta_{\mathcal{S}}$ as follows. For any label $\sigma \in \Sigma$, let $\boldsymbol{s}^\sigma \in \mathcal{S}$ be a signal profile such that the signal sent to receiver $r_v \in \mathcal{R}$ is $s_{r_v}^\sigma = \{k_\sigma\}$, *i.e.*, only a receiver of the type k_σ is told to play a_1, while all the other types are recommended to play a_0. Moreover, let $\boldsymbol{s}^\pi \in \mathcal{S}$ be a signal profile in which the signal sent to receiver $r_v \in \mathcal{R}$ is $s_{r_v}^\pi = \{k_\sigma\}$ with $\sigma \in \Sigma : \sigma = \pi(v)$, *i.e.*, each receiver r_v is told to play action a_1 only if her/his type is k_σ for the label σ assigned to vertex v by the labeling π, otherwise she/he is recommended to play a_0. Then, we define $\phi_{\theta_\sigma}(\boldsymbol{s}^\sigma) = 1$ for every state of nature $\theta_\sigma \in \Theta$, while $\phi_{\theta_0}(\boldsymbol{s}^\pi) = 1$. Notice that the signaling scheme ϕ is deterministic, since each state of nature is mapped to only one signal profile (with probability one). As a first step, we prove that the signaling scheme ϕ is *persuasive*. Let us fix a receiver $r_v \in \mathcal{R}$. After receiving a signal $s = \{k_\sigma\} \in \mathcal{S}_{r_v}$ with $\sigma \in \Sigma : \sigma \neq \pi(v)$, by definition of ϕ, the receiver's posterior belief is such that state of nature θ_σ is assigned probability one. Thus, if the receiver has type k_σ, then she/he is incentivized to play action a_1, since $u_{\theta_\sigma}^{r_v,k_\sigma} = \frac{1}{2} > 0$ (recall that $u_{\theta_\sigma}^{r_v,k_\sigma}$ is the utility different "action a_1 minus action a_0" when the state is θ_σ). Instead, if the receiver has type $k \in \mathcal{K}_{r_v} : k \neq k_\sigma$, then she/he is incentivized to play action a_0, since either $k = k_0$ and $u_{\theta_\sigma}^{r_v,k_0} = -1 < 0$ or $k = k_{\sigma'}$ with $\sigma' \in \Sigma : \sigma' \neq \sigma$ and $u_{\theta_\sigma}^{r_v,k_{\sigma'}} = -1 < 0$. After receiving a signal $s = \{k_\sigma\} \in \mathcal{S}_{r_v}$ with $\sigma = \pi(v)$, the receiver's posterior belief is such that the states of nature θ_σ and θ_0 are assigned probabilities proportional to their corresponding prior probabilities, respectively μ_{θ_σ} and μ_{θ_0} (she/he cannot tell whether $\boldsymbol{s}^\sigma$ or $\boldsymbol{s}^\pi$ has been selected by the sender). Thus, if the receiver has type k_σ, then she/he is incentivized to play action a_1, since her expected utility difference "action a_1 minus action a_0" is the following:

$$\frac{\mu_{\theta_\sigma}}{\mu_{\theta_\sigma} + \mu_{\theta_0}} u_{\theta_\sigma}^{r_v,k_\sigma} + \frac{\mu_{\theta_0}}{\mu_{\theta_\sigma} + \mu_{\theta_0}} u_{\theta_0}^{r_v,k_\sigma} = \frac{1}{\mu_{\theta_\sigma} + \mu_{\theta_0}} \left[\frac{\epsilon}{|\Sigma|^2} \frac{1}{2} - \left(1 - \frac{\epsilon}{|\Sigma|}\right) \frac{\epsilon}{2|\Sigma|^2} \right] > \frac{1}{\mu_{\theta_\sigma} + \mu_{\theta_0}} \left[\frac{\epsilon}{2|\Sigma|^2} - \frac{\epsilon}{2|\Sigma|^2} \right] = 0.$$

If the receiver has a type different from k_σ, simple arguments show that the expected utility difference is negative, incentivizing action a_0. This proves that the signaling scheme ϕ is persuasive. Next, we bound the sender's expected utility in ϕ. Notice that, when the state of nature is θ_0, if the receivers' type profile is $\boldsymbol{k}^{uv,\sigma} \in K$ with $\sigma = \pi(u)$ for some edge $e = (u, v) \in E$, then both receivers r_u and r_v play action a_1. This is readily proved since $k_{r_u}^{uv,\sigma} = k_\sigma$ and $k_{r_v}^{uv,\sigma} = k_{\sigma'}$ with $\sigma = \pi(u)$ and $\sigma' = \pi(v)$ (recall that $\pi(v) = \Pi_e(u)$ as ϕ satisfies all the edge constraints), and, thus, both r_u and r_v are recommended to play a_1 when the state is θ_0. As a result, under signaling scheme ϕ, when the receivers' type profile is $\boldsymbol{k}^{uv,\sigma} \in K$, then the sender's resulting payoff is one (recall the definition of functions f_θ). By recalling that each type profile $\boldsymbol{k}^{uv,\sigma} \in K$ with $\sigma = \pi(u)$ (for each edge $e = (u, v) \in E$) occurs with probability $\lambda_{\boldsymbol{k}^{uv,\sigma}} = \frac{1}{|E||\Sigma|}$, we can lower bound the sender's expected utility (see the objective of Problem (1)) as follows:

$$\sum_{\boldsymbol{k} \in K} \lambda_{\boldsymbol{k}} \sum_{\theta \in \Theta} \mu_\theta \sum_{\boldsymbol{s} \in \mathcal{S}} \phi_\theta(\boldsymbol{s}) f_\theta \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) \geq \mu_{\theta_0} \sum_{\boldsymbol{k}^{uv,\sigma} \in K : \sigma = \pi(u)} \lambda_{\boldsymbol{k}^{uv,\sigma}} = \mu_{\theta_0} \frac{1}{|\Sigma|} = \left(1 - \frac{\epsilon}{|\Sigma|}\right) \frac{1}{|\Sigma|}.$$

Soundness By contradiction, suppose that there exists a direct and persuasive signaling scheme $\phi : \Theta \to \Delta_{\mathcal{S}}$ that provides the sender with an expected utility greater than $\frac{2\epsilon}{|\Sigma|}$. Since the sender can extract an expected utility at most of $\frac{\epsilon}{|\Sigma|}$ from states of nature $\theta \in \Theta$ with $\theta \neq \theta_0$ (as $\sum_{\theta \in \Theta : \theta \neq \theta_0} \mu_\theta = \frac{\epsilon}{|\Sigma|}$ and the maximum value of functions f_θ is one), then it must be the case that the expected utility contribution due to state θ_0 is greater than $\frac{\epsilon}{|\Sigma|}$. Let us consider the distribution over signal profiles $\phi_{\theta_0} \in \Delta_{\mathcal{S}}$ induced by state of nature θ_0. We prove that, for each signal profile $\boldsymbol{s} \in \mathcal{S}$ such that $\phi_{\theta_0}(\boldsymbol{s}) > 0$ and each receiver $r_v \in \mathcal{R}$, it must hold that $|s_r| \leq 1$, *i.e.*, at most one type of receiver r_v is recommended to play a_1. First, notice that a receiver of type k_0 cannot be incentivized to play a_1, since $u_\theta^{r_v,k_0} = -1$ for all $\theta \in \Theta$. By contradiction, suppose that there are two receiver's types $k_\sigma, k_{\sigma'} \in \mathcal{K}_{r_v}$ with $k_\sigma \neq k_{\sigma'}$ such that $k_\sigma, k_{\sigma'} \in s_r$ (*i.e.*, they are both recommended to play a_1). By letting $\boldsymbol{\xi} \in \Delta_\Theta$ be the posterior belief of receiver r_v induced by s_r, for type k_σ it must be the case that:

$$\xi_{\theta_\sigma} u_{\theta_\sigma}^{r_v,k_\sigma} + \sum_{\theta_{\sigma''} \in \Theta : \theta_{\sigma''} \neq \theta_\sigma} \xi_{\theta_{\sigma''}} u_{\theta_{\sigma''}}^{r_v,k_\sigma} + \xi_{\theta_0} u_{\theta_0}^{r_v,k_\sigma} = \frac{1}{2} \xi_{\theta_\sigma} - \sum_{\theta_{\sigma''} \in \Theta : \theta_{\sigma''} \neq \theta_\sigma} \xi_{\theta_{\sigma''}} - \frac{\epsilon}{2|\Sigma|^2} \xi_{\theta_0} > 0,$$

since the signaling scheme is persuasive, and, thus, a receiver of type k_σ must be incentivized to play action a_1. This implies that $\xi_{\theta_\sigma} > 2 \sum_{\theta_{\sigma''} \in \Theta : \theta_{\sigma''} \neq \theta_\sigma} \xi_{\theta_{\sigma''}} \geq 2 \xi_{\theta_{\sigma'}}$. Analogous arguments for type $k_{\sigma'}$ imply that $\xi_{\theta_{\sigma'}} > 2 \xi_{\theta_\sigma}$, reaching a contradiction. This shows that, for each $\boldsymbol{s} \in \mathcal{S}$ such that $\phi_{\theta_0}(\boldsymbol{s}) > 0$ and each $r_v \in \mathcal{R}$, it must be the case that $|s_r| \leq 1$. Next, we provide the last contradiction proving the result. Let us recall that, by assumption, the sender's expected utility contribution due to θ_0 is $\sum_{\boldsymbol{k} \in K} \lambda_{\boldsymbol{k}} \sum_{\boldsymbol{s} \in \mathcal{S}} \phi_{\theta_0}(\boldsymbol{s}) f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) \geq \frac{\epsilon}{|\Sigma|}$. By an averaging argument, this implies that there must exist a signal profile $\boldsymbol{s} \in \mathcal{S}$ such that $\phi_{\theta_0}(\boldsymbol{s}) > 0$ and $\sum_{\boldsymbol{k} \in K} \lambda_{\boldsymbol{k}} f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) \geq \frac{\epsilon}{|\Sigma|}$. Let $\boldsymbol{s} \in \mathcal{S}$ be such signal profile. Let us define a corresponding labeling $\pi : U \cup V \to \Sigma$ of the graph G such that, for every vertex $v \in U \cup V$, it holds $\pi(v) = \sigma$, where $\sigma \in \Sigma$ is the label corresponding to the unique type k_σ of receiver r_v that is recommended to play action a_1 under $\boldsymbol{s}$ (if any, otherwise any label is fine). Since $\sum_{\boldsymbol{k} \in K} \lambda_{\boldsymbol{k}} f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) \geq \frac{\epsilon}{|\Sigma|}$ and it holds $\lambda_{\boldsymbol{k}} = \frac{1}{|E||\Sigma|}$ and $f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) \in \{0, 1\}$ for every $\boldsymbol{k} \in K$, it must be the case that there are at least $\epsilon|E|$ type profiles $\boldsymbol{k} \in K$ such that $f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) = 1$. Since a receiver of type k_0 cannot be incentivized to play action a_1, the value of $f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right)$ can be one only if there are at least two receivers with types different from k_0 that play action a_1. Thus, it must hold that $f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right) = 0$ for all the type profiles $\boldsymbol{k}^{uv,\sigma} \in K$ such that $\sigma \neq \pi(u)$ (as $k_{r_u}^{uv,\sigma}$ would be equal to k_σ with $\sigma \neq \pi(u)$ and $k_\sigma \notin s_{r_u}$). For the type profiles $\boldsymbol{k}^{uv,\sigma} \in K$ such that $\sigma = \pi(u)$ (one per edge $e = (u, v) \in E$ of the graph G), the value of $f_{\theta_0} \left(R_{\boldsymbol{s}}^{\boldsymbol{k}}\right)$ is one if and only if $\pi(v) = \Pi_e(u)$, so that both receivers r_u and r_v are told to play action a_1. As a result, this implies that there must be at least $\epsilon|E|$ edges $e \in E$ for which the labeling π satisfies the corresponding edge constraint Π_e, which is a contradiction. □

B. Proofs Omitted from Section 5

Theorem 3. *Given an oracle φ_α (as in Definition 1) for some $0 < \alpha \leq 1$, a learning rate $\eta \in (0, 1]$, and an approximation error $\epsilon \in [0, 1]$, Algorithm 1 has α-regret*

$$R_\alpha^T \leq \frac{|E^T|}{2\eta} + \frac{\eta T}{2} + \frac{\epsilon T}{2\eta},$$

with a per-iteration running time $\mathsf{poly}(t)$.

Proof. First, we bound the per-iteration running time of Algorithm 1. For any $t \in [T]$, we have $E^t = \bigcup_{t' \in [t]} e^{t'}$, which represents the set of feedbacks observed up to iteration t. Thus, it holds $|E^t| \leq t$. At iteration $t \in [T]$, the algorithm works with vectors $\boldsymbol{x}^t$ and $\boldsymbol{y}^{t+1}$. The first one belongs to $\mathcal{X}_{E^{t-1}}$ (as it is returned by φ_α at iteration $t-1$), and, thus, it has at most $t-1$ non-zero components. Similarly, since $\boldsymbol{y}^{t+1} = \boldsymbol{x}^t + \eta 1_{e^t}$, it holds that $\boldsymbol{y}^{t+1} \in [0, 2]^p$ and $y_e^{t+1} = 0$ for all $e \notin E^t$, which implies that $\boldsymbol{y}^{t+1}$ has at most t non-zero components. As a result, we can sparsely represent vectors $\boldsymbol{x}^t$ and $\boldsymbol{y}^{t+1}$ so that Algorithm 1 has a per-iteration running time bounded by t for any iteration $t \in [T]$, independently of the actual size p of the vectors. Moreover, notice that $\boldsymbol{y}^{t+1}$ satisfies the conditions required by the inputs of the oracle φ_α.

Next, we bound the α-regret of Algorithm 1. For the ease of notation, in the following, for any vector $\boldsymbol{x} \in \mathcal{X}$ and subset $E \subseteq \mathcal{E}$, we let $\boldsymbol{x}_E := \tau_E(\boldsymbol{x})$. Moreover, for any $t \in [T]$, we let $\mathbb{I}_t := \mathbb{I}\left\{e^t \notin E^{t-1}\right\}$, which is the indicator function that is equal to 1 if and only if $e^t \notin E^{t-1}$, *i.e.*, when the feedback e^t at iteration t has never been observed before. Fix $\boldsymbol{x} \in \alpha\mathcal{X}$. Then, the following relations hold:

$$
\begin{aligned}
\left|\left|\boldsymbol{x}_{E^t} - \boldsymbol{x}^{t+1}\right|\right|^2 &\leq \left|\left|\boldsymbol{x}_{E^t} - \boldsymbol{y}^{t+1}\right|\right|^2 + \epsilon && \text{(4a)}\\
&= \left|\left|\boldsymbol{x}_{E^t} - \boldsymbol{x}^t - \eta \mathbf{1}_{e^t}\right|\right|^2 + \epsilon && \text{(4b)}\\
&= \left|\left|\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t - \eta \mathbf{1}_{e_t}\right|\right|^2 + \epsilon && \text{(4c)}\\
&= \left|\left|\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t\right|\right|^2 + \eta^2 - 2\eta \mathbf{1}_{e^t}^\top\left(\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t\right) + \epsilon && \text{(4d)}\\
&= \left|\left|\boldsymbol{x}_{E^{t-1}} - \boldsymbol{x}^t_{E^{t-1}}\right|\right|^2 + \mathbb{I}_t \left|x_{e^t} - x^t_{e^t}\right|^2 + \eta^2 - 2\eta \mathbf{1}_{e^t}^\top\left(\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t\right) + \epsilon && \text{(4e)}\\
&\leq \left|\left|\boldsymbol{x}_{E^{t-1}} - \boldsymbol{x}^t_{E^{t-1}}\right|\right|^2 + \mathbb{I}_t + \eta^2 - 2\eta \mathbf{1}_{e^t}^\top\left(\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t\right) + \epsilon. && \text{(4f)}
\end{aligned}
$$

Notice that Equation (4b) holds by definition of φ_α since $\boldsymbol{x}_{E^t} \in \alpha\mathcal{X}_{E^t}$, Equation (4d) follows from $\boldsymbol{x}_{E^t} = \boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t}$, while Equation (4e) can be derived by decomposing the first squared norm in the preceding expression. By using the last relation above, we can write the following:

$$
\begin{aligned}
\sum_{t\in[T]} \mathbf{1}_{e^t}^\top\left(\boldsymbol{x} - \boldsymbol{x}^t\right) &= \sum_{t\in[T]} \mathbf{1}_{e^t}^\top\left(\boldsymbol{x}_{E^{t-1}} + \mathbb{I}_t\, x_{e^t}\mathbf{1}_{e^t} - \boldsymbol{x}^t\right) && \text{(5a)}\\
&\leq \frac{1}{2\eta}\sum_{t\in[T]}\left(\left|\left|\boldsymbol{x}_{E^{t-1}} - \boldsymbol{x}^t_{E^{t-1}}\right|\right|^2 - \left|\left|\boldsymbol{x}_{E^t} - \boldsymbol{x}^{t+1}\right|\right|^2 + \mathbb{I}_t + \eta^2 + \epsilon\right) && \text{(5b)}\\
&= \frac{1}{2\eta}\sum_{t\in[T]}\left(\mathbb{I}_t + \eta^2 + \epsilon\right) && \text{(5c)}\\
&= \frac{1}{2\eta}\left(|E^T| + T\eta^2 + T\epsilon\right), && \text{(5d)}
\end{aligned}
$$

where Equation (5c) is obtained by telescoping the sum. Then, the following concludes the proof:

$$
\begin{aligned}
R^T_\alpha &:= \alpha \max_{y\in\mathcal{Y}} \sum_{t\in[T]} u(y, e^t) - \sum_{t\in[T]} u(y^t, e^t) \leq \alpha \max_{\boldsymbol{x}\in\mathcal{X}} \sum_{t\in[T]} x_{e^t} - \sum_{t\in[T]} x^t_{e^t} = \alpha \max_{\boldsymbol{x}\in\mathcal{X}} \sum_{t\in[T]} \mathbf{1}_{e^t}^\top\left(\boldsymbol{x} - \boldsymbol{x}^t\right)\\
&= \max_{\boldsymbol{x}\in\alpha\mathcal{X}} \sum_{t\in[T]} \mathbf{1}_{e^t}^\top\left(\boldsymbol{x} - \boldsymbol{x}^t\right) \leq \frac{1}{2\eta}\left(|E^T| + T\eta^2 + T\epsilon\right).
\end{aligned}
$$

□

C. Proofs Omitted from Section 6.1

Theorem 4. *Given $\epsilon \in \mathbb{R}_+$ and an approximate separation oracle $\mathcal{O}_\alpha$, with $0 < \alpha \leq 1$, there exists a polynomial-time approximation algorithm for* BAYESIAN-OPT-SIGNAL *returning a signaling scheme with sender's utility at least α*OPT*$-\epsilon$, where* OPT *is the value of an optimal signaling scheme. Moreover, the algorithm works in time* $\mathsf{poly}(\frac{1}{\epsilon})$.

Proof of theorem 4. The dual problem of LP (1) reads as follows:

$$
\begin{aligned}
&\min_{\boldsymbol{z},\boldsymbol{d}} \quad \sum_{\theta\in\Theta} d_\theta && && \text{(6a)}\\
&\text{s.t. } \mu_\theta \sum_{r\in\mathcal{R}}\sum_{k\in s_r} u_\theta^{r,k} z_{r,s_r,k} + d_\theta \geq \mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) && \forall\theta\in\Theta, \forall \boldsymbol{s}\in\mathcal{S} && \text{(6b)}\\
&\quad z_{r,s,k} \leq 0 && \forall r\in\mathcal{R}, \forall s\in\mathcal{S}_r, \forall k\in\mathcal{K}_r : k\in s, && \text{(6c)}
\end{aligned}
$$

where $\boldsymbol{d} \in \mathbb{R}^{|\Theta|}$ is the vector of dual variable corresponding to the primal Constraints (1c), and $\boldsymbol{z} \in \mathbb{R}_{-}^{|\mathcal{R} \times \mathcal{S}_r \times \mathcal{K}_r|}$ is the vector of dual variable corresponding to Constraints (1b) in the primal. We rewrite the dual LP (6) so as to highlight the relation between an approximate separation oracle for Constraints (6b) and the oracle $\mathcal{O}_\alpha$. Specifically, we have

$$\min_{\boldsymbol{z} \geq 0, \boldsymbol{d}} \quad \sum_{\theta \in \Theta} d_\theta \tag{7a}$$

$$\text{s.t. } d_\theta \geq \mu_\theta \left(\sum_{\boldsymbol{k} \in \mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \sum_{r \in \mathcal{R}} \sum_{k \in s_r} u_\theta^{r,k} z_{r,s_r,k} \right) \qquad \forall \theta \in \Theta, \forall \boldsymbol{s} \in \mathcal{S}. \tag{7b}$$

Now, we show that it is possible to build a binary search scheme to find a value $\gamma^\star \in [0,1]$ such that the dual problem with objective $\gamma^\star$ is feasible, while the dual with objective $\gamma^\star - \beta$ is infeasible. The constant $\beta \geq 0$ will be specified later in the proof. The algorithm requires $\log(\beta)$ steps and works by determining, for a given value $\bar{\gamma} \in [0,1]$, whether there exists a feasible pair $(\boldsymbol{d}, \boldsymbol{z})$ for the following feasibility problem Ⓕ:

$$\text{Ⓕ} \begin{cases} \sum_{\theta \in \Theta} d_\theta \leq \bar{\gamma} \\ d_\theta \geq \mu_\theta \left(\sum_{\boldsymbol{k} \in \mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \sum_{r \in \mathcal{R}} \sum_{k \in s_r} u_\theta^{r,k} z_{r,s_r,k} \right) & \forall \theta \in \Theta, \forall \boldsymbol{s} \in \mathcal{S} \\ \boldsymbol{z} \geq 0. \end{cases}$$

At each iteration of the bisection algorithm, the feasibility problem Ⓕ is solved via the ellipsoid method. The algorithm is inizialized with $l = 0$, $h = 1$, and $\bar{\gamma} = \frac{1}{2}$. If Ⓕ is infeasible for $\bar{\gamma}$, the algorithm sets $l \leftarrow (l+h)/2$ and $\bar{\gamma} \leftarrow (h + \bar{\gamma})/2$. Otherwise, if Ⓕ is (approximately) feasible, it sets $h \leftarrow (l+h)/2$ and $\bar{\gamma} \leftarrow (l + \bar{\gamma})/2$. Then, the procedure is repeated with the updated value of $\bar{\gamma}$. The bisection procedure terminates when it determines a value $\gamma^\star$ such that Ⓕ is feasible for $\bar{\gamma} = \gamma^\star$, while it is infeasible for $\bar{\gamma} = \gamma^\star - \beta$. In the following, we present the approximate separation oracle which is employed at each iteration of the ellipsoid method.

Separation Oracle Given a point $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ in the dual space, and $\bar{\gamma} \in [0,1]$, we design an approximate separation oracle to determine if the point $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ is approximately feasible, or to determine a constraint of Ⓕ that is violated by such point. For each $\theta \in \Theta$, $r \in \mathcal{R}$, and $s \in \mathcal{S}_r$, let

$$w_{r,s}^\theta := \mu_\theta \sum_{k \in s} u_\theta^{r,k} \bar{z}_{r,s,k}.$$

When the magnitude of the weights $|w_{r,s}^\theta|$ is small, we show that it is enough to employ the optimization oracle $\mathcal{O}_\alpha$ in order to find a violated constraint, or to certify that all the constraints are approximately satisfied. On the other hand, when the weights $|w_{r,s}^\theta|$ are large (in particular, when the largest weight has exponential size in the size of the problem instance), the optimization oracle $\mathcal{O}_\alpha$ loses its polynomial time guarantees (see Definition 2). We show how to handle those specific settings in the following case analysis:

- Equation (7b) implies that $d_\theta \geq 0$ for each $\theta \in \Theta$. Then, if there exists a $\theta \in \Theta$ such that $\bar{d}_\theta < 0$, we return the violated constraint $(\theta, \varnothing)$ (that is, $d_\theta \geq 0$).
- If there exists $\theta \in \Theta$ such that $\bar{d}_\theta > 1$, then the first constraint of Ⓕ must be violated as $\bar{\gamma} \in [0,1]$.
- If there exists a receiver $r \in \mathcal{R}$ and a signal $s \in \mathcal{S}_r$ such that $w_{r,s}^\theta > 1$, then the constraint of Ⓕ corresponding to the pair (θ, s) is violated, because $d_\theta \leq 1$.
- If no violated constraint was found in the previous steps, we proceed by checking if there exists a state $\theta' \in \Theta$, a receiver $r' \in \mathcal{R}$, and a signal $s' \in \mathcal{S}_r$, such that $w_{r',s'}^{\theta'} \leq -|\mathcal{R}|$. If this is the case, we observe that for any pair $(\theta', \boldsymbol{s})$, with $\boldsymbol{s} \in \mathcal{S} : s_r = s'$, the corresponding constraint in Ⓕ reads

  $$\mu_\theta \sum_{\boldsymbol{k} \in \mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \sum_{r \in \mathcal{R} \setminus \{r'\}} w_{r,s_r}^{\theta'} + w_{r',s'}^{\theta'} \leq 0,$$

  since $\bar{\boldsymbol{d}} \geq 0$ if the current step is reached. For $w_{r',s'}^{\theta'} \leq -|\mathcal{R}|$ the above constraints are trivially satisfied, and therefore we can safely manage (for the current iteration of the ellipsoid method) any such constraint by setting $w_{r',s'}^{\theta'} = -|\mathcal{R}|$.

If none of the previous steps returned a violated constraint, we can safely assume that $0 \le d_\theta \le 1$ and $-|\mathcal{R}| \le w^\theta_{r,s} \le 1$, for each $\theta \in \Theta$, $r \in \mathcal{R}$, and $s \in \mathcal{S}_r$. Moreover, we observe that, by definition, for each $r \in \mathcal{R}$ and $\theta \in \Theta$, it holds $w^\theta_{r,\varnothing} = 0$. Since the magnitude of the weights is guaranteed to be small (that is, weights are guaranteed to be in the range $[-|\mathcal{R}|, 1]$), for each $\theta \in \Theta$ we can invoke $\mathcal{O}_\alpha(\theta, \mathcal{K}, \lambda, \boldsymbol{w}^\theta, \delta)$ to determine an $\boldsymbol{s}^\theta \in \mathcal{S}$ such that

$$\mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}^\theta}) + \sum_{r\in\mathcal{R}} w^\theta_{r,s^\theta_r} \ge \max_{\boldsymbol{s}\in\mathcal{S}} \left\{ \alpha\mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) + \sum_{r\in\mathcal{R}} w^\theta_{r,s_r} \right\} - \delta,$$

where δ is an approximation error that will be defined in the following. If at least one $\boldsymbol{s}^\theta$ is such that $(\theta, \boldsymbol{s}^\theta)$ is violated, we output that constraint, otherwise the algorithm returns that the LP is feasible.

Putting It All Together The bisection algorithm computes a $\gamma^\star \in [0, 1]$ and a pair $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ such that the approximate separation oracle does not find a violated constraint. The following lemma defines a modified LP and shows that $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ is a feasible solution for this problem and has value at most $\gamma^\star$.

Lemma 9. *The pair $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ is a feasible solution to the following LP and has value at most $\gamma^\star$:*

$$\begin{aligned}
&\min_{\boldsymbol{z}\ge 0, \boldsymbol{d}} \quad \sum_{\theta\in\Theta} d_\theta \\
&\text{s.t. } d_\theta \ge \alpha\mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) + \mu_\theta \sum_{r\in\mathcal{R}} \sum_{k\in s_r} u^{r,k}_\theta z_{r,s_r,k} - \delta \qquad \forall\theta\in\Theta, \forall \boldsymbol{s}\in\mathcal{S}.
\end{aligned}$$

Proof. The value is at most $\gamma^\star$ by assumption (that is, the separation oracle does not find a violated constraint for $(\bar{\boldsymbol{d}}, \bar{\boldsymbol{z}})$ in Ⓕ with objective $\gamma^\star$). Analogously, it holds that $\bar{d}_\theta \in [0, 1]$ for each $\theta \in \Theta$, and $w^\theta_{r,s} \le 1$ for each $r \in \mathcal{R}$, $s \in \mathcal{S}_r$, and $\theta \in \Theta$. Suppose, by contradiction, that $(\theta, \boldsymbol{s}')$ is a violated constraint of the modified LP above. Then, given $\bar{\boldsymbol{d}}$, oracle $\mathcal{O}_\alpha$ would have found an $\boldsymbol{s} \in \mathcal{S}$ such that

$$\mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f(R^{\boldsymbol{k}}_{\boldsymbol{s}}) + \mu_\theta \sum_{r\in\mathcal{R}} \sum_{k\in s_r} u^{r,k}_\theta \bar{z}_{r,s_r,k} \ge \alpha \sum_{\theta\in\Theta} \ \mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}'}) + \mu_\theta \sum_{r\in\mathcal{R}} \sum_{k\in s'_r} u^{r,k}_\theta \bar{z}_{r,s'_r,k} - \delta > \bar{d}_\theta,$$

where the first inequality follows by Definition 2, and the second from the assumption that the modified dual is infeasible. Hence, $\mathcal{O}_\alpha$ would return a violated constraint, reaching a contradiction. □

The dual problem of the LP of Lemma 9 reads as follows:

$$\begin{aligned}
&\max_{\phi} \sum_{\boldsymbol{s}\in\mathcal{S}} \sum_{\theta\in\Theta} \phi_\theta(\boldsymbol{s}) \left(\alpha\mu_\theta \sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}}\, f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) - \delta \right) \\
&\text{s.t. } \sum_{\theta\in\Theta} \mu_\theta \sum_{\boldsymbol{s}: s_r = s'} \phi_\theta(\boldsymbol{s})\, u^{r,k}_\theta \ge 0 \qquad \forall r \in R, \forall s' \in \mathcal{S}_r, \forall k \in \mathcal{K}_r : k \in s' \\
&\qquad \sum_{\boldsymbol{s}\in\mathcal{S}} \phi_\theta(\boldsymbol{s}) = 1 \qquad \forall \theta\in\Theta \\
&\qquad \phi_\theta(\boldsymbol{s}) \ge 0 \qquad \forall\theta\in\Theta, \boldsymbol{s}\in\mathcal{S}.
\end{aligned}$$

By strong duality, Lemma 9 implies that the value of the above problem is at most $\gamma^\star$. Then, let OPT be value of the optimal solution to LP (1). The same solution is feasible for the LP we just described, where it has value

$$\alpha\text{OPT} - |\Theta|\delta \le \gamma^\star. \tag{10}$$

Now, we show how to find a solution to the original problem (LP (1)) with value at least $\gamma^\star - \beta$. Let $\mathcal{H}$ be the set of constraints returned by the ellipsoid method run on the feasibility problem Ⓕ with objective $\gamma^\star - \beta$.

Lemma 10. *LP (1) with variables restricted to those corresponding to dual constraints $\mathcal{H}$ returns a signaling scheme with value at least $\gamma^\star - \beta$. Moreover, the solution can be determined in polynomial time.*

Proof. By construction of the bisection algorithm, Ⓕ is infeasible for value $\gamma^\star - \beta$. Hence, the following LP has value at least $\gamma^\star - \beta$:

$$\min_{\boldsymbol{z}\geq 0, \boldsymbol{d}} \quad \sum_{\theta\in\Theta} d_\theta$$
$$\text{s.t.} \quad d_\theta \geq \mu_\theta \left(\sum_{\boldsymbol{k}\in\mathcal{K}} \lambda_{\boldsymbol{k}} f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \sum_{r\in\mathcal{R}} \sum_{k\in s_r} u_\theta^{r,k} z_{r,s_r,k} \right) \qquad \forall (\theta, \boldsymbol{s}) \in \mathcal{H}.$$

Notice that the primal of the above LP is exactly LP (1) with variables restricted to those corresponding to dual constraints in $\mathcal{H}$, and that the former (restricted) LP has value at least $\gamma^\star - \beta$ by strong duality. To conclude the proof, the ellipsoid method guarantees that $\mathcal{H}$ is of polynomial size. Hence, the LP can be solved in polynomial time. □

Let APX be the value of an optimal solution to LP (1) restricted to variables corresponding to dual constraints in $\mathcal{H}$. Then,

$$\begin{aligned} \text{APX} &\geq \gamma^\star - \beta \\ &\geq \alpha \text{OPT} - |\Theta|\delta - \beta \\ &\geq \alpha \text{OPT} - \epsilon, \end{aligned}$$

where the first inequality holds by Lemma 10, the second inequality follows from Equation (10), and the last inequality is obtained by setting $\delta = \frac{\epsilon}{2|\Theta|}$ and $\beta = \frac{\epsilon}{2}$. □

D. Proofs Omitted from Section 6.2

Theorem 5. *Given a subset $K \subseteq \mathcal{K}$, a vector $\boldsymbol{y} \in [0,2]^{|\mathcal{K}|}$ such that $y_{\boldsymbol{k}} = 0$ for all $\boldsymbol{k} \notin K$, and an approximation error $\epsilon \in \mathbb{R}_+$, for any $0 < \alpha \leq 1$, the approximate projection oracle $\varphi_\alpha(K, \boldsymbol{y}, \epsilon)$ can be computed in polynomial time by querying the approximate separation oracle $\mathcal{O}_\alpha$.*

Proof. The problem of computing the projection of point $\boldsymbol{y}$ on $\mathcal{X}_K$ (see Equation (2)) can be formulated via the following convex programming problem, which we denote by Ⓟ:

$$\text{Ⓟ} \begin{cases} \min\limits_{\phi, \boldsymbol{x}} \displaystyle\sum_{\boldsymbol{k}\in K} (x_{\boldsymbol{k}} - y_{\boldsymbol{k}})^2 & \\ \text{s.t.} \displaystyle\sum_{\theta\in\Theta} \mu_\theta \left(\sum_{\substack{\boldsymbol{s}\in\mathcal{S}: \\ s_r = s'}} \phi_\theta(\boldsymbol{s}) u_\theta^{r,k} \right) \geq 0 & \forall r \in \mathcal{R}, \forall s' \in \mathcal{S}_r, \forall k \in \mathcal{K}_r : k \in s' \\ \displaystyle\sum_{\boldsymbol{s}\in\mathcal{S}} \phi_\theta(\boldsymbol{s}) = 1 & \forall \theta \in \Theta \\ \phi_\theta(\boldsymbol{s}) \geq 0 & \forall \theta \in \Theta, \forall \boldsymbol{s} \in \mathcal{S} \\ x_{\boldsymbol{k}} \leq \displaystyle\sum_{\theta\in\Theta} \sum_{\boldsymbol{s}\in\mathcal{S}} \mu_\theta \, \phi_\theta(\boldsymbol{s}) f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) & \forall \boldsymbol{k} \in K. \end{cases}$$

Then, we compute the Lagrangian of Ⓟ by introducing dual variables $z_{r,s,k} \leq 0$ for each $r \in \mathcal{R}$, $s \in \mathcal{S}_r$, and $k \in s$, $d_\theta \in \mathbb{R}$ for each $\theta \in \Theta$, $v_{\theta,\boldsymbol{s}} \leq 0$ for each $\theta \in \Theta$, $\boldsymbol{s} \in \mathcal{S}$, and $\nu_{\boldsymbol{k}} \geq 0$ for each $\boldsymbol{k} \in K$. Specifically, the Lagrangian of Ⓟ reads as follows

$$\begin{aligned} L(\phi, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v}, \boldsymbol{\nu}, \boldsymbol{d}) := \sum_{\boldsymbol{k}\in K} (x_{\boldsymbol{k}} - y_{\boldsymbol{k}})^2 + \sum_{r\in\mathcal{R}} \sum_{s'\in\mathcal{S}_r} \sum_{k\in s'} z_{r,s,k} \left(\sum_{\theta\in\Theta} \mu_\theta \sum_{\boldsymbol{s}: s_r = s'} \phi_\theta(\boldsymbol{s}) \, u_\theta^{r,k} \right) \\ + \sum_{\theta\in\Theta, \boldsymbol{s}\in\mathcal{S}} v_{\theta,\boldsymbol{s}} \phi_\theta(\boldsymbol{s}) + \sum_{\theta\in\Theta} d_\theta \left(\sum_{\boldsymbol{s}\in\mathcal{S}} \phi_\theta(\boldsymbol{s}) - 1 \right) \\ + \sum_{\boldsymbol{k}\in K} \nu_{\boldsymbol{k}} \left(x_{\boldsymbol{k}} - \sum_{\theta\in\Theta, \boldsymbol{s}\in\mathcal{S}} \mu_\theta \phi_\theta(\boldsymbol{s}) f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) \right). \end{aligned}$$

We observe that Slater's condition holds for Ⓟ (all constraints are linear, and by setting $\boldsymbol{x} = 0$ any signaling scheme ϕ constitutes a feasible solution). Therefore, by strong duality, an optimal dual solution must satisfy the KKT conditions. In particular, in order for stationarity to hold, it must be $\mathbf{0} \in \partial_{\phi_\theta(\boldsymbol{s})}(L)$ for each $\boldsymbol{s}$ and θ. Then, for each $\theta \in \Theta$ and $\boldsymbol{s} \in \mathcal{S}$, we have

$$\partial_{\phi_\theta(\boldsymbol{s})}(L) = \sum_{r \in \mathcal{R}} \sum_{k \in s_r} \mu_\theta z_{r,s_r,k} u_\theta^{r,k} + v_{\theta,\boldsymbol{s}} + d_\theta - \sum_{\boldsymbol{k} \in K} \nu_{\boldsymbol{k}} \mu_\theta f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) = 0.$$

Then, for each $\theta \in \Theta$ and $\boldsymbol{s} \in \mathcal{S}$, we obtain

$$\sum_{r \in \mathcal{R}} \sum_{k \in s_r} \mu_\theta z_{r,s_r,k} u_\theta^{r,k} + d_\theta - \sum_{\boldsymbol{k} \in K} \nu_{\boldsymbol{k}} \mu_\theta f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}})) \geq 0. \tag{12}$$

Moreover, stationarity has to hold with respect to variables $\boldsymbol{x}$. Formally, for each $\boldsymbol{k} \in K$,

$$\partial_{x_{\boldsymbol{k}}}(L) = 2(x_{\boldsymbol{k}} - y_{\boldsymbol{k}})\nu_{\boldsymbol{k}} = 0.$$

Therefore, for each $\boldsymbol{k} \in K$,

$$x_{\boldsymbol{k}} = y_{\boldsymbol{k}} - \frac{\nu_{\boldsymbol{k}}}{2}. \tag{13}$$

By Equations (12) and (13), we obtain the following dual quadratic program

$$\text{Ⓓ} \begin{cases} \max\limits_{\boldsymbol{z},\boldsymbol{v},\boldsymbol{\nu},\boldsymbol{d}} \ \sum\limits_{\boldsymbol{k} \in K} \left(\nu_{\boldsymbol{k}} y_{\boldsymbol{k}} - \frac{\nu_{\boldsymbol{k}}^2}{4} \right) - \sum\limits_{\theta \in \Theta} d_\theta & \\ \text{s.t.} \quad d_\theta \geq \sum\limits_{\boldsymbol{k} \in K} \nu_{\boldsymbol{k}} \mu_\theta f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \sum\limits_{r \in \mathcal{R}} \sum\limits_{k \in s_r} \mu_\theta z_{r,s_r,k} u_\theta^{r,k} & \forall \theta \in \Theta, \forall \boldsymbol{s} \in \mathcal{S} \\ \quad z_{r,s,k} \geq 0 & \forall r \in \mathcal{R}, \forall s \in \mathcal{S}_r, \forall k \in \mathcal{K}_r : k \in s \\ \quad \nu_{\boldsymbol{k}} \geq 0 & \forall \boldsymbol{k} \in K, \end{cases}$$

in which the objective function is obtained by observing that each term $\phi_\theta(\boldsymbol{s})$ in the definition of L is multiplied by $\partial_{\phi_\theta(\boldsymbol{s})}(L)$, which has to be equal to zero by stationarity. Similarly to what we did in the proof of Theorem 4, we repeatedly apply the ellipsoid method equipped with an approximate separation oracle to problem Ⓓ. In this case, the analysis is more involved than what happens in Theorem 4, because we are interested in computing an approximate projection on $\alpha\mathcal{X}_K$ rather than an approximate solution of Ⓟ. We proceed by casting Ⓓ as a feasibility problem with a certain objective (analogously to Ⓕ in Theorem 4). In particular, given objective $\gamma \in [0, 1]$, the objective function of Ⓓ becomes the following constraint in the feasibility problem

$$\sum_{\boldsymbol{k} \in K} \left(\nu_{\boldsymbol{k}} y_{\boldsymbol{k}} - \frac{\nu_{\boldsymbol{k}}^2}{4} \right) - \sum_{\theta \in \Theta} d_\theta \geq \gamma. \tag{14}$$

Then, given an approximation oracle $\mathcal{O}_\alpha$ which will be specified later, we apply to the feasibility problem the search algorithm described in Algorithm 2.

Algorithm 2 SEARCH ALGORITHM

Input: Error ϵ, $\boldsymbol{y} \in \mathbb{R}_+^{|\mathcal{K}|}$, subspace $K \subseteq \mathcal{K}$.

1: **Initialization**: $\beta \leftarrow \frac{\epsilon}{2}$, $\delta \leftarrow \frac{\epsilon}{2|\Theta|}$, $\gamma \leftarrow |K| + \beta$, and $\mathcal{H} \leftarrow \varnothing$.
2: **repeat**
3: $\quad \gamma \leftarrow \gamma - \beta$
4: $\quad \mathcal{H}_{\text{UNF}} \leftarrow \mathcal{H}$
5: $\quad \mathcal{H} \leftarrow \{$violated constraints returned by the ellipsoid method on Ⓓ with objective γ and constraints $\mathcal{H}_{\text{UNF}}\}$
6: **until** Ⓓ is feasible with objective γ (see Equation (14))
7: return $\mathcal{H}_{\text{UNF}}$

At each iteration of the main loop, given an objective value γ, Algorithm 2 checks whether the problem Ⓓ is approximately feasible or unfeasible, by applying the ellipsoid algorithm with separation oracle $\mathcal{O}_\alpha$. Let $\mathcal{H}$ be the set of constraints returned by the separation oracle (the separating hyperplanes due to the linear inequalities). At each iteration, the ellipsoid method is applied on the problem with explicit constraints in the current set $\mathcal{H}_{\text{UNF}}$ (that is, each constraint in $\mathcal{H}_{\text{UNF}}$ is explicitly checked for feasibility), while the other constraints are checked through the approximate separation oracle. Algorithm 2 returns the set of violated constraints $\mathcal{H}_{\text{UNF}}$ corresponding to the last value of γ for which the problem was unfeasible. Now, we describe how to implement the approximate separation oracle employed in Algorithm 2. Then, we conclude the proof by showing how to build an approximate projection starting from the set $\mathcal{H}_{\text{UNF}}$ computed as we just described.

Approximate Separation Oracle Let $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{d}})$ be a point in the space of dual variables. Then, let, for each $\theta \in \Theta$, $r \in \mathcal{R}$, and $s \in \mathcal{S}_r$,

$$w^{\theta}_{r,s} := \sum_{k \in s} \bar{z}_{r,s,k} \mu_\theta u^{r,k}_\theta .$$

First, we can check in polynomial time if one of the constraint in $\mathcal{H}$ is violated. If at least one of those constraint is violated, we output that constraint. Moreover, if the constraint corresponding to the objective is violated, we can output a separation hyperplane in polynomial time since the constraint has a polynomial number of variables. Then, by following the same rationale of the proof of Theorem 4 (offline setting), we proceed with a case analysis in which we ensure it is possible to output a violated constraint when $|\nu_{\boldsymbol{k}}|$ or $|w^{\theta}_{r,s}|$ are too large to guarantee polynomial-time sovability by Definition 2.

- First, it has to hold $d_\theta \in [0, 4|K|]$ for each $\theta \in \Theta$. Indeed, if $d_\theta < 0$, then the constraint relative to $(\theta, \varnothing)$ would be violated. Otherwise, suppose that there exists a θ with $\bar{d}_\theta > 4|K|$. Two cases are possible: (i) the constraint corresponding to the objective is violated, which allows us to output a separation hyperplane; (ii) it holds

  $$\sum_{\boldsymbol{k} \in K} \left(\bar{\nu}_{\boldsymbol{k}} y_{\boldsymbol{k}} - \frac{\bar{\nu}^2_{\boldsymbol{k}}}{4} \right) > 4|K|,$$

  which implies that there exists a $\boldsymbol{k} \in K$ such that $\bar{\nu}_{\boldsymbol{k}} y_{\boldsymbol{k}} - \bar{\nu}^2_{\boldsymbol{k}}/4 > 4$. However, we reach a contradiction since, by assumption, $y_{\boldsymbol{k}} \leq 2$ for each $\boldsymbol{k} \in K$, and therefore it must hold $\bar{\nu}_{\boldsymbol{k}} y_{\boldsymbol{k}} - \bar{\nu}^2_{\boldsymbol{k}}/4 \leq 2\bar{\nu}_{\boldsymbol{k}} - \bar{\nu}^2_{\boldsymbol{k}}/4 \leq 4$.

- Second, we show how to determine a violated constraint when $\bar{\nu}_{\boldsymbol{k}} \notin [0, |K| + 10]$. Specifically, if there exists a $\boldsymbol{k} \in K$ for which $\bar{\nu}_{\boldsymbol{k}} < 0$, then the objective is negative, and we can return a separation hyperplane (corresponding to Equation (14)). If there exists a $\nu_{\boldsymbol{k}} > |K| + 10$, then

  $$\begin{aligned} \sum_{\boldsymbol{k}' \in K} \left(\bar{\nu}_{\boldsymbol{k}'} y_{\boldsymbol{k}'} - \frac{\bar{\nu}^2_{\boldsymbol{k}'}}{4} \right) &\leq 2\nu_{\boldsymbol{k}} - \frac{\bar{\nu}^2_{\boldsymbol{k}}}{4} + \sum_{\boldsymbol{k}' \in K \setminus \{\boldsymbol{k}\}} \left(2\bar{\nu}_{\boldsymbol{k}'} - \frac{\bar{\nu}^2_{\boldsymbol{k}'}}{4} \right) \\ &\leq 2|K| + 20 - \frac{|K|^2}{4} - 5|K| - 25 + 4|K| \\ &= -\frac{|K|^2}{4} + |K| - 5 \\ &< 0, \end{aligned}$$

  where the first inequality follows by the assumption that $y_{\boldsymbol{k}} \leq 2$ for each $\boldsymbol{k} \in K$, and the second inequality follows from the fact that $2\nu_{\boldsymbol{k}} - \bar{\nu}^2_{\boldsymbol{k}}/4$ has its maximum in $\bar{\nu}_{\boldsymbol{k}} = 4$ and, when $\bar{\nu}_{\boldsymbol{k}} \geq |K| + 10$, the maximum is at $\bar{\nu}_{\boldsymbol{k}} = |K| + 10$ since the function in concave. Hence, we obtain that Constraint (14) is violated.

- Finally, suppose that there exists a $\theta \in \Theta$, $r \in \mathcal{R}$, $s \in \mathcal{S}_r$ such that $w^{\theta}_{r,s} > 4|K|$. Then, the constraint corresponding to (θ, s) is violated (because $d_\theta \leq 4|K|$, otherwise we would have already determined a violated constraint in the first case of our analysis). If, instead, there exists a $\theta \in \Theta$, $r \in \mathcal{R}$, $s \in \mathcal{S}_r$ such that $w^{\theta}_{r,s} < -4|K||\mathcal{R}| - 10$, then, for all the inequalities $(\theta, \boldsymbol{s}')$ with $s'_r = s$, it holds $\bar{d}_\theta \geq 0$ and

  $$\mu_\theta \sum_{\boldsymbol{k} \in K} \bar{\nu}_{\boldsymbol{k}} f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) + \sum_{r' \in \mathcal{R} \setminus \{r\}} w^{\theta}_{r', s'_{r'}} + w^{\theta}_{r, s'_r} \leq 0.$$

  In this last case, all the inequalities corresponding to $(\theta, \boldsymbol{s}')$ with $s'_r = s$ are guaranteed to be satisfied. Then, we can safely manage all the inequalities comprising of $w^{\theta}_{r,s} \leq -4|K||\mathcal{R}| - 10$ by setting $w^{\theta}_{r,s} = -4|K||\mathcal{R}| - 10$.

After the previous steps, it is guaranteed that $|w^\theta_{r,s}| \le 4|K||\mathcal{R}|+10$ for each θ, r, s, and $\nu_{\boldsymbol{k}} \in [0, |K|+10]$ for each $\boldsymbol{k}$. Hence, we can employ an oracle $\mathcal{O}_\alpha$ with $|w^\theta_{r,s}|$ and $\lambda^\theta_{\boldsymbol{k}} = \nu_{\boldsymbol{k}}\mu_\theta$, which is guaranteed to be polynomial in the size of the instance by Definition 2. Let δ be an error parameter which will be defined in the remainder of the proof. For each $\theta \in \Theta$, we call the oracle $\mathcal{O}_\alpha(\theta, K, \{\nu_{\boldsymbol{k}}\}_{\boldsymbol{k}\in K}, \boldsymbol{w}^\theta, \delta)$. Each query to the oracle returns an $\boldsymbol{s}^\theta$. If at least one of the constraints corresponding to a pair $(\theta, \boldsymbol{s}^\theta)$ is violated, we output that constraint. Otherwise, if for each $\theta \in \Theta$ the constraint $(\theta, \boldsymbol{s}^\theta)$ is satisfied, we conclude that the point is in the feasible region.

Putting It All Together Algorithm 2 terminates at objective $\gamma^\star$. It is easy to see that the algorithm terminates in polynomial time because it must return *feasible* when $\gamma = 0$. Our proof proceeds in two steps. First, we prove that a particular problem obtained from (P) has value at least $\gamma^\star$. Then, we prove that the solution of (P) with only variables in $\mathcal{H}_{\text{UNF}}$ has value close to $\gamma^\star$. Finally, we show that the two solutions are, respectively, the projection and an approximate projection on a set that includes $\alpha\mathcal{X}_K$. This will complete the proof.

If the algorithm terminates at objective γ^*, the following convex optimization problem is feasible (see Theorem 4).[10]

$$\begin{cases} \sum_{\boldsymbol{k}\in K} \left(\nu_{\boldsymbol{k}} y_{\boldsymbol{k}} - \nu_{\boldsymbol{k}}^2/4\right) - \sum_{\theta\in\Theta} d_\theta \ge \gamma^\star \\ d_\theta \ge \sum_{\boldsymbol{k}\in K} \nu_{\boldsymbol{k}}\mu_\theta f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) - \sum_{r\in\mathcal{R}, k\in s_r} z_{r,s_r,k}\, \mu_\theta u^{r,k}_\theta & \forall(\theta,\boldsymbol{s}) \in \mathcal{H}_{\text{UNF}} \\ d_\theta \ge \sum_{\boldsymbol{k}\in K} \alpha\nu_{\boldsymbol{k}}\mu_\theta f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) - \sum_{r\in\mathcal{R}, k\in s_r} z_{r,s_r,k}\, \mu_\theta u^{r,k}_\theta - \delta & \forall(\theta,\boldsymbol{s}) \notin \mathcal{H}_{\text{UNF}}. \end{cases}$$

By strong duality, the following convex optimization problem has value at least $\gamma^\star$

$$\text{(Pf)}\begin{cases} \min_{\phi,\boldsymbol{x}} \ \sum_{\boldsymbol{k}\in K} (x_{\boldsymbol{k}} - y_{\boldsymbol{k}})^2 + \delta \sum_{(\theta,\boldsymbol{s})\notin\mathcal{H}_{\text{UNF}}} \phi_\theta(\boldsymbol{s}) \\ \text{s.t.} \ \sum_{\theta\in\Theta} \mu_\theta \left(\sum_{\boldsymbol{s}': s'_r = s} \phi_\theta(\boldsymbol{s}') u^{r,k}_\theta \right) \ge 0 & \forall r\in\mathcal{R}, \forall s\in\mathcal{S}_r, \forall k\in\mathcal{K}_r : k \in s \\ \sum_{\boldsymbol{s}\in\mathcal{S}} \phi_\theta(\boldsymbol{s}) = 1 & \forall\theta\in\Theta \\ \phi_\theta(\boldsymbol{s}) \ge 0 & \forall\theta\in\Theta, \forall\boldsymbol{s}\in\mathcal{S} \\ x_{\boldsymbol{k}} \le \sum_{\theta\in\Theta} \left(\sum_{\boldsymbol{s}:(\theta,\boldsymbol{s})\in\mathcal{H}_{\text{UNF}}} \mu_\theta\, \phi_\theta(\boldsymbol{s}) f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) + \alpha \sum_{\boldsymbol{s}:(\theta,\boldsymbol{s})\notin\mathcal{H}_{\text{UNF}}} \mu_\theta\, \phi_\theta(\boldsymbol{s}) f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) \right) & \forall\boldsymbol{k}\in K. \end{cases}$$

Moreover, since the algorithm did not terminate at value $\gamma^\star + \beta$, problem (D) with value $\gamma^\star + \beta$ is unfeasible when restricting the set of constraints to $\mathcal{H}_{\text{UNF}}$. The primal problem (P) restricted to primal variables corresponding to dual constraints in $\mathcal{H}_{\text{UNF}}$ reads as follows

$$\begin{cases} \min_{\phi,\boldsymbol{x}} \ \sum_{\boldsymbol{k}\in K} (x_{\boldsymbol{k}} - y_{\boldsymbol{k}})^2 \\ \text{s.t.} \ \sum_{\theta\in\Theta} \mu_\theta \left(\sum_{\substack{\boldsymbol{s}:(\theta,\boldsymbol{s})\in\mathcal{H}_{\text{UNF}}, \\ s_r = s'}} \phi_\theta(\boldsymbol{s}) u^{r,k}_\theta \right) \ge 0 & \forall r\in\mathcal{R}, s'\in\mathcal{S}_r, \forall k\in\mathcal{K}_r : k \in s' \\ \sum_{\boldsymbol{s}:(\theta,\boldsymbol{s})\in\mathcal{H}_{\text{UNF}}} \phi_\theta(\boldsymbol{s}) = 1 & \forall\theta\in\Theta \\ \phi_\theta(\boldsymbol{s}) \ge 0 & \forall(\theta,\boldsymbol{s})\in\mathcal{H}_{\text{UNF}} \\ x_{\boldsymbol{k}} \le \sum_{\theta\in\Theta} \sum_{\boldsymbol{s}:(\theta,\boldsymbol{s})\in\mathcal{H}_{\text{UNF}}} \mu_\theta\, \phi_\theta(\boldsymbol{s}) f_\theta(R^{\boldsymbol{k}}_{\boldsymbol{s}}) & \forall\boldsymbol{k}\in K. \end{cases}$$

[10] In the following, we will refer to the proof of Theorem 4 when the steps of the two proofs are analogous.

By strong duality, the above problem has value at most $\gamma^\star + \beta$. Moreover, it has a polynomial number of variables and constraints because the ellipsoid method returns a set of constraints $\mathcal{H}_{\text{UNF}}$ of polynomial size. Therefore, the above problem can be solved in polynomial time.

A solution to the above problem is a feasible signaling scheme. Let $(\boldsymbol{x}^\epsilon, \phi)$ be its solution. We have that $\boldsymbol{x}^\epsilon \in \bar{\mathcal{X}}_K$, with

$$\bar{\mathcal{X}}_K = \left\{ \boldsymbol{x} : x_{\boldsymbol{k}} \leq \sum_{\theta \in \Theta} \left(\sum_{\boldsymbol{s}:(\theta,\boldsymbol{s}) \in \mathcal{H}_{\text{UNF}}} \mu_\theta \, \phi_\theta(\boldsymbol{s}) f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) + \alpha \sum_{\boldsymbol{s}:(\theta,\boldsymbol{s}) \notin \mathcal{H}_{\text{UNF}}} \mu_\theta \, \phi_\theta(\boldsymbol{s}) f_\theta(R_{\boldsymbol{s}}^{\boldsymbol{k}}) \right) \quad \forall \boldsymbol{k} \in K, \phi \in \Phi \right\}.$$

It holds $\alpha \mathcal{X}_K \subseteq \bar{\mathcal{X}}_K$. Now, we show that $\boldsymbol{x}^\epsilon$ is *close* to $\boldsymbol{x}^\star$, where $\boldsymbol{x}^\star$ is the projection of $\boldsymbol{y}$ on $\bar{\mathcal{X}}_K$ (that is the solution of Pf with $\delta = 0$). Since $x^\star$ is a feasible solution of Pf and the minimum of Pf is at least $\gamma^\star$, it holds $||\boldsymbol{x}^\star - \boldsymbol{y}||^2 + \delta|\Theta| \geq \gamma^\star$. Then,

$$\begin{aligned} ||\boldsymbol{x}^\star - \boldsymbol{y}||^2 + \delta|\Theta| + \beta &\geq \gamma^\star + \beta \\ &\geq ||\boldsymbol{x}^\epsilon - \boldsymbol{y}||^2 \\ &= ||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star + \boldsymbol{x}^\star - \boldsymbol{y}||^2 \\ &= ||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star||^2 + ||\boldsymbol{x}^\star - \boldsymbol{y}||^2 + 2\langle \boldsymbol{x}^\epsilon - \boldsymbol{x}^\star, \boldsymbol{x}^\star - \boldsymbol{y} \rangle \\ &\geq ||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star||^2 + ||\boldsymbol{x}^\star - \boldsymbol{y}||^2, \end{aligned}$$

where the last inequality follows from $\langle \boldsymbol{x}^\epsilon - \boldsymbol{x}^\star, \boldsymbol{x}^\star - \boldsymbol{y} \rangle \geq 0$, because $\boldsymbol{x}^\star$ is the projection of $\boldsymbol{y}$ on $\bar{\mathcal{X}}_K$ and $\boldsymbol{x}^\epsilon \in \bar{\mathcal{X}}_K$. Hence, $||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star||^2 \leq \delta|\Theta| + \beta$. Finally, let $\boldsymbol{x}$ be a point in $\alpha\mathcal{X}_K$. Then,

$$\begin{aligned} ||\boldsymbol{x}^\epsilon - \boldsymbol{x}||^2 &\leq ||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star||^2 + ||\boldsymbol{x}^\star - \boldsymbol{x}||^2 \\ &\leq ||\boldsymbol{x}^\epsilon - \boldsymbol{x}^\star||^2 + ||\boldsymbol{y} - \boldsymbol{x}||^2 \\ &\leq ||\boldsymbol{y} - \boldsymbol{x}||^2 + \delta|\Theta| + \beta, \end{aligned}$$

where the second inequality follow from the fact that $\boldsymbol{x}^\star$ is the projection of $\boldsymbol{y}$ on a superset of $\alpha\mathcal{X}_K$. Setting $\delta = \frac{\epsilon}{2|\Theta|}$ and $\beta = \frac{\epsilon}{2}$ concludes the proof. □

E. Proofs Omitted from Section 7

In this section, we provide the complete proof of Theorem 11.

Firs, we introduce some preliminary, known results concerning the optimization over matroids. Given a *non-decreasing submodular* set function $f : 2^{\mathcal{G}} \to \mathbb{R}_+$ and a *linear* set function $\ell : 2^{\mathcal{G}} \ni I \mapsto \sum_{i \in I} w_i$ defined for finite ground set $\mathcal{G}$ and weights $\boldsymbol{w} = (w_i)_{i \in \mathcal{G}}$ with $w_i \in \mathbb{R}$ for each $i \in \mathcal{G}$, let us consider the problem of maximizing the sum $f(I) + \ell(I)$ over the bases $I \in \mathcal{B}(\mathcal{M})$ of a given matroid $\mathcal{M} := (\mathcal{G}, \mathcal{I})$. We make use of a theorem due to Sviridenko et al. (2017), which, by letting $v_f := \max_{I \in 2^{\mathcal{G}}} f(I)$, $v_\ell := \max_{I \in 2^{\mathcal{G}}} |\ell(I)|$, and $v := \max\{v_f, v_\ell\}$, reads as follows:

Theorem 11 (Essentially Theorem 3.1 by Sviridenko et al. (2017))**.** *For every $\epsilon > 0$, there exists an algorithm running in time* poly $\left(|\mathcal{G}|, \frac{1}{\epsilon}\right)$ *that produces a basis $I \in \mathcal{B}(\mathcal{M})$ satisfying $f(I) + \ell(I) \geq \left(1 - \frac{1}{e}\right) f(I') + \ell(I') - O(\epsilon)v$ for every $I' \in \mathcal{B}(\mathcal{M})$ with high probability.*

Next, we provide the complete proof of Theorem 11.

Theorem 6. *If the sender's utility is such that function f_θ is* submodular *for each $\theta \in \Theta$, then there exists a polynomial-time separation oracle $\mathcal{O}_{1-\frac{1}{e}}$.*

Proof. We show how to implement an approximation oracle $\mathcal{O}_\alpha(\theta, K, \boldsymbol{\lambda}, \boldsymbol{w}, \epsilon)$ (see Definition 2) running in time poly $\left(n, |K|, \max_{r,s} |w_{r,s}|, \max_{\boldsymbol{k}} \lambda_{\boldsymbol{k}}, \frac{1}{\epsilon}\right)$ for $\alpha = 1 - \frac{1}{e}$. Let $\mathcal{M}_{\mathcal{S}} := (\mathcal{G}_{\mathcal{S}}, \mathcal{I}_{\mathcal{S}})$ be a matroid defined as in Section 7 for direct signal profiles $\mathcal{S}$. Let us recall that, given the relation between the bases of $\mathcal{M}_{\mathcal{S}}$ and direct signals, each direct signal profiles $\boldsymbol{s} \in \mathcal{S}$ corresponds to a basis $I \in \mathcal{B}(\mathcal{M}_{\mathcal{S}})$, which is defined as $I := \{(r, s_r) \mid r \in \mathcal{R}\}$. In the following, given a subset $I \subseteq \mathcal{G}_{\mathcal{S}}$ and a type profile $\boldsymbol{k} \in K$, we let $R_I^{\boldsymbol{k}} \subseteq \mathcal{R}$ be the set of receivers $r \in \mathcal{R}$ such that there exits a pair $(r, s) \in I$ (for some signal $s \in \mathcal{S}_r$) with the receiver's type k_r being recommended to play a_1 under signal s; formally,

$$R_I^{\boldsymbol{k}} := \{r \in \mathcal{R} \mid \exists (r, s) \in I : k_r \in s\}.$$

First, we show that, when using matroid notation, the left-hand side of Equation (3) can be expressed as the sum of a non-decreasing submodular set function and a linear set function. To this end, let $f_\theta^{\boldsymbol{\lambda}} : 2^{\mathcal{G}_\mathcal{S}} \to \mathbb{R}_+$ be defined as $f_\theta^{\boldsymbol{\lambda}}(I) = \sum_{\boldsymbol{k}\in K} \lambda_{\boldsymbol{k}} f_\theta(R_I^{\boldsymbol{k}})$ for every subset $I \subseteq \mathcal{G}_\mathcal{S}$. We prove that $f_\theta^{\boldsymbol{\lambda}}$ is submodular. Since $f_\theta^{\boldsymbol{\lambda}}$ is a suitably defined weighted sum of the functions f_θ, it is sufficient to prove that, for each type profile $\boldsymbol{k} \in K$, the function $f_\theta : 2^\mathcal{R} \to [0,1]$ is submodular in the sets $R_I^{\boldsymbol{k}}$. For every pair of subsets $I \subseteq I' \subseteq \mathcal{G}_\mathcal{S}$, and for every receiver $r \in \mathcal{R}$ and signal $s \in \mathcal{S}_r$, the marginal contribution to the value of function f_θ due to the addition of element (r,s) to the set I is:

$$\begin{aligned} f_\theta(R^{\boldsymbol{k}}_{I\cup(r,s)}) - f_\theta(R^{\boldsymbol{k}}_I) &= \mathbb{I}\left\{k_r \in s \wedge \nexists (r,s') \in I : k_r \in s'\right\} \left(f_\theta(R^{\boldsymbol{k}}_I \cup \{r\}) - f_\theta(R^{\boldsymbol{k}}_I)\right) \geq \\ &\geq \mathbb{I}\left\{k_r \in s \wedge \nexists (r,s') \in I' : k_r \in s'\right\} \left(f_\theta(R^{\boldsymbol{k}}_I \cup \{r\}) - f_\theta(R^{\boldsymbol{k}}_I)\right) \geq \\ &\geq \mathbb{I}\left\{k_r \in s \wedge \nexists (r,s') \in I' : k_r \in s'\right\} \left(f_\theta(R^{\boldsymbol{k}}_{I'} \cup \{r\}) - f_\theta(R^{\boldsymbol{k}}_{I'})\right) = \\ &= f_\theta(R^{\boldsymbol{k}}_{I'\cup(r,s)}) - f_\theta(R^{\boldsymbol{k}}_{I'}), \end{aligned}$$

where the last inequality holds since the functions f_θ are submodular by assumption. Since the last expression is the marginal contribution to the value of function f_θ due to the addition of element (r,s) to the set I', the relations above prove that the function $f_\theta^{\boldsymbol{\lambda}}$ is submodular. Let $\ell^{\boldsymbol{w}} : 2^{\mathcal{G}_\mathcal{S}} \to \mathbb{R}_+$ be a linear function such that $\ell^{\boldsymbol{w}}(I) = \sum_{r\in\mathcal{R}} w_{r,s_r}$ for every basis $I \subseteq \mathcal{B}(\mathcal{M}_\mathcal{S})$, with each $s_r \in \mathcal{S}_r$ being the signal of receiver $r \in \mathcal{R}$ specified by the signal profile corresponding to the basis, namely $(r, s_r) \in I$. Then, we have that finding a signal profile $\boldsymbol{s} \in \mathcal{S}$ satisfying Equation (3) is equivalent to finding a basis $I \in \mathcal{B}(\mathcal{M}_\mathcal{S})$ of the matroid $\mathcal{M}_\mathcal{S}$ (representing a direct signal profile) such that:

$$f_\theta^{\boldsymbol{\lambda}}(I) + \ell^{\boldsymbol{w}}(I) \geq \max_{I^\star \in \mathcal{B}(\mathcal{M}_\mathcal{S})} \left\{ \alpha \sum_{\boldsymbol{k}\in K} f_\theta^{\boldsymbol{\lambda}}(I^\star) + \ell^{\boldsymbol{w}}(I^\star) \right\} - \epsilon.$$

Notice that, for $\epsilon' > 0$, the algorithm of Theorem 11 by Sviridenko et al. (2017) can be employed to find a basis $I \in \mathcal{B}(\mathcal{M}_\mathcal{S})$ such that $f_\theta^{\boldsymbol{\lambda}}(I) + \ell^{\boldsymbol{w}}(I) \geq \left(1 - \frac{1}{e}\right) f_\theta^{\boldsymbol{\lambda}}(I') + \ell^{\boldsymbol{w}}(I') - O(\epsilon')v$ for every $I' \in \mathcal{B}(\mathcal{M})$ with high probability, employing time polynomial in $|\mathcal{G}_\mathcal{S}|$ and $\frac{1}{\epsilon}$. Since $|\mathcal{G}_\mathcal{S}|$ is polynomial in n and v is polynomial in $|K|$, $\max_{r,s} |w_{r,s}|$ and $\max_{\boldsymbol{k}} \lambda_{\boldsymbol{k}}$, by setting $\epsilon' = O(\frac{\epsilon}{v})$ and $\alpha = 1 - \frac{1}{e}$, we get the result. $\square$