Regret Minimization in Single-Dimensional Contract-Design with Binary Actions

Riccardo Poiani
Bocconi University
riccardo.poiani@unibocconi.it

Martino Bernasconi
Bocconi University
martino.bernasconi@unibocconi.it

Andrea Celli
Bocconi University
andrea.celli2@unibocconi.it

Abstract

We study principal-agent problems in which a principal commits to an outcome-dependent payment scheme (*i.e.*, a *contract*) in order to induce an agent to take a costly action leading to a favorable outcome. We consider the online extension of the classical (one-shot) principal-agent problem, in which the principal repeatedly interacts with agents by proposing contracts over multiple rounds. The principal has no information about the agents and, crucially, does not observe their actions. As a result, the principal must learn an optimal contract using only the realized outcomes observed at each round. We focus on the setting with binary actions and single-dimensional agent types, where the agent's private type represents their cost per unit-of-effort. For adversarial-type sequences, we provide tight $\Theta(T^{2/3})$ regret guarantees. Remarkably, this rate is completely independent of the number of outcomes m. The upper bound is based on two key components: 1) a reduction to a one-dimensional threshold optimization problem and 2) a non-uniform discretization to handle the non-Lipschitz nature of the problem. Moreover, in the case of a single (fixed) hidden type, we show that it is possible to improve the rates and provide a tight $\widetilde{\Theta}(\sqrt{T})$ regret bound. Our algorithm is based on an explore-then-commit strategy where we first approximately learn the hidden type via a stochastic binary search, and then we commit to a "robustified" near-optimal contract.

1 Introduction

Principal-agent problems model scenarios in which a principal seeks to influence the behavior of a self-interested agent capable of performing costly actions. The principal commits to an outcome-dependent payment rule (*i.e.*, a *contract*), after which the agent selects one of the available actions. In the standard hidden-action model, the principal does not observe the agent's action directly, but only a stochastic outcome whose distribution depends on the chosen action and determines the principal's reward. The principal's objective is then to design a contract that incentivizes the agent to take actions leading to favorable outcomes. This model is central to modern economic theory and has recently given rise to a thriving body of research studying it through a computational lens. Such interest is motivated by applications like online labor platforms [39], strategic data provisioning [18], pay-for-performance healthcare [10, 11], blockchain [28] and delegation to AI agents [37, 44, 45, 48, 46].

We study the generalization of the classical (single-round) model, in which the principal faces uncertainty about the agents' types and must learn an optimal contract from data. In particular, the

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

principal repeatedly interacts with agents over multiple rounds. At each round, the principal commits to a new contract and observes the outcome produced by the agent's response. The goal of the principal is minimizing the regret with respect to the best contract in hindsight (*i.e.*, given knowledge of the sequence of agent types). The best known regret guarantees for general online contract design problems are nearly linear [49] or exponential in the instance size [6]. Stronger sublinear-regret guarantees are known under additional structural assumptions, such as regularity conditions relating action costs and outcome distributions, bounded-density assumptions, or restrictions on the class of admissible contracts (see, e.g., [49, 26, 15] and the discussion in Section 1.2). In this work, we ask whether there are natural settings in which one can obtain tight sublinear regret for general classes of contracts, *without* imposing further assumptions on the type or outcome distributions.

Motivated by known hardness results for the Bayesian setting with multi-dimensional types [21, 36, 22], we study the single-dimensional type setting introduced by Alon et al. [1], in which the agent's type is a scalar representing a cost per unit of effort (*i.e.*, a parameter that scales the cost of the chosen action). We consider the case in which agents have binary actions, so they can either exert effort or not. This is a canonical model in economics, capturing the standard "work / shirk" formulation of moral hazard (see, *e.g.*, [38, 4, 31, 35]). A useful way to interpret this model is through a simple delegated-learning example in the spirit of Saig et al. [44], where a principal outsources the training of a classifier to a provider. The provider can either invest *high effort* (a_1), for example by performing extensive data annotation and hyperparameter tuning, or training on a more expensive pipeline, or choose *low effort* (a_2), corresponding to a cheaper baseline procedure. The provider's private type θ represents its cost per unit-of-effort, so that the cost of action a is $\theta \cdot C_a$. After training, the principal does not observe which pipeline was actually used, but only a stochastic performance outcome (*e.g.*, validation accuracy), whose distribution depends on the chosen action. A second example is a crowdsourcing or data-labeling platform. In each round, a new worker arrives to complete a task, and the platform offers an outcome-contingent payment rule, for example, a base payment together with a bonus tied to an observable quality signal (*e.g.*, agreement with other workers, expert review, or the performance of a downstream model trained on the collected data). The worker can either exert high effort, corresponding to careful completion of the task, or low effort, corresponding to a cheap but lower-quality submission. Again, the platform does not observe the worker's effort directly, but only the realized quality outcome, whose distribution depends on the chosen action. In both settings, a contract specifies a payment as a function of the realized performance outcome. In this binary-action setting, every contract trades off incentives and payments by determining which types find it worthwhile to choose the high-effort action, thereby inducing a threshold on the hidden type space. Learning a good contract then amounts to learning the right threshold from outcome-only feedback.

1.1 Our Results and Techniques

Main Results. We establish a tight $\Theta(T^{2/3})$ regret bound when the sequence of agent types is chosen adversarially (Section 3). We then turn to the setting in which the agent type is fixed throughout the interaction (though it is hidden) and show that the optimal rate improves to $\widetilde{\Theta}(\sqrt{T})$ (Section 4).

Key Features of Our Bounds. A notable feature of these results is that, unlike previous guarantees, they are completely independent of the number of outcomes m. This is particularly desirable in environments with large outcome spaces, where existing bounds may deteriorate substantially with m (see, *e.g.*, the $\widetilde{O}(T^{1-1/(2m+1)})$ bound by Zhu et al. [49]). Our guarantee for adversarial-type sequences also improves over the $\widetilde{O}(m^2T^{4/5})$ bound of Bacchiocchi et al. [6] (specialized to the two-action setting), not only in its dependence on the horizon T, but also in its dependence on the instance parameters, since our rate does not scale with m at all. Moreover, our results are surprising if we view online contract design as a special case of repeated Stackelberg games with myopic agents [42]. Indeed, in general, known hardness results show that there exist instances for which learning an optimal commitment requires a number of samples exponential in the number of actions of the leader (*i.e.*, the principal in our terminology) [43]. In contrast, we can exploit the specific structure of our problem to obtain sublinear regret rates, despite the principal having a continuous action space and observing only outcome feedback. Finally, our bounds provide the first evidence that no-regret learning is possible in the single-dimensional-type setting without bounded-density assumptions, addressing a question recently posed by Bernasconi et al. [15] in the special case of binary actions.

Techniques. We briefly highlight here the main technical components of our analysis, and defer a broader discussion to the later sections.

- The first key idea to obtain the $O(T^{2/3})$ upper bound is showing that, although contracts live in an m-dimensional space, the search for the best fixed contract can be reduced to a one-dimensional optimization problem over the expected payment gap between the two actions. The resulting one-dimensional objective is not $O(1)$-Lipschitz, so standard uniform discretization arguments fail. To overcome this, we develop a non-uniform discretization tailored to the geometry of the linear program that defines the cheapest contract implementing a given incentive level. This yields a discretization with only $O(\varepsilon^{-1})$ relevant points while preserving enough control on the discretization error to obtain the desired regret bound.
- The lower bound shows that every algorithm suffers regret $\Omega(T^{2/3})$ under stochastic type generation with an absolutely continuous type distribution. The main challenges to obtain this result are the learner acts in a continuous contract space and the observed feedback does not coincide with the principal's utility. To overcome this, we construct a family of carefully perturbed continuous instances in which the utility improves only on a small interval of contracts while the induced outcome distributions remain hard to distinguish, and then apply a change-of-measure argument directly in the continuous action space.
- Finally, in order to obtain the $\widetilde{O}(\sqrt{T})$ upper bound in the case of a single fixed type, we design an algorithm exploiting the binary-action structure to perform a stochastic binary search over the unknown threshold induced by the hidden type. The main challenge is to design exploratory contracts that are informative enough to localize the threshold while remaining close enough to optimal to keep regret under control. We address this by combining threshold-testing contracts with a robustification step in the exploitation phase.

1.2 Related Work

In the following, we review studies most directly related to our work. For a comprehensive overview of algorithmic contract theory, readers are referred to the recent survey by Dütting et al. [33].

Typed Agents. Contracts for typed agents have been studied in both multi- and single-dimensional models. In the multi-dimensional setting, where the costs and outcome distribution can depend arbitrarily on the type, a series of works showed that the problem of approximating optimal (deterministic) contracts is APX-hard [21, 36, 22]. The single-dimensional model admits stronger positive results. In particular, Alon et al. [1, 2] show tractability for a constant number of actions, and also obtain positive approximation guarantees for simple contracts. Moreover, Bernasconi et al. [15] showed that the problem admits an additive PTAS. However, for multiplicative approximations, the single-dimensional problem remains as hard as the general multi-dimensional case [23].

Online Contract Design. The problem of learning optimal contracts was introduced by Ho et al. [38] and later extended in Cohen et al. [27]. The work most closely related to ours on learning optimal contracts in general principal-agent settings is Zhu et al. [49]. They provide an algorithm guaranteeing cumulative regret of $\widetilde{O}(\sqrt{m}T^{1-1/(2m+1)})$. Moreover, they show that it is possible to learn linear contracts (which are a weaker class of contracts in which the payment is proportional to the principal's reward) incurring regret of order $\widetilde{\Theta}(T^{2/3})$. They also provide an almost-matching lower bound of $\Omega(T^{1-1/(2m+1)})$ for the case of general bounded contracts. Their lower-bound construction employs instances with a number of actions exponential in m. The question of whether it is possible to do better in settings with small action spaces was answered by Bacchiocchi et al. [6], who showed that for bounded contracts it is possible to achieve regret $\widetilde{O}(m^n\, T^{4/5})$, which depends polynomially on the instance size if the number of actions n is constant. Moreover, for linear contracts, Bacchiocchi et al. [7] showed that $\widetilde{O}(\sqrt{ndT})$ regret is attainable in settings with a finite number of types d and a number of actions n that is small relative to T (*i.e.*, $n \leq T^{1/3}$).

Learning under Additional Assumptions. In order to circumvent the negative results of Zhu et al. [49], subsequent works have explored learning optimal contracts under additional assumptions. Chen et al. [26] show that if the problem satisfies First-Order Stochastic Dominance (FOSD) and the Concavity of Distribution Function Property (CDFP), then it is possible to achieve regret of order $\widetilde{O}(m^{11/21} \cdot T^{20/21} \cdot \log(1/\delta))$. Recently, Bernasconi et al. [15] studied a single-dimensional-type

setting similar to ours but requiring types to be generated according to a distribution admitting density bounded by some constant β. Under this assumption, they provide a learning algorithm with regret of order $\widetilde{O}(\beta\,\mathrm{poly}(n,m)\,T^{3/4})$. Here, the bounded-density assumption plays a crucial role, as it provides a way to control the discretization error and thereby enables a direct reduction from the continuous problem to a discrete one.

Other Feedback Models. Dütting et al. [32] study a stronger feedback model where the principal can observe their expected utility instead of a sampled outcome. In this setting, linear contracts can be learned, incurring regret at most $O(\log\log T)$. Finally, we mention the recent work of Dütting et al. [34] which uses the notion of pseudo-dimension to bound the sample complexity of approximately optimal contracts. While providing fundamental barriers to learnability in the case of unbounded contracts, they provide polynomial sample complexity results for bounded contracts, which translate to a regret of order $\widetilde{O}(\sqrt{mT})$ under a full feedback model (i.e., the learner observes the type after committing to a contract).

Our problem also shares connections with the more general problem of learning an optimal commitment, in repeated Stackelberg games [42, 43, 16, 8, 9] and online Bayesian persuasion [19, 20, 12].

2 Preliminaries

The single-dimensional Bayesian principal-agent problem is defined as follows. There is a set of outcomes Ω of size $|\Omega| = m$. Each possible outcome ω provides to the principal a reward $r_\omega \in [0,1]$. There are two actions a_1 and a_2 which are the high- and low-effort actions, respectively. Each action is associated with a cost $C_a \in [0,1]$, and we can assume without loss of generality that $C_{a_1} \geq C_{a_2}$.[1] Each action $a \in A := \{a_1, a_2\}$ is associated to a distribution over outcomes $F_a \in \Delta(\Omega)$. We denote with $F_{a,\omega}$ the probability that action a leads to outcome ω. We consider the single-dimensional-type problem introduced by Alon et al. [1] where the agent is associated with a hidden type $\theta \in [0,1]$, and the agent's cost of performing action a is $\theta \cdot C_a$. In this setting, the uncertainty is entirely encapsulated in the private types, the outcome distributions and costs are part of the publicly known environment.

The principal's goal is to choose a contract $p \in [0,1]^m$, where each component p_ω specifies the payment made in outcome ω. Based on the proposed contract p, the agent takes an action $a \in A$ and an outcome $\omega \sim F_a$ is reached. When an agent of type θ takes action a, the expected utility of the agent is $U^A_\theta(p,a) = \sum_{\omega\in\Omega} F_{a,\omega} p_w - \theta \cdot C_a$, while the principal has expected utility $U^P(p,a) = \sum_{\omega\in\Omega} F_{a,\omega}(r_\omega - p_\omega)$. After committing to a contract, the principal observes only the realized outcome ω, and observes neither the action $a \in A$ taken nor the type θ.

When an agent of type θ observes a contract p, we assume that it will play an action $a \in A$ that is both (1) *incentive compatible* (IC), *i.e.*, it maximizes the expected utility over actions, and (2) *individually rational* (IR), *i.e.*, $U^A_\theta(p,a) \geq 0$.

To conveniently incorporate IR into IC constraints, it is standard to assume that there exists an action a with cost $C_a = 0$. When multiple actions are IC, we assume that ties are broken in favor of the principal. More formally, let $\mathcal{B}(\theta,p) = \mathrm{argmax}_{a\in A} U^A_\theta(p,a)$ be the set of actions that are IC for type θ and contract p. Then, we define $b(\theta,p)$ as the action within $\mathcal{B}(\theta,p)$ that maximizes the principal's expected utility under p.[2] Finally, we conclude by introducing the following notation. For $\epsilon > 0$, we denote by $\mathcal{B}_\epsilon(\theta,p)$ as the sets of actions that are ϵ-IC for the agent, *i.e.*,

$$\mathcal{B}_\epsilon(\theta,p) = \left\{ a \in A : U^A(p,a) \geq \max_{b\in A} U^A(p,b) - \epsilon \right\}.$$

Online Setting. In this work, we focus on the online adversarial setting where the principal interacts with an arbitrary sequence of $T \in \mathbb{N}$ agents. During each round $t \in [T]$, a type θ_t arrives, and the principal selects a contract $p_t \in [0,1]^m$. The agent θ_t best responds, *i.e.*, it plays an action $a_t \in b(\theta_t, p_t)$ and the principal observes $\omega_t \sim F_{a_t}$ and receives $r_{\omega_t} - p_{w_t}$. The principal's goal is to minimize the regret against the best contract in hindsight for any fixed sequence of types, *i.e.*,

$$R_T = \sup_{\{\theta_t\}_{t\in[T]}} \sup_{p\in[0,1]^m} \left\{ \sum_{t=1}^T U^P(p, b(\theta_t,p)) - \mathbb{E}\left[\sum_{t=1}^T U^P(p_t, b(\theta_t,p_t)) \right] \right\},$$

[1] Without loss of generality, we assume that $C_{a_1} > 0$, otherwise the problem admits a trivial solution.

[2] Whenever the utilities of a_1 and a_2 are both optimal for the principal and the agent, we assume the agent plays a_1.

where the expectation is taken with respect to the stochasticity of the algorithm. In the following, whenever it is clear from context, we write $U^P(p,\theta) = U^P(p, b(\theta,p))$. Moreover, for a fixed sequence of types, we define $\textsc{Opt} := \sup_{p\in[0,1]^m} \sum_{t=1}^T U^P(p,\theta_t)$, omitting the dependence on the type sequence when it is clear from context. Throughout this document, we assume that $F_{a_1}^\top r > F_{a_2}^\top r$. If this condition does not hold, then there exists a trivial algorithm with zero regret.[3]

3 Adversarial Types: $\Theta(T^{2/3})$ is Tight

In this section, we consider the general adversarial setting in which an arbitrary sequence of types $\{\theta_t\}_{t=1}^T$ arrives. In Section 3.1, we propose an algorithm that attains $O(T^{2/3})$ regret. Then, in Section 3.2, we show that this rate is tight in the worst-case sense by proving a matching lower bound.

3.1 The Adversarial $O(T^{2/3})$ Upper Bound

In this section, we prove the following upper bound on adversarial types and provide a high-level overview of the algorithm. The formal proof is deferred to Section A.

Theorem 1. *There exists an algorithm for the online contract-design problem with binary actions and single-dimensional adversarial types, such that $R_T \le O(T^{2/3})$.*

A notable feature of the upper bound is that it is independent of the number of outcomes m. This is surprising since the action space is continuous and standard discretization arguments typically incur an exponential dependence on m. Our approach avoids this through two main ideas. First, we show that the search for an optimal contract can be reduced from an m-dimensional space to a one-dimensional one. The key parameter is the threshold type, *i.e.*, the highest type that can still be induced to choose action a_1, together with the cheapest contract that implements it (see Section 3.1.1). This reduces the problem to optimization over the interval $[0,1]$. Second, once we obtained the one-dimensional problem, the standard way of handling it would be to discretize the $[0,1]$ interval uniformly with step $\epsilon = T^{-1/3}$. If the underlying reward function was $O(1)$-Lipschitz, this would give a $O(\sqrt{T/\epsilon} + \epsilon T) = O(T^{2/3})$ upper bound (this also holds for one-sided Lipschitz functions [32]). However, when reformulated as a one-dimensional problem, the reward is not $O(1)$-Lipschitz. As a result, a clever discretization strategy is needed to avoid dependence on the number of outcomes m. This motivates the second key idea: a non-uniform discretization that exploits additional structure of the problem, which we discuss in more detail in Section 3.1.2.

3.1.1 The One-Dimensional Problem

To show that the problem can be reduced to a one-dimensional one, we start by analyzing the utility of the principal when it plays a contract p. To this end, we will only focus on the subset of contracts for which $F_{a_1}^\top(r-p) \ge F_{a_2}^\top(r-p)$ holds, *i.e.*, the subset where ties in $\mathcal{B}(\theta,p)$ are broken in favor of a_1, which yields larger utility for the principal. We denote this subset by $\mathcal{P}_1$. Indeed, we show in Lemma 1 that $\textsc{Opt}$ can be rewritten as an optimization over $\mathcal{P}_1$ only.[4] The next observation is that, for a fixed $\theta \in [0,1]$, the utility of any contract in $\mathcal{P}_1$ can be rewritten as follows:

$$U^P(p,\theta) = \mathbb{1}\{\theta \le \tilde{\theta}_{\max}(\lambda_p)\} F_{a_1}^\top(r-p) + \mathbb{1}\{\theta > \tilde{\theta}_{\max}(\lambda_p)\} F_{a_2}^\top(r-p) \tag{1}$$

$$= \mathbb{1}\{\theta \le \tilde{\theta}_{\max}(\lambda_p)\}((F_{a_1} - F_{a_2})^\top r - \lambda_p) + F_{a_2}^\top r + \lambda_p - F_{a_1}^\top p, \tag{2}$$

where λ_p is a shorthand for $(F_{a_1} - F_{a_2})^\top p$ and $\tilde{\theta}_{\max}(\lambda) = \lambda C_{a_1}^{-1}$. The rationale behind this rewriting is that, when the contract is p, all types $\theta \le \tilde{\theta}_{\max}(\lambda_p)$ will best respond by playing a_1, while types larger than $\tilde{\theta}_{\max}(\lambda_p)$ will play a_2. We can already see that, to induce the agent to choose action a_1, the principal will need to increase the expected payment gap λ_p between the two actions. Equation (2) explicitly shows the dependence of the utility on λ_p. This is important as it allows us to reduce the problem to one dimension. Indeed, if we fix the expected difference in payment λ, Equation (2)

[3] In this case, the null contract is optimal for every agent $\theta \in [0,1]$. We prove this formally in Lemma 17.

[4] From an intuitive perspective, the reason is that for contracts that do not belong to $\mathcal{P}_1$, $F_{a_2}^\top(r-p) \ge F_{a_1}^\top(r-p)$ holds. Hence, for any θ, their utility is at most $F_{a_2}^\top(r-p) \le F_{a_2}^\top r$, which is always smaller or equal than the utility of the contract $\bar{p} = (0,\dots,0)$, which belongs to $\mathcal{P}_1$.

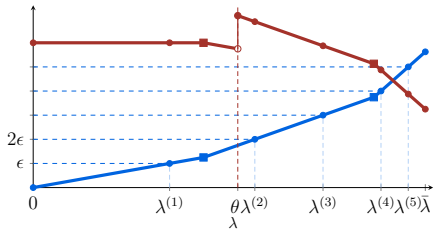


Figure 1: Visual representation of the instance given by 4 outcomes, $F_{a_1} = (\frac{1}{5}, \frac{1}{10}, \frac{1}{5}, \frac{1}{2})$, $F_{a_2} = (\frac{1}{2}, 0, \frac{1}{10}, \frac{2}{5})$, $r = (0, \frac{1}{2}, \frac{4}{5}, 1)$ and $C_{a_1} = 1$. In the plot we can see represented $f(\lambda)$ in **blue**, and $\lambda \mapsto U^P(\lambda, \theta)$ in **red** for $\theta = 0.12$ and $\epsilon = 0.08$. In the plot, we can also observe the type threshold θ where the agent switches from playing a_1 to a_2. Square markers indicate switches in the active constraints in the LP solution defining $f(\lambda)$, and therefore changes in its slope. The round marker indicates the result of the discretization Equation (5).

highlights how the optimal contract for that λ is simply the one that minimizes the expected payment to a_1, *i.e.*, the solution of the following linear program parametrized in λ:

$$f(\lambda) := \min_{p \in \mathcal{P}_1} F_{a_1}^\top p \quad \text{s.t.} \quad (F_{a_1} - F_{a_2})^\top p = \lambda. \tag{3}$$

Now we can rewrite the principal utility entirely as a function of λ and the value of the LP $f(\lambda)$, *i.e.*,

$$\widetilde{U}^P(\lambda, \theta) = \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda)\}((F_{a_1} - F_{a_2})^\top r - \lambda) + F_{a_2}^\top r + \lambda - f(\lambda).$$

In other words, we restricted the principal utility to a one-dimensional function where the λs of interests are those for which there exists $p \in \mathcal{P}_1$ that yields λ as the expected payment gap. We show in Lemma 3 that the only relevant range of λ is $[0, \bar{\lambda}]$ with $\bar{\lambda} := (F_{a_1} - F_{a_2})^\top r$. The upper endpoint is obtained from the definition of $\mathcal{P}_1$, while the lower endpoint is clipped to 0. The intuitive reason is that it is never convenient to incentivize a_2 over a_1, since a_1 is the rewarding action.

3.1.2 The Non-Uniform Discretization

At this point, ideally we would like to discretize $[0, \bar{\lambda}]$ with ϵ^{-1} points, and apply a no-regret adversarial bandit algorithm using as the action set the contracts yielding the discretized set of λs. To this end, we start by analyzing the utility difference when going from λ_1 to $\lambda_2 > \lambda_1$. Precisely, given type θ, one can show (Lemma 4) that

$$\widetilde{U}^P(\lambda_1, \theta) - \widetilde{U}^P(\lambda_2, \theta) \leq f(\lambda_2) - f(\lambda_1). \tag{4}$$

Observe that $f(\lambda)$ is the value of a linear program where λ is the right-hand side of the equality constraint. As a result $f(\lambda_2) - f(\lambda_1)$ cannot be directly upper bounded by $c|\lambda_2 - \lambda_1|$ for some universal constant $c \in \mathbb{R}$.[5] However, it is not hard to see that $f(\lambda)$ is a strictly increasing continuous function of λ, and is bounded in $[0, 1]$. As a consequence, we propose the following discretization of $[0, \bar{\lambda}]$ that meets our requirements. Let $\epsilon > 0$ and consider the following sequence. Set $\lambda^{(0)} = 0$ and for all $n \geq 1$ define $\lambda^{(n+1)}$ as follows:

$$\lambda^{(n+1)} = \begin{cases} \inf\{\lambda : \lambda \geq \lambda^{(n)}, f(\lambda) = f(\lambda^{(n)}) + \epsilon\} & \text{if } f(\lambda^{(n)}) + \epsilon < f(\bar{\lambda}) \\ \bar{\lambda} & \text{otherwise} \end{cases}. \tag{5}$$

We stop the sequence as soon as $\lambda^{(n)} = (F_{a_1} - F_{a_2})^\top r$ and denote by $\Lambda^{(\epsilon)}$ the collection of points in this sequence. By taking a closer look, we can see that Equation (5) generates points in $[0, \bar{\lambda}]$ in a way that the distance in $f(\cdot)$ between two subsequent points is always exactly ϵ.[6] Thus, by Equation (5) and Equation (4), we have that $\Lambda^{(\epsilon)}$ approximates the utility of any λ with accuracy ϵ. Furthermore, since f is bounded by $[0, 1]$ for $\lambda \in [0, \bar{\lambda}]$, we have $|\Lambda^{(\epsilon)}| = O(\epsilon^{-1})$. In Figure 1 we can see a visual representation of the non-uniform discretization described in this section.

Remark: computing $\Lambda^{(\epsilon)}$ is efficient. We observe that to compute the $(n+1)$-term of $\Lambda^{(\epsilon)}$ efficiently, we only need to solve the following linear feasibility problem in p and λ:

$$p \in \mathcal{P}_1, \lambda \in [\lambda^{(n)}, (F_{a_1} - F_{a_2})^\top r] \ \wedge \ (F_{a_1} - F_{a_2})^\top p = \lambda \ \wedge \ F_{a_1}^\top p = f(\lambda^{(n)}) + \epsilon. \tag{6}$$

[5] Note that this is what is needed by uniform discretization to guarantee that ϵ^{-1} points approximates OPT with precision ϵ.

[6] This is true except for the last point, where the distance can actually be smaller than ϵ.

When a solution exists, it is equivalent to the first case of Equation (5), and the point λ that satisfies Equation (6) is exactly $\lambda^{(n+1)}$. When it does not exist, it means $\lambda^{(n+1)} = (F_{a_1} - F_{a_2})^\top r$, and we stop the discretization procedure. Observe that the procedure that we described also yields a set of contracts whose expected payment gaps are exactly those in $\Lambda^{(\epsilon)}$. Indeed, these are the contracts satisfying Equation (6). In the following, we will denote these contracts with $\mathcal{P}^{(\epsilon)}$, and we recall that $|\mathcal{P}^{(\epsilon)}| = |\Lambda^{(\epsilon)}| \approx \epsilon^{-1}$.

3.1.3 Putting Everything Together

At this point, we just need to combine all the pieces together. First, we compute $\Lambda^{(\epsilon)}$ and the corresponding sets of contracts $\mathcal{P}^{(\epsilon)}$, and then we run the adversarial bandit algorithm of Audibert and Bubeck [3] using as arms the contracts in $\mathcal{P}^{(\epsilon)}$. This procedure, for any $\epsilon > 0$, guarantees that $R_T \leq \mathcal{O}(\sqrt{T|\mathcal{P}^{(\epsilon)}|} + \epsilon T)$, where the first term has to be interpreted as the regret of the principal against the best contract in $\mathcal{P}^{(\epsilon)}$ while the second one, instead, comes from the fact that contracts in $\mathcal{P}^{(\epsilon)}$ approximate the optimal constant fixed in hindsight up to accuracy ϵT (Lemma 6). Indeed, this is the key property which we obtained with the discretization described above. Then, since $|\mathcal{P}^{(\epsilon)}| \approx \epsilon^{-1}$, picking $\epsilon = T^{-1/3}$ yields the desired result.

3.2 A Matching $\Omega(T^{2/3})$ Lower Bound

In this section, we will prove that Theorem 1 is tight up to constants. In particular, we provide a lower bound which holds even for stochastic type generation. Importantly, we show that the lower bound holds even when the type distribution is absolutely continuous (with respect to Lebesgue measure).

The main difference from standard Multi-Armed Bandit lower bounds is that the feedback received by the learner and the reward suffered by the learner do not coincide, and that the action space of the learner is continuous. This is a characteristic of lower bounds for similar economic models, such as dynamic pricing [40, 5], bilateral trade [25, 13], and market making [24]. A principled way to deal with such problems is to reduce them to an instance with a finite set of "arms" and apply standard KL chain rule techniques (see [47, 41]). This is done by carefully identifying arms that are optimal with respect to regret or information acquired. However, this would require introducing point masses into the distribution. Instead, here, to produce an absolutely continuous distribution, we do not reduce to the finite case and use the KL chain rule with continuous actions.

More formally, we create $K = \Theta(1/\epsilon)$ instances $\{\mathfrak{I}_1, \dots, \mathfrak{I}_K\}$ such that

- each instance $\mathfrak{I}_k$ is identical to a base instance $\mathfrak{I}_0$, apart from an interval J_k of size $O(\epsilon)$;
- the principal utility on the interval J_k is $\Omega(\epsilon)$ larger than the utility outside J_k;
- the Kullback-Leibler divergence between the instance $\mathfrak{I}_0$ and instance $\mathfrak{I}_k$ is less than $O(\epsilon^2)$ times the number of times we played in expectation (under the base instance) in the interval J_k.

Then, by standard change of measure arguments, we can prove that for each algorithm, there is an instance in which the regret is at least $\Omega(\epsilon T - \epsilon^2\sqrt{T/K})$, and by choosing $K = \Theta(1/\epsilon)$ and calibrating $\epsilon = \Theta(T^{-1/3})$ gives the desired $\Omega(T^{2/3})$ lower-bound (the full proof is in Section B).

Theorem 2. *For any algorithm, there exists an instance with stochastic types such that the expected regret is lower bounded by $\Omega(T^{2/3})$.*

4 Improved Rates Against a Single Hidden Type

In this section, we discuss how we can improve the theoretical guarantees when the principal is repeatedly playing against a fixed, but unknown, type $\bar{\theta}$. In Section 4.1 we propose an algorithm that attains $\widetilde{O}(\sqrt{T})$ regret and in Section 4.2 we prove a $\Omega(\sqrt{T})$ lower bound.

4.1 An Explore-and-Commit Strategy for $\widetilde{O}(\sqrt{T})$ Regret

We first state the main result of this section and then we explain how it is obtained.

Algorithm 1 Regret minimization against a single type

Require: Threshold β, robustification α, number of phases $L \in \mathbb{N}$, samples per phase $n \in \mathbb{N}$

if $\|F_{a_1} - F_{a_2}\|_1 \le \beta$ **then**
  Play $p_t = (0, \dots, 0)$ for all $t \in [T]$
else
  if $\theta_{\text{lazy}} < 1$ **then** ▷ Test if $\bar{\theta} > \theta_{\text{lazy}}$
    Play any $p^\star_{\theta_{\text{lazy}}}$ for n rounds; observe $\{\omega_i^{(0)}\}_{i=1}^n$; compute $Z_0 \coloneqq \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\omega_i^{(0)} \in \Omega_1\}$
    if $Z_0 \le \tau$ **then**
      Play $\bar{p} = (0, \dots, 0)$ for all the remaining rounds
  Set $s_1 = 0$, $e_1 = \theta_{\text{lazy}}$, $\hat{\theta}_1 = (e_1 + s_1)/2$ ▷ Binary search in $[0, \theta_{\text{lazy}}]$
  for $l \in \{1, \dots, L-1\}$ **do**
    Play $p^\star_{\hat{\theta}_l}$ for n rounds; observe $\{\omega_i^{(l)}\}_{i=1}^n$; compute $Z_l \coloneqq \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\omega_i^{(l)} \in \Omega_1\}$
    if $Z_l \le \tau$ **then**
      Set $s_{l+1} = \hat{\theta}_l$, $e_{l+1} = e_l$, $\hat{\theta}_{l+1} = (e_{l+1} + s_{l+1})/2$
    else
      Set $s_{l+1} = s_l$, $e_{l+1} = \hat{\theta}_l$, $\hat{\theta}_{l+1} = (e_{l+1} + s_{l+1})/2$
  Compute $p^\star_{\hat{\theta}_L}$ and play $\tilde{p} = p^\star_{\hat{\theta}_L} + \alpha(r - p^\star_{\hat{\theta}_L})$ until the end of the horizon ▷ Exploit

Theorem 3. *There exists an algorithm such that $R_T \le \widetilde{O}(\sqrt{T})$ for any sequence $\{\bar{\theta}\}_{t=1}^T$ where $\bar{\theta}$ is any unknown type in $[0, 1]$.*

Outline. We start by observing that, if the principal knew the hidden type $\bar{\theta}$, then it could simply play an optimal contract for that type. Suppose, for the moment, that an oracle provided an ϵ-accurate estimate $\hat{\theta}$ of $\bar{\theta}$. However, simply playing the optimal contract $p^\star_{\hat{\theta}}$ for $\hat{\theta}$ might be greatly suboptimal; indeed, even infinitesimal errors in the type $\hat{\theta}$ might change the action played by the agent and lead to very suboptimal outcomes. However, we can instead compute the optimal contract $p^\star_{\hat{\theta}}$ for type $\hat{\theta}$ and play its "robustified" version $\tilde{p}_{\hat{\theta}} = p^\star_{\hat{\theta}} + \alpha(r - p^\star_{\hat{\theta}})$ for some $\alpha > 0$. The role of robustification is to make the contract insensitive to small errors in the estimated type at the cost of only a small loss in utility. By the standard robustification argument of Dütting et al. [29] (see also Bernasconi et al. [14]), choosing $\alpha = \sqrt{\epsilon}$ guarantees $U^P(\tilde{p}_{\hat{\theta}}, \theta) \ge U^P(p^\star_{\hat{\theta}}, \theta) - 2\sqrt{\epsilon}$ for nearby types (see Lemma 19). Therefore, an ϵ-accurate estimate of $\bar{\theta}$ translates into $O(\sqrt{\epsilon})$ per-round exploitation regret. The goal of the algorithm is then to obtain such an estimate using a short exploration phase. To do this, we show how to build suitable threshold tests that we will use inside a stochastic binary search procedure. Crucially, the exploration contracts themselves are chosen so that each test incurs only small regret.

Testing a Threshold. Fix a contract p_θ such that all types $\theta' \le \theta$ choose a_1 and all types $\theta' > \theta$ choose a_2. Playing p_θ for n rounds yields samples from F_{a_1} if $\bar{\theta} \le \theta$ and from F_{a_2} otherwise. We distinguish these two cases using the statistic

$$Z = \frac{1}{n}\sum_{i=[n]} \mathbb{1}\{\omega_i \in \Omega_1\}, \qquad \Omega_1 = \{\omega \in \Omega : F_{a_1,\omega} > F_{a_2,\omega}\}.$$

We compare Z with $\tau = \sum_{\omega\in\Omega_1}(F_{a_1,\omega} + F_{a_2,\omega})/2$. If $\bar{\theta} > \theta$, then Z concentrates below τ, and otherwise above τ. We show the error probability of this test decays as $O(\exp(-n\|F_{a_1} - F_{a_2}\|_1^2))$.[7]

Exploring with Low-Regret Contracts. The contracts used for threshold tests must satisfy two properties: 1) they must not incur large regret, and 2) they must reveal whether $\bar{\theta}$ lies above or below the tested threshold. Consider any contract p, changing the type from θ to θ' can only change the principal's utility if the induced action switches between a_1 and a_2. In that case, the loss is controlled by the distance between the two outcome distributions, *i.e.*, $|U^P(p, \theta) - U^P(p, \theta')| \le O(\|F_{a_1} - F_{a_2}\|_1)$. Consequently, the optimal value function is also $O(\|F_{a_1} - F_{a_2}\|_1)$-stable across types. Indeed, if $p^\star_{\theta'}$ is optimal for type θ', then

$$\max_{p\in[0,1]^m} U^P(p, \bar{\theta}) - U^P(p^\star_{\theta'}, \bar{\theta}) \le O(\|F_{a_1} - F_{a_2}\|_1).$$

[7] Note that the rate is independent of m, unlike a test based on estimating the full outcome distribution.

Hence, even if $p^\star_{\theta'}$ is not optimal for the true type $\bar{\theta}$, its per-round regret against $\bar{\theta}$ is at most $O(\|F_{a_1} - F_{a_2}\|_1)$. Thus, property (1) is satisfied. Now, let θ_{lazy} be the largest type for which it is still optimal for the principal to incentivize the high-effort action a_1 rather than induce the low-effort action a_2 with the null contract.[8] For every $\theta > \theta_{\text{lazy}}$, it is optimal to induce a_2, and the null contract $\bar{p} = (0, \dots, 0)$ is optimal. This case is therefore easy: it is sufficient to detect that $\bar{\theta} > \theta_{\text{lazy}}$ and commit to $\bar{p}$ without needing to estimate $\bar{\theta}$ more accurately. For every $\theta \le \theta_{\text{lazy}}$, instead, the optimal contract $p^\star_\theta$ can be used for exploration because it implements a sharp threshold at θ. Indeed, $p^\star_\theta$ is the cheapest contract that induces type θ to play a_1, and hence satisfies $(F_{a_1} - F_{a_2})^\top p^\star_\theta = \theta C_{a_1}$. Then, any for type $\theta' \le \theta$ action a_1 is IC, and it is not IC for $\theta' > \theta$. Thus, when $p^\star_\theta$ is played, the observed outcomes are drawn from F_{a_1} if $\bar{\theta} \le \theta$ and from F_{a_2} if $\bar{\theta} > \theta$. This makes $p^\star_\theta$ an informative threshold test for binary search, and thus, property (2) is satisfied.

Algorithm. Algorithm 1 first handles the "nearly indistinguishable" case $\|F_{a_1} - F_{a_2}\|_1 \le \beta$ (Line 2). Since the null contract is then suboptimal by at most $O(\|F_{a_1} - F_{a_2}\|_1)$ per round, playing it throughout gives regret $O(\beta T)$, and by choosing $\beta = T^{-1/2}$ we have the desired $\sqrt{T}$ bound. When $\|F_{a_1} - F_{a_2}\|_1 > \beta$, the algorithm first tests whether $\bar{\theta} > \theta_{\text{lazy}}$ by playing $p^\star_{\theta_{\text{lazy}}}$ for n rounds. If the test indicates $\bar{\theta} > \theta_{\text{lazy}}$, it commits to the null contract, which is optimal for all such types (Line 7). Otherwise, it performs $L - 1$ phases of stochastic binary search on $[0, \theta_{\text{lazy}}]$. In phase l, the midpoint $\hat{\theta}_l$ is tested by playing $p^\star_{\hat{\theta}_l}$ for n rounds and comparing the statistic Z_l with τ (Line 10). After L phases, the algorithm obtains an estimate $\hat{\theta}_L$ with error at most 2^{-L} on the event that all tests are correct. It then commits to the robustified contract $\tilde{p} = p^\star_{\hat{\theta}_L} + \alpha(r - p^\star_{\hat{\theta}_L})$, this ensures that we are playing a contract that actually yields good utility for types close to $\hat{\theta}_L$.

Proof Sketch. We sketch the regret bound for the main case $\bar{\theta} \le \theta_{\text{lazy}}$ and $\|F_{a_1} - F_{a_2}\|_1 \ge T^{-1/2}$, the other cases are simpler (the complete proofs can be found in Section C). During the exploration phase, the algorithm plays either the null contract or contracts $p^\star_\theta$ that are optimal for some type. By the suboptimality bound above, the regret accumulated during exploration is $O(nL\|F_{a_1} - F_{a_2}\|_1)$. For the exploitation phase, let $\mathcal{E} := \{\bar{\theta} \in [s_L, e_L]\}$ be the event that the binary search has localized the true type. On $\mathcal{E}$, we have $|\bar{\theta} - \hat{\theta}_L| \le 2^{-L}$, and robustification yields per-round regret of $O(2^{-L}\alpha^{-1} + \alpha)$. On the complement $\mathcal{E}^\complement$, a union bound over the L threshold tests gives $\mathbb{P}(\mathcal{E}^\complement) \le O(L\exp(-n\|F_{a_1} - F_{a_2}\|_1^2))$. Even on $\mathcal{E}^\complement$, the committed contract is close to an optimal contract for some type, so the per-round regret is bounded by $O\left((\|F_{a_1} - F_{a_2}\|_1 + \alpha)L\exp(-n\|F_{a_1} - F_{a_2}\|_1^2)\right)$. Combining the two terms we obtain the following regret upper bound

$$R_T = O\left(nL\|F_{a_1} - F_{a_2}\|_1 + T\left[2^{-L}\alpha^{-1} + \alpha + (\|F_{a_1} - F_{a_2}\|_1 + \alpha)L\exp(-n\|F_{a_1} - F_{a_2}\|_1^2)\right]\right).$$

Choosing $\alpha \approx \sqrt{2^{-L}}$, $L \approx \log T$, $n \approx \|F_{a_1} - F_{a_2}\|_1^{-2}\log(T\|F_{a_1} - F_{a_2}\|_1^2)$ proves Theorem 3.

4.2 $\Omega(\sqrt{T})$ is Unavoidable Against a Fixed Type

We complement the upper bound with a matching lower bound. We construct instances that allow us to reduce the problem to a multi-armed bandits with two arms. Precisely, we construct two instances such that, in each instance, the optimal arm has $\Omega(\epsilon)$ higher utility than it does in the other, while the KL divergence between the feedback distributions of the two instances is at most $O(\epsilon^2)$. A change-of-measure argument then yields the result. The complete proof is reported in Section D.

Theorem 4. *Any (possibly randomized) algorithm suffers at least regret $R_T = \Omega(\sqrt{T})$.*

5 Future Work

Several questions on the learnability of contracts remain open. A first direction is to strengthen lower bounds for more general online contract-design settings, especially by constructing hard instances whose dependence on the instance size is only polynomial. Another important direction is to move beyond the binary-action case: with multiple actions, the type space induces a richer arrangement

[8] We defer the exact expression of θ_{lazy} to Equation (18) in Section C. Here, we only mention that it is the solution of a linear program, and thus it can be computed efficiently.

of incentive regions, and it is unclear whether similar outcome-independent regret guarantees are possible. More broadly, it would be interesting to study intermediate regimes between adversarial and fixed types, such as slowly drifting types.

Acknowledgments and Disclosure of Funding

This work was partially funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

This work is supported by an ERC grant (Project 101165466 — PLA-STEER)

References

[1] Tal Alon, Paul Dütting, and Inbal Talgam-Cohen. Contracts with private cost per unit-of-effort. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 52–69, 2021.

[2] Tal Alon, Paul Duetting, Yingkai Li, and Inbal Talgam-Cohen. Bayesian analysis of linear contracts. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC '23, page 66, 2023.

[3] Jean-Yves Audibert and Sébastien Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT - The 22nd Conference on Learning Theory*, 2009.

[4] Moshe Babaioff, Michal Feldman, Noam Nisan, and Eyal Winter. Combinatorial agency. *J. Econ. Theory*, 147(3):999–1034, 2012.

[5] Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg, and Aleksandrs Slivkins. Dynamic pricing with limited supply, 2015.

[6] Francesco Bacchiocchi, Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Learning optimal contracts: How to exploit small action spaces. In *The Twelfth International Conference on Learning Representations*, 2023.

[7] Francesco Bacchiocchi, Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Regret minimization for piecewise linear rewards: Contracts, auctions, and beyond. *arXiv preprint arXiv:2503.01701*, 2025.

[8] Maria-Florina Balcan, Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Commitment without regrets: Online learning in stackelberg security games. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 61–78, 2015.

[9] Maria-Florina Balcan, Martino Bernasconi, Matteo Castiglioni, Andrea Celli, Keegan Harris, and Zhiwei Steven Wu. Nearly-optimal bandit learning in stackelberg games with side information. *arXiv preprint arXiv:2502.00204*, 2025.

[10] Hamsa Bastani, Mohsen Bayati, Mark Braverman, Ramki Gummadi, and Ramesh Johari. Analysis of medicare pay-for-performance contracts. *Available at SSRN 2839143*, 2016.

[11] Hamsa Bastani, Joel Goh, and Mohsen Bayati. Evidence of upcoding in pay-for-performance programs. *Management Science*, 65(3):1042–1060, 2019.

[12] Martino Bernasconi, Matteo Castiglioni, Andrea Celli, Alberto Marchesi, Francesco Trovò, and Nicola Gatti. Optimal rates and efficient algorithms for online bayesian persuasion. In *International Conference on Machine Learning*, pages 2164–2183. PMLR, 2023.

[13] Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. No-regret learning in bilateral trade via global budget balance. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 247–258, 2024.

[14] Martino Bernasconi, Matteo Castiglioni, and Alberto Marchesi. Regret-minimizing contracts: Agency under uncertainty. *arXiv preprint arXiv:2402.13156*, 2024.

[15] Martino Bernasconi, Matteo Castiglioni, and Andrea Celli. Single-dimensional contract design: Efficient algorithms and learning. *arXiv preprint arXiv:2502.11661*, 2025.

[16] Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Learning optimal commitment to overcome insecurity. *Advances in Neural Information Processing Systems*, 27, 2014.

[17] Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. In *Summer school on machine learning*, pages 208–240. Springer, 2003.

[18] Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou. Optimum statistical estimation with strategic data sources. In *Conference on Learning Theory*, pages 280–296. PMLR, 2015.

[19] Matteo Castiglioni, Andrea Celli, Alberto Marchesi, and Nicola Gatti. Online bayesian persuasion. *Advances in neural information processing systems*, 33:16188–16198, 2020.

[20] Matteo Castiglioni, Alberto Marchesi, Andrea Celli, and Nicola Gatti. Multi-receiver online bayesian persuasion. In *International Conference on Machine Learning*, pages 1314–1323. PMLR, 2021.

[21] Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Bayesian agency: Linear versus tractable contracts. *Artificial Intelligence*, 307, 2022.

[22] Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Designing menus of contracts efficiently: The power of randomization. *Artificial Intelligence*, 318:103881, 2023.

[23] Matteo Castiglioni, Junjie Chen, Minming Li, Haifeng Xu, and Song Zuo. A reduction from multi-parameter to single-parameter bayesian contract design. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1795–1836. SIAM, 2025.

[24] Nicolo Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Luigi Foscari, and Vinayak Pathak. Market making without regret. *arXiv preprint arXiv:2411.13993*, 2024.

[25] Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Regret analysis of bilateral trade with a smoothed adversary. *Journal of Machine Learning Research*, 25(234):1–36, 2024.

[26] Yurong Chen, Zhaohua Chen, Xiaotie Deng, and Zhiyi Huang. Are bounded contracts learnable and approximately optimal? In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 315–344, 2024.

[27] Alon Cohen, Argyrios Deligkas, and Moran Koren. Learning approximately optimal contracts. In *Algorithmic Game Theory: 15th International Symposium, SAGT 2022, Colchester, UK, September 12–15, 2022, Proceedings*, page 331–346, Berlin, Heidelberg, 2022. Springer-Verlag.

[28] Lin William Cong and Zhiguo He. Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797, 2019.

[29] Paul Dütting, Tim Roughgarden, and Inbal Talgam-Cohen. Simple versus optimal contracts. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 369–387, 2019.

[30] Paul Dutting, Tim Roughgarden, and Inbal Talgam-Cohen. The complexity of contracts. *SIAM Journal on Computing*, 50(1):211–254, 2021.

[31] Paul Dütting, Tomer Ezra, Michal Feldman, and Thomas Kesselheim. Multi-agent contracts. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1311–1324, 2023.

[32] Paul Dütting, Guru Guruganesh, Jon Schneider, and Joshua Ruizhi Wang. Optimal no-regret learning for one-sided lipschitz functions. In *International Conference on Machine Learning*, pages 8836–8850. PMLR, 2023.

[33] Paul Dütting, Michal Feldman, Inbal Talgam-Cohen, et al. Algorithmic contract theory: A survey. *Foundations and Trends® in Theoretical Computer Science*, 16(3-4):211–412, 2024.

[34] Paul Dütting, Michal Feldman, Tomasz Ponitka, and Ermis Soumalias. The pseudo-dimension of contracts. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pages 514–539, 2025.

[35] Michal Feldman. Combinatorial contract design: Recent progress and emerging frontiers. *arXiv preprint arXiv:2510.15065*, 2025.

[36] Guru Guruganesh, Jon Schneider, and Joshua R Wang. Contracts under moral hazard and adverse selection. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 563–582, 2021.

[37] Dylan Hadfield-Menell and Gillian K Hadfield. Incomplete contracting and ai alignment. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 417–422, 2019.

[38] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research*, 55:317–359, 2016.

[39] Nur Kaynar and Auyon Siddiq. Estimating effects of incentive contracts in online labor platforms. *Management Science*, 69(4):2106–2126, 2023.

[40] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

[41] Tor Lattimore, Csaba Szepesvari, and Gellert Weisz. Learning with good feature representations in bandits and in rl with a generative model. In *International conference on machine learning*, pages 5662–5670. PMLR, 2020.

[42] Joshua Letchford, Vincent Conitzer, and Kamesh Munagala. Learning and approximating the optimal strategy to commit to. In *Algorithmic Game Theory: Second International Symposium, SAGT 2009*, pages 250–262. Springer, 2009.

[43] Binghui Peng, Weiran Shen, Pingzhong Tang, and Song Zuo. Learning optimal strategies to commit to. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2149–2156, 2019.

[44] Eden Saig, Inbal Talgam-Cohen, and Nir Rosenfeld. Delegated classification. *Advances in Neural Information Processing Systems*, 36:13200–13236, 2023.

[45] Eden Saig, Ohad Einav, and Inbal Talgam-Cohen. Incentivizing quality text generation via statistical contracts. *Advances in Neural Information Processing Systems*, 37:51196–51222, 2024.

[46] Eden Saig, Tamar Garbuz, Ariel D Procaccia, Inbal Talgam-Cohen, and Jamie Tucker-Foltz. Adaptive contracts for cost-effective ai delegation. *arXiv preprint arXiv:2603.17212*, 2026.

[47] Aleksandrs Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

[48] Ander Artola Velasco, Stratis Tsirtsis, Nastaran Okati, and Manuel Gomez-Rodriguez. Is your llm overcharging you? tokenization, transparency, and incentives. *arXiv preprint arXiv:2505.21627*, 2025.

[49] Banghua Zhu, Stephen Bates, Zhuoran Yang, Yixin Wang, Jiantao Jiao, and Michael I Jordan. The sample complexity of online contract design. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 1188–1188, 2023.

A Proof of Theorem 1

In this section we prove Theorem 1. This section is structured as follows. First, in Section A.1, we show that, to reach optimality, we can focus on a one-dimensional problem. Then, in Section A.2, we exploit this result using an ad-hoc non uniform discretization. Finally, in Section A.3, we combine these results and prove Theorem 1.

A.1 A one-dimensional problem

For convenience, we define $\mathcal{P}_1, \mathcal{P}_2$ as follows:

$$\mathcal{P}_1 = \{p \in [0,1]^m : (F_{a_1} - F_{a_2})^\top p \leq (F_{a_1} - F_{a_2})^\top r\}$$
$$\mathcal{P}_2 = \{p \in [0,1]^m : (F_{a_1} - F_{a_2})^\top p \geq (F_{a_1} - F_{a_2})^\top r\}.$$

Then, in the following Lemma, we first show that OPT can be reduced to an optimization problem over $\mathcal{P}_1$.

Lemma 1 (Optimize only over $\mathcal{P}_1$)**.** *Suppose that $F_{a_1}^\top r > F_{a_2}^\top r$, then, for any sequence $\{\theta_t\}_{t=1}^T$ it holds that:*

$$\textsc{Opt} = \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p)).$$

Proof. Recall that $\mathcal{P}_2 = \{p \in [0,1]^m : F_{a_2}^\top (r-p) \geq F_{a_1}^\top (r-p)\}$. Thus, take any $p \in \mathcal{P}_2$ and any $\theta \in [0,1]$ it holds that:

$$U^P(p, b(\theta, p)) = \mathbb{1}\{a_1 \in b(\theta, p)\} F_{a_1}^\top (r-p) + \mathbb{1}\{b(\theta, p) = a_2\} F_{a_2}^\top (r-p) \leq F_{a_2}^\top r. \tag{7}$$

Next, observe that $\bar{p} = (0, \dots, 0)$ belongs to $\mathcal{P}_1$ since $F_{a_1}^\top (r - \bar{p}) = F_{a_1}^\top r > F_{a_2}^\top r = F_{a_2}^\top (r - \bar{p})$ and furthermore:

$$U^P(\bar{p}, b(\theta, \bar{p})) = \mathbb{1}\{a_1 \in b(\theta, \bar{p})\} F_{a_1}^\top r + \mathbb{1}\{b(\theta, \bar{p}) = a_2\} F_{a_2}^\top r > F_{a_2}^\top r. \tag{8}$$

Chaining these results, we have that:

$$\begin{aligned}
\textsc{Opt} &= \sup_{p \in [0,1]^m} \sum_{t=1}^T U^P(p, b(\theta_t, p)) \\
&= \max\left\{ \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p)), \sup_{p \in \mathcal{P}_2} \sum_{t=1}^T U^P(p, b(\theta_t, p)) \right\} \\
&\leq \max\left\{ \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p)), F_{a_2}^\top r \right\} && \text{(Equation (7))} \\
&= \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p)) && \text{(Equation (8))} \\
&\leq \textsc{Opt}.
\end{aligned}$$

Hence, it must hold that $\textsc{Opt} = \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p))$, which concludes the proof. □

Next, we start by rewriting the utility of contracts that belong to $\mathcal{P}_1$.

Lemma 2 (Utility rewriting)**.** *Consider $p \in \mathcal{P}_1$ and let $\lambda_p = (F_{a_1} - F_{a_2})^\top p$. Then, for any $\theta \in [0,1]$ it holds that:*

$$U^P(p, b(\theta, p)) = \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_p)\} \left((F_{a_1} - F_{a_2})^\top r - \lambda_p\right) + F_{a_2}^\top r + \lambda_p - F_{a_1}^\top p.$$

Proof. Observe that $a_1 \in \mathcal{B}(\theta, p)$ holds if $F_{a_1}^\top p - \theta C_{a_1} \geq F_{a_2}^\top p$, *i.e.*, if $\theta \leq \tilde{\theta}_{\max}(\lambda_p)$. At equality, ties break in favor of the principal and therefore, since $p \in \mathcal{P}_1$, at $\theta = \tilde{\theta}_{\max}(\lambda_p)$ the agent will play

action a_1. Thus, we have that:

$$\begin{aligned}
U^P(p, b(\theta, p)) &= \mathbb{1}\{a_1 \in b(\theta, p)\} F_{a_1}^\top (r - p) + \mathbb{1}\{a_2 = b(\theta, p)\} F_{a_2}^\top (r - p) \\
&= \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_p)\} F_{a_1}^\top (r - p) + \mathbb{1}\{\theta > \tilde{\theta}_{\max}(\lambda_p)\} F_{a_2}^\top (r - p) \\
&= \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_p)\} \left(F_{a_1}^\top (r - p) - F_{a_2}^\top (r - p)\right) + F_{a_2}^\top (r - p) \\
&= \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_p)\} \{(F_{a_1} - F_{a_2})^\top r - \lambda_p\} + F_{a_2}^\top r + \lambda_p - F_{a_1}^\top p,
\end{aligned}$$

which concludes the proof. □

We now make an important remark that is a consequence of Lemma 2. Before that, we need to introduce some additional quantities. First, given $\lambda \in \mathbb{R}$, we recall the definition of $f(\lambda)$ as the expected payment to a_1 of cheapest contract having an expected payment gap between the two actions of λ as a solution to the following LP. More precisely, we define $f(\lambda)$ as follows

$$f(\lambda) \coloneqq \left\{ \begin{array}{ll} \min\limits_{p \in \mathcal{P}_1} & F_{a_1}^\top p \\ \text{s.t.} & (F_{a_1} - F_{a_2})^\top p = \lambda \end{array} \right. \tag{9}$$

Then, for $\lambda \in \mathbb{R}$ and $\theta \in [0, 1]$, we define the expected utility of the principal as a function of λ as

$$\widetilde{U}^P(\lambda, \theta) \coloneqq \mathbb{1}\{\theta_t \leq \tilde{\theta}_{\max}(\lambda)\}[(F_{a_1} - F_{a_2})^\top r - \lambda] + F_{a_2}^\top r + \lambda - f(\lambda),$$

and we let

$$\Lambda_1 \coloneqq \{\lambda \in \mathbb{R} : \exists p \in \mathcal{P}_1 \wedge (F_{a_1} - F_{a_2})^\top p = \lambda\}$$

be the set of expected payment gaps that can be induced by contracts in $\mathcal{P}_1$. Then, Lemma 2 directly implies that, for any $\lambda \in \Lambda_1$:

$$\min_{p \in \mathcal{P}_1 : (F_{a_1} - F_{a_2})^\top p = \lambda} U^P(p, b(\theta, p)) = \widetilde{U}^P(\lambda, \theta).$$

This simple observation is important to obtain the following key lemma. This lemma allows us to turn the original optimization problem in the m-dimensional space of contracts into a one-dimensional problem over λ.

Lemma 3 (From p's to λ's)**.** *Suppose that $F_{a_1}^\top r > F_{a_2}^\top r$. For any sequence of types $\{\theta_t\}_{t=1}^T$, we have that:*

$$\textsc{Opt} = \max_{\lambda \in [0, (F_{a_1} - F_{a_2})^\top r]} \sum_{t=1}^T \widetilde{U}^P(\lambda, \theta_t). \tag{10}$$

Proof. First, we observe that

$$\begin{aligned}
\Lambda_1 &= \{\lambda \in \mathbb{R} : \exists p \in \mathcal{P}_1 \wedge (F_{a_1} - F_{a_2})^\top p = \lambda\} \\
&= \{\lambda \in \mathbb{R} : \exists p \in [0, 1]^m \wedge F_{a_1}^\top (r - p) \geq F_{a_2}^\top (r - p) \wedge (F_{a_1} - F_{a_2})^\top p = \lambda\} \\
&= \{\lambda \in \mathbb{R} : \exists p \in [0, 1]^m \wedge (F_{a_1} - F_{a_2})^\top r \geq \lambda \wedge (F_{a_1} - F_{a_2})^\top p = \lambda\} \\
&= \left[\sum_{\omega : F_{a_2, \omega} > F_{a_1, \omega}} F_{a_1, \omega} - F_{a_2, \omega}, (F_{a_1} - F_{a_2})^\top r \right].
\end{aligned}$$

Then, we can rewrite the optimal value that can be extracted by the principal as follows:

$$\begin{aligned}
\textsc{Opt} &= \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T U^P(p, b(\theta_t, p)) && \text{(Lemma 1)} \\
&= \sup_{p \in \mathcal{P}_1} \sum_{t=1}^T \widetilde{U}^P((F_{a_1} - F_{a_2})^\top p, \theta_t) \\
&= \sup_{\lambda \in \Lambda_1} \sum_{t=1}^T \widetilde{U}^P(\lambda, \theta_t),
\end{aligned}$$

where the second equality holds by the remark above. At this point, consider $\lambda \in \Lambda_1$ such that $\lambda < 0$. Then, for any $\theta \in [0, 1]$, it holds that:

$$\begin{aligned}
\widetilde{U}^P(\lambda, \theta) &= \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda)\}[(F_{a_1} - F_{a_2})^\top r - \lambda] + F_{a_2}^\top r + \lambda - f(\lambda) \\
&= F_{a_2}^\top r + \lambda - f(\lambda) && (\tilde{\theta}_{\max}(\lambda) < 0) \\
&\leq F_{a_2}^\top r. && (\lambda < 0 \text{ and } f(\lambda) \geq 0)
\end{aligned}$$

Furthermore, at $\bar{\lambda} = 0$, we have that:

$$\begin{aligned}
\widetilde{U}^P(\bar{\lambda}, \theta) &= \mathbb{1}\{\theta = 0)[(F_{a_1} - F_{a_2})^\top r - \bar{\lambda}\} + F_{a_2}^\top r + \bar{\lambda} - f(\bar{\lambda}) \\
&= \mathbb{1}\{\theta = 0)[(F_{a_1} - F_{a_2})^\top r\} + F_{a_2}^\top r && (f(\bar{\lambda}) = 0 \text{ and } \bar{\lambda} = 0) \\
&\geq F_{a_2}^\top r.
\end{aligned}$$

Thus, for all $\lambda < 0$ and any θ, $\widetilde{U}^P(\lambda, \theta) \leq \widetilde{U}^P(\bar{\lambda}, \theta)$. Therefore, we obtained that:

$$\textsc{Opt} = \sup_{\lambda \in \Lambda_1} \sum_{t=1}^{T} \widetilde{U}^P(\lambda, \theta_t) = \sup_{\lambda \in [0, (F_{a_1} - F_{a_2})^\top r]} \sum_{t=1}^{T} \widetilde{U}^P(\lambda, \theta_t).$$

Finally, we prove that the sup is attained. To this end, fix $t \in [T]$ and observe that $\lambda \to \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda)\}$ is upper semicontinuous for any θ. [9] Furthermore, the product of two positive upper semi-continuous functions is upper semicontinuous, and hence $\mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda)\}\left[(F_{a_1} - F_{a_2})^\top r - \lambda\right]$ is upper semicontinuous since $F_{a_1}^\top r > F_{a_2}^\top r$ and $\lambda \in [0, (F_{a_1} - F_{a_2})^\top r]$. Moreover, the sum of two upper semicontinuous functions is upper semicontinuous and thus $\sum_{t=1}^{T} \widetilde{U}^P(\lambda, \theta_t)$ is upper semicontinuous as well. Any upper semicontinuous function on a compact domain attains a maximum on that domain (Theorem 6), and therefore the sup is attained. This concludes the proof. □

Finally, we prove two important properties of the utility as a function of λ. First, the next lemma shows that the variation of the one-dimensional objective $\widetilde{U}^P(\lambda, \theta)$ as λ changes is controlled by the variation of the function $f(\lambda)$. As we discussed in Section 3.1, this property is crucial in the discretization argument.

Lemma 4 (Objective Function Differences)**.** *Consider $\lambda_2 \geq \lambda_1 \geq 0$ and any $\theta \in [0, 1]$. It holds that:*

$$\widetilde{U}^P(\lambda_1, \theta) - \widetilde{U}^P(\lambda_2, \theta) \leq f(\lambda_2) - f(\lambda_1).$$

Proof. The proof follows by using the definition of $\widetilde{U}^P(\lambda, \theta)$. Specifically:

$$\begin{aligned}
\widetilde{U}^P(\lambda_1, \theta) - \widetilde{U}^P(\lambda_2, \theta) = \lambda_1 - \lambda_2 + f(\lambda_2) - f(\lambda_1) \\
+ \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_1)\}\left[(F_{a_1} - F_{a_2})^\top r - \lambda_1\right] \\
- \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_2)\}\left[(F_{a_1} - F_{a_2})^\top r - \lambda_2\right].
\end{aligned}$$

Now, since $\lambda_1 \leq \lambda_2$, we have that $\mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_1)\} \leq \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_2)\}$, thus:

$$\begin{aligned}
\widetilde{U}^P(\lambda_1, \theta) - \widetilde{U}^P(\lambda_2, \theta) &\leq \lambda_1 - \lambda_2 + f(\lambda_2) - f(\lambda_1) \\
&\quad + \mathbb{1}\{\theta \leq \tilde{\theta}_{\max}(\lambda_2)\}\left[\lambda_2 - \lambda_1\right] \\
&\leq f(\lambda_2) - f(\lambda_1).
\end{aligned}$$

which concludes the proof. □

Finally, we conclude by highlighting some properties of $f(\lambda)$ that we use in our discretization argument in the next section.

Lemma 5 (Characterization of $f(\lambda)$)**.** *Consider $f(\lambda)$ and suppose $F_{a_1}^\top r > F_{a_2}^\top r$. Then, it holds that $f(0) = 0$. Moreover $f(\lambda) \leq 1$ for all λ's for which the problem is feasible. Finally, $f(\lambda)$ is well-defined, continuous and strictly increasing in the interval $[0, (F_{a_1} - F_{a_2})^\top r]$.*

[9] See Section E for the definition of upper semicontinuity.

Proof. At 0, $f(0)$ reduces to the following optimization problem:

$$\min_{p \in \mathcal{P}_1} F_{a_1}^\top p \quad \text{s.t. } (F_{a_1} - F_{a_2})^\top p = 0. \tag{11}$$

Observe that, since $p \geq 0$ and $F_{a_1} \in \Delta(\Omega)$, $f(\lambda)$ is always greater than 0. Furthermore, since $F_{a_1}^\top r > F_{a_2}^\top r$, the null contract $\bar{p} = (0, \dots, 0)$ belongs to $\mathcal{P}_1$. Indeed:

$$(F_{a_1} - F_{a_2})^\top \bar{p} = 0 < (F_{a_1} - F_{a_2})^\top r.$$

Since the null contract belongs to the feasible region of (11) and since it yields 0 as objective function, it is optimal.

Next, suppose that the optimization problem that defines $f(\lambda)$ admits a non-empty feasible region. Then, for any $p \in \mathcal{P}_1$, $F_{a_1}^\top p \leq 1$. Thus, $f(\lambda)$ is always bounded by 1.

Finally, let $\lambda \in [0, (F_{a_1} - F_{a_2})^\top r]$ and observe that the feasible region of (9) is non-empty. This is guaranteed by the fact that $p \in \mathcal{P}_1$ if $p \in [0,1]^m$ and $(F_{a_1} - F_{a_2})^\top p \leq (F_{a_1} - F_{a_2})^\top r$. Thus, for any $\lambda \in [0, (F_{a_1} - F_{a_2})^\top r]$, we can always find p such that $(F_{a_1} - F_{a_2})^\top p = \lambda$. Next, we show that f is continuous. To this end, we show that f is convex. Consider $\lambda_1, \lambda_2 \in [0, (F_{a_1} - F_{a_2})^\top r]$ and let p_1, p_2 be optimal contracts for λ_1 and λ_2 respectively, *i.e.*, $f(\lambda_1) = F_{a_1}^\top p_1$ and $f(\lambda_2) = F_{a_1}^\top p_2$. Let $p_\alpha = \alpha p_1 + (1-\alpha) p_2$ for $\alpha \in [0,1]$. Then, by linearity, p_α is feasible point of (9) with $\lambda = \alpha\lambda_1 + (1-\alpha)\lambda_2$. Therefore, it holds that:

$$f(\lambda_\alpha) \leq F_{a_1}^\top p_\alpha = \alpha F_{a_1}^\top p_1 + (1-\alpha) F_{a_1}^\top p_2 = \alpha f(\lambda_1) + (1-\alpha) f(\lambda_2),$$

and, thus f is convex. Concerning the strict monotonicity, instead, consider $\lambda_1 < \lambda_2$, and let p_2 be an optimal solution for λ_2, *i.e.*, $f(\lambda_2) = F_{a_1}^\top p_2$. Let $\alpha = \frac{\lambda_1}{\lambda_2}$ and let $p = \alpha \cdot p_2$. Then, observe that

$$(F_{a_1} - F_{a_2})^\top p = \frac{\lambda_1}{\lambda_2} (F_{a_1} - F_{a_2})^\top p_2 = \lambda_1.$$

Furthermore, since $p_2 \in \mathcal{P}_1$ also $p \in \mathcal{P}_1$. Hence, p is a feasible point of (9) with λ set to λ_1. Therefore, it holds that:

$$f(\lambda_1) \leq F_{a_1}^\top p = \frac{\lambda_1}{\lambda_2} F_{a_1}^\top p_2 = \frac{\lambda_1}{\lambda_2} f(\lambda_2) < f(\lambda_2),$$

where in the last point, we have used that for $\lambda_2 > 0$, $f(\lambda_2)$ is strictly positive due to the constraint $(F_{a_1} - F_{a_2})^\top p = \lambda_2$. This concludes the proof. □

A.2 Properties of the non-uniform discretization

We now prove some key properties of the non-uniform discretization that we introduced in Section 3.1.

Lemma 6 (Discretization Error)**.** *Let $\epsilon > 0$ and suppose that $F_{a_1}^\top r > F_{a_2}^\top r$. Then, for any sequence $\{\theta_t\}_{t=1}^T$, the following holds:*

$$|\Lambda^{(\epsilon)}| \leq \frac{1}{\epsilon} + 1 \tag{12}$$

$$\textsc{Opt} - \max_{\lambda \in \Lambda^{(\epsilon)}} \sum_{t=1}^T \widetilde{U}^P(\lambda, \theta_t) \leq T\epsilon. \tag{13}$$

Proof. We first prove that $|\Lambda^{(\epsilon)}| \leq \frac{1}{\epsilon} + 1$. By definition, it holds that:

$$(|\Lambda^{(\epsilon)}| - 1)\epsilon \leq \sum_{i=1}^{|\Lambda^{(\epsilon)}|} \left(f(\lambda^{(i)}) - f(\lambda^{(i-1)}) \right) = f(\lambda^{(|\Lambda^{(\epsilon)}|)}) - f(\lambda^{(0)}) \leq 1,$$

where in the last step we used $f(\lambda^{(0)}) = 0$ and $f(\cdot) \leq 1$ (Lemma 5). Thus, Equation (12) holds.

Next, we continue by proving Equation (13). By applying Lemma 3, we have that there exists $\lambda^\star \in [0, (F_{a_1} - F_{a_2})^\top r]$ such that $\textsc{Opt} = \sum_{t=1}^T \widetilde{U}^P(\lambda^\star, \theta_t)$. Let $\tilde{\lambda} = \min\{\lambda \in \Lambda^{(\epsilon)} : \lambda \geq \lambda^\star\}$. Observe

that, by construction, $\tilde{\lambda}$ always exists as $\max_{\lambda \in \Lambda^{(\epsilon)}} \lambda = (F_{a_1} - F_{a_2})^\top r$ by definition. Now, suppose that $\tilde{\lambda} \neq \lambda^\star$ (otherwise Equation (13) is direct) and observe that the set $\overline{\Lambda}^{(\epsilon)} = \{\lambda \in \Lambda^{(\epsilon)} : \lambda \leq \lambda^\star\}$ is always non-empty, since $0 \in \Lambda^{(\epsilon)}$ for all $\epsilon > 0$ and $\lambda^\star \geq 0$. Denote by $\overline{\lambda}$ the following quantity $\overline{\lambda} = \max_{\lambda \in \overline{\Lambda}^{(\epsilon)}} \lambda$ and observe that:

$$f(\tilde{\lambda}) - f(\overline{\lambda}) \leq \epsilon. \tag{14}$$

The proof of Equation (14) is direct once we realize that, by definition, there exists $n \in \mathbb{N}$ such that $\overline{\lambda} = \lambda^{(n)}$ and $\tilde{\lambda} = \lambda^{(n+1)}$ (*i.e.*, $\overline{\lambda}$ and $\tilde{\lambda}$ are adjacent elements of the sequence we constructed).

Then, we have that:

$$\begin{aligned}
\textsc{Opt} - \max_{\lambda \in \Lambda^{(\epsilon)}} \sum_{t=1}^{T} \widetilde{U}^P(\lambda, \theta_t) &\leq \sum_{t=1}^{T} \widetilde{U}^P(\lambda^\star, \theta_t) - \widetilde{U}^P(\tilde{\lambda}, \theta_t) \\
&\leq T\left(f(\tilde{\lambda}) - f(\lambda^\star)\right) && \text{(Lemma 4 and } \tilde{\lambda} \geq \lambda^\star\text{)} \\
&\leq T\left(f(\tilde{\lambda}) - f(\overline{\lambda})\right) && \text{(} \overline{\lambda} \leq \lambda^\star \text{ and } \lambda \to f(\lambda) \text{ is increasing)} \\
&\leq T\epsilon, && \text{(Equation (14))}
\end{aligned}$$

This concludes the proof. □

A.3 Combining the results

Finally, we can combine these results and prove Theorem 1.

Proof of Theorem 1. If $F_{a_2}^\top r \geq F_{a_1}^\top r$, the result is direct from Lemma 17.

Next, consider $F_{a_1}^\top r > F_{a_2}^\top r$. Let $\epsilon > 0$ and consider the following set of contracts:

$$\mathcal{P}^{(\epsilon)} = \{p \in [0,1]^m : \exists \lambda \in \Lambda^{(\epsilon)} \text{ and } F_{a_1}^\top p = f(\lambda)\}.$$

Consider any minimax optimal algorithm for finite-armed adversarial bandits and run it using as arm space $\mathcal{P}^{(\epsilon)}$, *i.e.*, Audibert and Bubeck [3]. Recall that the algorithm in Audibert and Bubeck [3] achieves $\mathcal{O}(\sqrt{TK})$ regret on adversarial oblivious bandits with K arms.

Then, it holds that:

$$\begin{aligned}
R_T &= \textsc{Opt} - \mathbb{E}\left[\sum_{t=1}^{T} U^P(p_t, b(\theta_t, p_t)\right] \\
&= \textsc{Opt} - \max_{p \in \mathcal{P}^{(\epsilon)}} \sum_{t=1}^{T} U^P(p, b(\theta_t, p)) + \max_{p \in \mathcal{P}^{(\epsilon)}} \sum_{t=1}^{T} U^P(p, b(\theta_t, p)) - \mathbb{E}\left[\sum_{t=1}^{T} U^P(p_t, b(\theta_t, p_t)\right] \\
&= \textsc{Opt} - \max_{\lambda \in \Lambda^{(\epsilon)}} \sum_{t=1}^{T} \widetilde{U}^P(\lambda, \theta_t) + \max_{p \in \mathcal{P}^{(\epsilon)}} \sum_{t=1}^{T} U^P(p, b(\theta_t, p)) - \mathbb{E}\left[\sum_{t=1}^{T} U^P(p_t, b(\theta_t, p_t)\right] \\
&\leq T\epsilon + \max_{p \in \mathcal{P}^{(\epsilon)}} \sum_{t=1}^{T} U^P(p, b(\theta_t, p)) - \mathbb{E}\left[\sum_{t=1}^{T} U^P(p_t, b(\theta_t, p_t)\right] && \text{(Lemma 6)} \\
&\in \mathcal{O}\left(T\epsilon + \sqrt{|\mathcal{P}^{(\epsilon)}|T}\right) && \text{(Minimax optimality)} \\
&= \mathcal{O}\left(T\epsilon + \sqrt{\frac{T}{\epsilon}}\right). && \text{(Lemma 6)}
\end{aligned}$$

where the second inequality follows from Lemma 2 together with the definition of $f(\lambda)$ and the fact that $p \in \mathcal{P}^{(\epsilon)}$ if and only if there exists $\lambda \in \Lambda^{(\epsilon)}$ such that $F_{a_1}^\top p = f(\lambda)$.

Now, choosing $\epsilon = T^{-1/3}$ yields the result. □

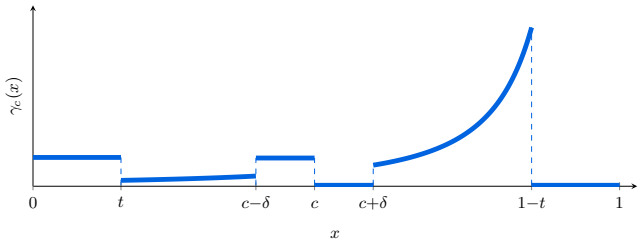


Figure 2: Type distribution density of the lower bound instance.

B Proof of Theorem 2

In this section, we give a formal proof of the $\Omega(T^{2/3})$ lower bound, which holds even when the type sequence is stochastic and comes from an absolutely continuous distribution.

B.1 Construction

We consider the following $K+1$ instances $\{\mathfrak{I}_0, \mathfrak{I}_1, \ldots, \mathfrak{I}_K\}$, where $\mathfrak{I}_0$ will be the *base* instance and $\{\mathfrak{I}_1, \ldots, \mathfrak{I}_K\}$ will be the *perturbed* instances. Let $\{\nu_0, \nu_1, \ldots, \nu_K\}$ be the joint probability distributions over the T turns of the interaction between the instances and the (possibly randomized) algorithm. Each perturbed instance will be parametrized by two parameters (c, δ), which we will fix in the following. We now describe a generic instance parametrized by $(c, \delta) \in [0,1]^2$. We have $\Gamma_{c,\delta}$, the type' distribution (that we will describe shortly), 2 outcomes ω_1 and ω_2 such that $r_{\omega_1} = 1$ and $r_{\omega_2} = 0$, the cost of performing action a_1 is $c_{a_1} = 1$ and the cost of performing action a_2 is $c_{a_2} = 0$, action a_1 leads deterministically to ω_1 and action a_2 leads deterministically to ω_2.

Fix $t = 1/5$ and any $\epsilon > 0$. To each instance $\mathfrak{I}_k$ with $k \geq 1$, we are going to associate a type distribution $\Gamma_k = \Gamma_{c_k,\delta_k}$, where $c_k = t + \bar{\delta}(2k-1)$, $\bar{\delta} := \epsilon\frac{1-t}{t+2\epsilon}$, and finally, $\delta_k = \epsilon\frac{1-c_k}{t+\epsilon}$. The base instance $\mathfrak{I}_0$ will be defined with $\Gamma_0 = \Gamma_{0,0}$, whose definition will be clear once we give a concrete parametrized definition of $\Gamma_{c,\delta}$.

First, we characterize the shape of the principal utility for a generic instance parametrized by any Γ.

Lemma 7. *For each contract $p \in [0,1]^2$, and any type distribution Γ, the principal' utility is:*

$$U^P(p) = \mathbb{P}_\Gamma(\theta \leq p_{\omega_1} - p_{\omega_2})(1 - (p_{\omega_1} - p_{\omega_2})) - p_{\omega_2}.$$

Proof. The agent will play action a_1 if and only if $p_{\omega_1} - \theta \cdot c_{a_1} = p_{\omega_1} - \theta \geq p_{\omega_2}$ so only if $\theta \leq p_{\omega_1} - p_{\omega_2}$. Thus, the utility of the principal is

$\mathbb{P}_\Gamma(\theta \leq p_{w_1} - p_{w_2})(1-p_{w_1}) + \mathbb{P}_\Gamma(\theta > p_{w_1} - p_{w_2})(-p_{w_2}) = \mathbb{P}_\Gamma(\theta \leq p_{w_1} - p_{w_2})(1 - p_{w_1} + p_{w_2}) - p_{w_2}$,

as claimed. □

Remark 1. *From this lemma, it is clear that we can focus on algorithms that play $p_{w_2} = 0$. Indeed, for every algorithm that plays $(p_{\omega_1}, p_{\omega_2})$, we can instantiate an algorithm that plays $(p_{\omega_1} - p_{\omega_2}, 0)$ and the feedback under any instance $\{\mathfrak{I}_k\}_{k=0}^K$, has the same distribution and larger utility.*

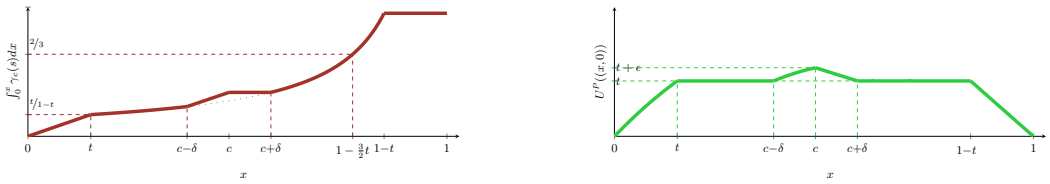
Thus, for every instance $\{\mathfrak{I}_0, \ldots, \mathfrak{I}_K\}$ we can define

$$U_k^P(x) := \mathbb{P}_{\Gamma_k}(\theta \leq x)(1-x),$$

as the principal expected utility of proposing contract $p = (x, 0)$ to the agent.

Now we can define the type distribution $\Gamma_{c,\delta}$ as follows. The type distribution is given by its probability density $\gamma_{c,\delta}$.

$$\gamma_{c,\delta}(x) = \begin{cases} \frac{1}{1-t} & \text{if } x \in [0, t) \\ \frac{t}{(1-x)^2} & \text{if } x \in [t, c-\delta) \\ \frac{2t}{(1-c-\delta)(1-c+\delta)} & \text{if } x \in [c-\delta, c) \\ 0 & \text{if } x \in [c, c+\delta) \\ \frac{t}{(1-x)^2} & \text{if } x \in [c+\delta, 1-t) \\ 0 & \text{if } x \in [1-t, 1] \end{cases} \tag{15}$$



(a) Feedback Function

(b) Reward function

Figure 3: Feedback distribution and reward function of the lower bound instances.

A visual representation of the density is shown in Figure 2. The pdf has two constant regions: $[0, t)$ and $[c - \delta, c)$, while on $[t, c - \delta)$ and $[c + \delta, 1 - t)$ it is a power-law. Everywhere else it is 0. From Lemma 7 and Remark 1, we know that the principal utility is the function $x \mapsto (1 - x) \cdot \int_0^x \gamma_{c,\delta}(s)ds$. We now characterize the main features of the instances, which are 1) the reward structure and 2) the information structure. An illustration can also be found in Figure 3.

Lemma 8. *For each $k \in [K]$, for a contract $p = (x, 0)$, the principal expected utility $U^P(p)$ in instance $\mathfrak{I}_k$ and the feedback distribution is a Bernoulli with probability $\mathbb{P}_{\Gamma_k}(\theta \leq x)$ are as follows:*

$$\mathbb{P}_{\Gamma_k}(\theta \leq x) = \begin{cases} \frac{x}{1-t} & \text{if } x \in [0, t), \\ \frac{t}{1-x} & \text{if } x \in [t, c_k - \delta_k), \\ \frac{t}{1-c_k+\delta_k} + (x - c_k + \delta_k)\frac{2t}{(1-c_k-\delta_k)(1-c_k+\delta_k)} & \text{if } x \in [c_k - \delta_k, c_k), \\ \frac{t}{1-c_k-\delta_k} & \text{if } x \in [c_k, c_k + \delta_k), \\ \frac{t}{1-x} & \text{if } x \in [c_k + \delta_k, 1 - t), \\ 1 & \text{if } x \in [1 - t, 1]. \end{cases}$$

Moreover, the base instance is characterized by:

$$\mathbb{P}_{\Gamma_0}(\theta \leq x) = \begin{cases} \frac{x}{1-t} & \text{if } x \in [0, t), \\ \frac{t}{1-x} & \text{if } x \in [t, 1 - t), \\ 1 & \text{if } x \in [1 - t, 1]. \end{cases}$$

We now observe some important features of our instances. For each instance $\mathfrak{I}_k$ for $k \geq 1$, we define the interval $J_k := [c_k - \bar{\delta}, c_k + \bar{\delta}]$ and $B_k := [c_k, c_k + \delta_k/2]$. In Figure 3 we have an illustration of the reward and feedback distribution.

Lemma 9. *For any $K \leq \frac{1}{20\epsilon}$ and $\epsilon < 1/20$, we have*

(i) $\mathbb{P}_{\Gamma_k}(\theta \leq x) = \mathbb{P}_{\Gamma_{k'}}(\theta \leq x)$ *for all* $x \notin J_k \cup J_{k'}$;

(ii) $\delta_k \leq \bar{\delta}$ *for all* $k \in [K]$;

(iii) $J_k \subseteq [t, 1 - \frac{3}{2}t]$ *for all* $k \in [K]$;

(iv) $U_k^P(x) \geq t + \epsilon/2$ *for all* $x \in B_k := [c_k, c_k + \delta_k/2]$;

(v) $|\mathbb{P}_{\Gamma_k}(\theta \leq x) - \mathbb{P}_{\Gamma_0}(\theta \leq x)| \leq 4\epsilon$ *for all* $x \in J_k$.

Proof. For (i) it is enough to observe that all type distributions coincide with the base instance outside of the intervals J_k.

For (ii) we can observe that δ_k is decreasing in c_k and thus it is enough to prove the statement for $k = 1$, and straightforward calculations show that $\bar{\delta} = \delta_1$.

For (iii) it is enough to observe that $c_k - \bar{\delta} = t + 2\bar{\delta}(k - 1) \geq t$ and thus $J_k \subseteq [t, \infty)$. Moreover, $c_k + \bar{\delta} \leq c_K + \bar{\delta} = t + 2K\bar{\delta}$. Since $t = 1/5$, $\bar{\delta} = \frac{4\epsilon}{1+10\epsilon}$, we get $2K\bar{\delta} \leq 2/5$ and thus $c_k + \bar{\delta} \leq 1 - \frac{3}{2}t$, proving that $J_k \subseteq [t, 1 - \frac{3}{2}t]$.

For (iv) it is sufficient to observe that $U_k^P(x)$ is decreasing in x in the interval B_k, and that $U_k^P(c_k + \delta_k/2) = t + \epsilon/2$.

For (v) we can observe that by (ii) and Lemma 8, for all $x \notin [c_k - \delta_k, c_k + \delta_k]$

$$\mathbb{P}_{\Gamma_k}(\theta \le x) = \mathbb{P}_{\Gamma_0}(\theta \le x).$$

Thus, it suffices to consider $x \in [c_k - \delta_k, c_k + \delta_k]$. On this interval, the difference

$$D(x) := \mathbb{P}_{\Gamma_k}(\theta \le x) - \mathbb{P}_{\Gamma_0}(\theta \le x)$$

is nonnegative, increasing on $[c_k - \delta_k, c_k)$, and decreasing on $[c_k, c_k + \delta_k]$, hence it is maximized at $x = c_k$. Therefore

$$|\mathbb{P}_{\Gamma_k}(\theta \le x) - \mathbb{P}_{\Gamma_0}(\theta \le x)| \le \frac{t}{1 - c_k - \delta_k} - \frac{t}{1 - c_k} = \frac{\epsilon}{1 - c_k}.$$

By item 3, $1 - c_k \ge 3t/2$, so the right-hand side is at most $\frac{2}{3t}\epsilon \le 4\epsilon$, concluding the proof. □

A standard change of measure argument concluded the proof.

Theorem 5. *For any (possibly randomized) learning algorithm, there exists $k \in [K]$ such that*

$$\textsc{Opt} - \mathbb{E}_{\nu_k}\left[\sum_{t=1}^{T} U^P(p_t)\right] = \Omega(T^{2/3}).$$

Proof. Because of Remark 1, we can focus on algorithms that play on the single-dimensional line $x_t \in [0, 1]$, and the associated contract is $p_t = (x_t, 0)$. For any instance $k \in [K]$ let $N_k(T) = \sum_{t\in[T]} \mathbb{I}(x_t \in J_k)$ and $M_k(T) = \sum_{t\in[T]} \mathbb{I}(x_t \in B_k)$.

Claim 1.

$$\mathrm{KL}(\nu_0, \nu_k) \le 10^4 \epsilon^2 \mathbb{E}_{\nu_0}[N_k(T)].$$

Proof. We can use the KL decomposition for continuous action spaces [41, Lemma 15.8 (a)].

$$\mathrm{KL}(\nu_0, \nu_k) = \int_0^1 \mathrm{kl}(\mathbb{P}_{\Gamma_k}(\theta \le x), \mathbb{P}_{\Gamma_0}(\theta \le x)) dG_{\nu_0}(x),$$

where $G_{\nu_0}(B) = \mathbb{E}_{\nu}[\sum_{t\in[T]} \mathbb{I}(x_t \in B)]$ and $\mathrm{kl}(a, b) = a \log(a/b) + (1 - a) \log(\frac{1-a}{1-b})$. By Lemma 9, we have that

$$\begin{aligned}
\mathrm{KL}(\nu_0, \nu_k) &= \int_{J_k} \mathrm{kl}(\mathbb{P}_{\Gamma_k}(\theta \le x), \mathbb{P}_{\Gamma_0}(\theta \le x)) dG_{\nu_0}(x) \\
&\le G_{\nu_0}(J_k) \sup_{x\in J_k} \mathrm{kl}(\mathbb{P}_{\Gamma_k}(\theta \le x), \mathbb{P}_{\Gamma_0}(\theta \le x)) \\
&= \mathbb{E}_{\nu_0}[N_k(T)] \mathrm{kl}(\mathbb{P}_{\Gamma_k}(\theta \le x), \mathbb{P}_{\Gamma_0}(\theta \le x)).
\end{aligned}$$

Now we upper bound $\mathrm{kl}(a, b) \le \frac{(a-b)^2}{2\ell(1-u)}$, which holds for all $a, b \in [\ell, u]$. By using Lemma 9 we know that $\mathbb{P}_{\Gamma_0}(\theta \le x) \in [1/4, 2/3]$, and from Lemma 8 (v) that $\mathbb{P}_{\Gamma_k}(\theta \le x) \in [1/4 - 4\epsilon, 2/3 + 4\epsilon]$, which, for $\epsilon \le 1/20$, is contained in $[1/20, 13/15] = [\ell, u]$. This lets us conclude that

$$\mathrm{KL}(\nu_0, \nu_k) \le 10^4 \epsilon^2 \mathbb{E}_{\nu_0}[N_k(T)],$$

concluding the proof. □

Moreover, the regret can be easily upper-bounded as follows:

$$\begin{aligned}
\textsc{Opt} - \mathbb{E}_{\nu_k}\left[\sum_{t=1}^{T} U^P(p_t)\right] &= T\,(t + \epsilon) - \mathbb{E}_{\nu_k}\left[\sum_{t\in[T]:x_t\in B_k(T)}^{T} U^P(p_t) + \sum_{t\in[T]:x_t\notin B_k(T)}^{T} U^P(p_t)\right] \\
&\ge T(t + \epsilon) - \mathbb{E}_{\nu_k}\left[M_k(T)\right](t + \epsilon) - (T - \mathbb{E}_{\nu_k}\left[M_k(T)\right])(t + \frac{\epsilon}{2}) \\
&= \frac{\epsilon}{2}(T - \mathbb{E}_{\nu_k}\left[M_k(T)\right]). \qquad (16)
\end{aligned}$$

The proof can now be concluded as follows. Since J_k are disjoint, we have $\sum_{k\in[K]} N_k(T) \leq T$ and thus there eixts k^* such that $\mathbb{E}_{\nu_0}[N_{k^*}(T)] \leq T/K$ and, since $M_k(T) \leq N_k(T)$, also $\mathbb{E}_{\nu_0}[M_{k^*}(T)] \leq T/K$ and by Markov $\mathbb{P}_{\nu_0}[M_{k^*}(T) \geq T/2] \leq 2/K$. Now, by Pinsker's inequality and Claim 1, we have

$$|\mathbb{P}_{\nu_0}(A) - \mathbb{P}_{\nu_k}(A)| \leq \sqrt{\frac{1}{2}\mathrm{KL}(\nu_0, \nu_k)} \leq 10^2\epsilon\sqrt{\mathbb{E}_{\nu_0}[N_k(T)]},$$

for any measurable event A and instance $k \in [K]$. We can apply this to the event $\{M_{k^*}(T) \geq T/2\}$, which gives

$$\mathbb{P}_{\nu_{k^*}}[M_{k^*} \leq T/2] \geq 1 - \frac{2}{K} - \frac{1}{2}\epsilon\sqrt{\mathbb{E}_{\nu_0}[N_{k^*}(T)]} \geq 1 - \frac{2}{K} - 10^2\epsilon\sqrt{\frac{T}{K}}, \tag{17}$$

where the last inequality comes from our choice of k^*. From, reverse Markov's inequality it is clear that $\mathbb{E}_{\nu_{k^*}}[M_{k^*}(T)] \leq T - \frac{T}{2}\mathbb{P}_{\nu_{k^*}}[M_{k^*}(T) \leq T/2]$, which plugged into Equation (16), gives

$$\begin{aligned}
\textsc{Opt} - \mathbb{E}_{\nu_{k^*}}\left[\sum_{t=1}^{T} U^P(p_t)\right] &\geq \frac{\epsilon T}{4}\mathbb{P}_{\nu_{k^*}}[M_{k^*}(T) \leq T/2] \\
&\geq \frac{\epsilon T}{4}\left(1 - \frac{2}{K} - 10^2\epsilon\sqrt{\frac{T}{K}}\right). \qquad \text{(Equation (17))}
\end{aligned}$$

Now we can take $K = \lfloor \frac{1}{15\epsilon} \rfloor \leq \frac{1}{15\epsilon}$, which is allowed by Lemma 9. This gives:

$$\textsc{Opt} - \mathbb{E}_{\nu_{k^*}}\left[\sum_{t=1}^{T} U^P(p_t)\right] \geq \frac{\epsilon T}{4}\left(1 - 32\epsilon - 10^2\epsilon\sqrt{32T\epsilon}\right),$$

choosing $\epsilon = T^{-1/3}/2^7$ gives

$$\textsc{Opt} - \mathbb{E}_{\nu_{k^*}}\left[\sum_{t=1}^{T} U^P(p_t)\right] \geq \frac{T^{2/3}}{2^7} - \frac{T^{1/3}}{2^{11}} - \frac{T^{2/3}}{2^{10}} \geq \frac{T^{2/3}}{200},$$

which proves the desired lower bound. □

C Proof of Theorem 3

Now, we focus on the setting where we want to minimize the regret against a single unknown type $\bar{\theta} \in [0, 1]$. When $F_{a_2}^\top r \geq F_{a_1}^\top r$ there is a trivial solution that yields regret 0 (Lemma 17). Therefore, in the following we focus only on the case in which $F_{a_1}^\top r > F_{a_2}^\top r$.

The section is structured as follows:

- First, we discuss the properties of contracts played by Algorithm 1. In particular, we first formally define θ_{lazy}, and then we show that contracts played by Algorithm 1 are optimal contracts that can be used to approximate $\bar{\theta}$. More precisely, we formally show that, for each $\theta \leq \theta_{\text{lazy}}$, there exists optimal contracts such that for all types smaller or equal than θ, the best-response is a_1, while for types larger than θ the best response is a_2. Furthermore, we also show that, for types larger than θ_{lazy}, it is optimal to play the null contract. Finally, we conclude these sections by showing that these contracts have instantaneous which is bounded by $\mathcal{O}(\|F_{a_1} - F_{a_2}\|_1)$.
- Then, in Section C.2, we provide proofs regarding the statistical tests that we use in Algorithm 1 to approximate correctly the hidden type $\bar{\theta}$.
- Finally, in Section C.3, we combine these results to analyze the regret of Algorithm 1 and prove Theorem 3.

C.1 Properties of contracts played by Algorithm 1

We start by presenting the formal definition of θ_{lazy}. Recall that we introduced θ_{lazy} as the a type such that, for all $\theta > \theta_{\text{lazy}}$, the optimal action to be induced is a_2, and hence, the optimal contract is the

null contract $\bar{p} = (0, \dots, 0)$. The rationale behind θ_{lazy} is that incentivizing a_1 for types larger than θ_{lazy} is not convenient for the principal. Specifically, we define θ_{lazy} as the value of the following linear program.

$$\theta_{\text{lazy}} \coloneqq \begin{cases} \max\limits_{\theta \in [0,1], p \in [0,1]^m} & \theta \\ \text{s.t.} & (F_{a_1} - F_{a_2})^\top p \geq \theta C_{a_1} \\ & F_{a_1}^\top (r - p) \geq F_{a_2}^\top r \end{cases} \tag{18}$$

First of all, we observe that Equation (18) is always well-defined since it has a non-empty and bounded feasible region. To see that this region is non-empty it is sufficient to consider $\theta = 0$ and $p = (0, \dots, 0)$. Then, all constraints are satisfied since $F_{a_1}^\top r > F_{a_2}^\top r$ holds by assumption. Furthermore, it is also easy to see that $\theta_{\text{lazy}} > 0$, as $F_{a_1}^\top r$ is strictly greater than $F_{a_2}^\top r$ and C_{a_1} is strictly greater than 0.

Next, we observe that for all (θ, p) in the feasible region, $(F_{a_1} - F_{a_2})^\top p \geq \theta C_{a_1}$, so that $a_1 \in \mathcal{B}(\theta, p)$, and, furthermore, $F_{a_1}^\top (r - p) \geq F_{a_2}^\top r$, so that the principal utility when playing p is larger than the optimal utility that can be obtained when playing contracts that induce a_2, *i.e.*, $F_{a_2}^\top r$. Note, moreover, that $F_{a_1}^\top (r - p) \geq F_{a_2}^\top r \geq F_{a_2}^\top (r - p)$ This ensures that, when playing p against θ, $b(\theta, p) = a_1$. It follows directly from this definition that, for all $\theta \leq \theta_{\text{lazy}}$, it is optimal for the principal to induce a_1, while for all types greater than θ_{lazy}, it is optimal to induce a_2 using the null contract. This result is proved in the following lemma.

Lemma 10 (Agent behaviors and optimal contracts)**.** *For all $\theta > \theta_{\text{lazy}}$, $p = (0, \dots, 0)$ is an optimal contract. Furthermore, for all $\theta \leq \theta_{\text{lazy}}$, there exists a contract $p_\theta^\star$ such that:*

$$\textit{(i)} \max_{p \in [0,1]^m} U^P(p, b(\theta, p)) = U^P(p_\theta^\star, b(\theta, p_\theta^\star))$$

$$\textit{(ii)}\ a_1 = b(\tilde{\theta}, p_\theta^\star) \quad \forall \tilde{\theta} \leq \theta$$

$$\textit{(iii)}\ a_2 = b(\tilde{\theta}, p_\theta^\star) \quad \forall \tilde{\theta} > \theta.$$

In particular, $p_\theta^\star$ is the solution of (9) with λ set to θC_{a_1}.

Proof. First, fix any $\theta > \theta_{\text{lazy}}$ and observe that the null contract yields utility $F_{a_2}^\top r$ to the principal. Furthermore, it is also easy to see that the null contract is optimal among all the contracts for which the agent plays a_2. Next, focus on contracts for which the agent plays a_1. Specifically, suppose that there exists $p \in [0,1]^m$ such that $a_1 = b(\theta, p)$ and $F_{a_1}^\top (r - p) \geq F_{a_2}^\top r$, *i.e.*, agent plays a_1 when offered p, and p yields the same (or more) utility to the principal than the null contract. This however is impossible, since $a_1 = b(\theta, p)$ implies that $(F_{a_1} - F_{a_2})^\top p \geq \theta C_{a_1}$, and, therefore, since, by assumption $F_{a_1}^\top (r - p) \geq F_{a_2}^\top r$ holds, (θ, p) would belong to the feasible region of (18). Then, since $\theta > \theta_{\text{lazy}}$, θ_{lazy} cannot be an optimal solution of (18). Thus, we proved that there is no contract which induces the agent to play a_1 while guaranteeing more utility for the principal. This concludes the first part of the proof.

Second, consider $\theta \leq \theta_{\text{lazy}}$. We start by proving (i). Observe that, due to Lemma 3, the maximum over contracts is well-defined.

$$\begin{aligned}
\max_{p \in [0,1]^m} U^P(p, b(\theta, p)) &= \max \left\{ \max_{p : a_1 = b(\theta,p)} U^P(p, b(\theta, p)), \max_{p : a_2 = b(\theta,p)} U^P(p, b(\theta, p)) \right\} \\
&= \max \left\{ \max_{p : a_1 = b(\theta,p)} U^P(p, b(\theta, p)), F_{a_2}^\top r \right\} \\
&= \max \left\{ \max_{p : (F_{a_1} - F_{a_2})^\top p \geq \theta C_{a_1}} F_{a_1}^\top (r - p), F_{a_2}^\top r \right\} \\
&= \max \left\{ F_{a_1}^\top r - \min_{p : (F_{a_1} - F_{a_2})^\top p = \theta C_{a_1}} F_{a_1}^\top p, F_{a_2}^\top r \right\} \\
&= \max \left\{ F_{a_1}^\top (r - p_\theta^\star), F_{a_2}^\top r \right\}.
\end{aligned}$$

where the last equality follows directly from the definition of $p_\theta^\star$. Now, observe that, since $p_\theta^\star$ is optimal for θ among all the contracts for which $a_1 \in \mathcal{B}(\theta, p)$, and since $\theta \leq \theta_{\text{lazy}}$, it must hold that

the pair $(\theta, p^\star_\theta)$ is a feasible solution of (18), and, therefore, $F_{a_1}^\top(r - p^\star_\theta) \geq F_{a_2}^\top r$. As a consequence, we have proved (i).

To prove (ii), it is sufficient to observe that $a_1 = b(\theta, p^\star_\theta)$. Hence, it must hold that $a_1 = b(\tilde{\theta}, p^\star_\theta)$ for all $\tilde{\theta} < \theta$ as well.

Finally, to prove (iii), we observe that, by definition $p^\star_\theta$ satisfies $(F_{a_1} - F_{a_2})^\top p^\star_\theta = \theta C_{a_1}$. Therefore, for all $\tilde{\theta} > \theta$, action a_1 cannot belong to $\mathcal{B}(\tilde{\theta}, p^\star_\theta)$. This concludes the proof. □

Next, we show that the instantaneous regret for the contracts that we designed above is upper bounded by $\|F_{a_1} - F_{a_2}\|_1$, and, therefore, they are suitable for exploration.

Lemma 11 (Sub-optimality gaps)**.** *Suppose $F_{a_1}^\top r > F_{a_2}^\top r$. Consider $\theta_1, \theta_2 \in [0,1]$ and let $p^\star_1, p^\star_2$ be optimal contracts for θ_1 and θ_2 respectively. Then, it holds that:*

$$U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_1, p^\star_2)) \leq 2\|F_{a_1} - F_{a_2}\|_1. \tag{19}$$

Furthermore, let $\alpha > 0$ and define $\tilde{p}_2 = p^\star_2 + \alpha(r - p^\star_2)$. Then, it holds that:

$$U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_1, \tilde{p}^\star_2)) \leq 2\|F_{a_1} - F_{a_2}\|_1 + \alpha. \tag{20}$$

Finally, let $\bar{p} = (0, \dots, 0)$. It holds that:

$$U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(\bar{p}, b(\theta_1, \bar{p})) \leq \|F_{a_1} - F_{a_2}\|_1. \tag{21}$$

Proof. We start by proving Equation (19). First, observe that $p^\star_1$ and $p^\star_2$ are well-defined due to Lemma 3.

Second, fix any contract $p \in [0,1]$ and any θ_1, θ_2. Denote by $a = b(\theta_1, p)$ and $\bar{a} = b(\theta_2, p)$. Then, it holds that:

$$U^P(p, b(\theta_1, p)) - U^P(p, b(\theta_2, p)) = \begin{cases} 0 & \text{if } a = a \\ F_a^\top(r - p) - F_{\bar{a}}^\top(r - p) & \text{if } a \neq \bar{a}, \end{cases}$$

from which it directly follows that:

$$U^P(p, b(\theta_1, p)) - U^P(p, b(\theta_2, p)) \leq \|F_{a_1} - F_{a_2}\|_1. \tag{22}$$

At this point, we are ready to prove our result.

$$\begin{aligned} U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_1, p^\star_2)) = U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_2, p^\star_2)) + \\ + U^P(p^\star_2, b(\theta_2, p^\star_2)) - U^P(p^\star_2, b(\theta_1, p^\star_2)) \end{aligned}$$

Focus on the second term of the r.h.s. of this equation. Due to Equation (22), it is smaller than $\|F_{a_1} - F_{a_2}\|_1$. Concerning the first term of the r.h.s., instead, we observe the following.

$$\begin{aligned} \max_{p \in [0,1]^m} U^P(p, b(\theta_1, p)) - \max_{p \in [0,1]^m} U^P(p, b(\theta_2, p)) &\leq \max_{p \in [0,1]^m} (U^P(p, b(\theta_1, p)) - U^P(p, b(\theta_2, p))) \\ &\leq \|F_{a_1} - F_{a_2}\|_1. \qquad \text{(by Equation (22))} \end{aligned}$$

Combining these arguments, yields Equation (19).

We now continue by proving Equation (20). To this end, we note that we can decompose $U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_1, \tilde{p}^\star_2))$ as the sum of $U^P(p^\star_1, b(\theta_1, p^\star_1)) - U^P(p^\star_2, b(\theta_1, p^\star_2))$ and $U^P(p^\star_2, b(\theta_1, p^\star_2)) - U^P(p^\star_2, b(\theta_1, \tilde{p}^\star_2))$. The first term can be bounded using Equation (19), yielding the $\|F_{a_1} - F_{a_2}\|_1$ term in Equation (20). For the second term, instead, we apply Lemma 19 with $\epsilon = 0$ and $\rho(\theta) = b(\theta, p^*_2)$. This yields:

$$U^P(p^\star_2, b(\theta_1, p^\star_2)) - U^P(p^\star_2, b(\theta_1, \tilde{p}^\star_2)) \leq \alpha,$$

which concludes the proof of Equation (20).

Finally, we prove Equation (21). If $b(\theta_1, p^\star_1) = a_2$, then, the null contract must be an optimal contract for θ_1 as all the other contracts that induce a_2 cannot yield larger utility to the principal. If $b(\theta_1, p^\star_1) = a_1$, instead, there are two cases. Either $b(\theta_1, \bar{p}) = a_1$, and again we have that the null contract must be an optimal contract for θ_1, or $b(\theta_1, \bar{p}) = a_2$. In this case:

$$F_{a_1}^\top(r - p^\star_1) - F_{a_2}^\top r \leq (F_{a_1} - F_{a_2})^\top r \leq \|F_{a_1} - F_{a_2}\|_1,$$

which concludes the proof. □

C.2 Statistical Test

In this section, we prove results regarding the statistical test that we use in Algorithm 1. Specifically, Lemma 12 shows how the Z statistic works when we apply to a stream of n i.i.d. outcomes realized from either F_{a_1} or F_{a_2}. Then, in Lemma 13, we apply such tests within Algorithm 1.

Lemma 12 (Statistical Test)**.** *Consider $\{\omega_i\}_{i=1}^n$ be a stream of n i.i.d. samples from F_{a_2}. Then, it holds that:*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\omega_i \in \Omega_1\} \geq \tau\right) \leq \exp\left(-\frac{1}{8}n\|F_{a_1} - F_{a_2}\|_1^2\right). \tag{23}$$

Similarly, if $\{\omega_i\}_{i=1}^n$ is a stream of n i.i.d. samples from F_{a_1}, it holds that:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\omega_i \in \Omega_1\} \leq \tau\right) \leq \exp\left(-\frac{1}{8}n\|F_{a_1} - F_{a_2}\|_1^2\right). \tag{24}$$

Proof. We first prove Equation (23). For brevity, we use Z to denote $\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\omega_i \in \Omega_1\}$. Observe that

$$\mathbb{E}[Z] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{1}\{\omega \in \Omega_1\}\right] = \mathbb{P}(\omega \in \Omega_1) = \sum_{\omega\in\Omega_1} F_{a_2,\omega}.$$

Then, we have that:

$$\begin{aligned}
\mathbb{P}\left(Z \geq \tau\right) &= \mathbb{P}\left(Z - \sum_{\omega\in\Omega_1} F_{a_2,\omega} \geq \tau - \sum_{\omega\in\Omega_1} F_{a_2,\omega}\right) \\
&= \mathbb{P}\left(Z - \sum_{\omega\in\Omega_1} F_{a_2,\omega} \geq \frac{\sum_{\omega\in\Omega_1}(F_{a_1,\omega} - F_{a_2,\omega})}{2}\right) && \text{(Def. of } \tau\text{)} \\
&= \mathbb{P}\left(Z - \sum_{\omega\in\Omega_1} F_{a_2,\omega} \geq \frac{\|F_{a_1} - F_{a_2}\|_1}{4}\right) \\
&\leq \exp\left(-\frac{1}{8}n\|F_{a_1} - F_{a_2}\|_1^2\right) && \text{(Lemma 18)}
\end{aligned}$$

The proof of Equation (24) follows from similar arguments. □

Lemma 13 (Correct Exploration)**.** *Consider Algorithm 1. Suppose $\|F_{a_1} - F_{a_2}\|_1 > \beta$ and $\bar{\theta} > \theta_{\text{lazy}}$. It holds that:*

$$\mathbb{P}\left(Z_0 \geq \tau\right) \leq \exp\left(-\frac{1}{8}n\|F_{a_1} - F_{a_2}\|_1^2\right). \tag{25}$$

Suppose that $\bar{\theta} \leq \theta_{\text{lazy}}$. It holds that:

$$\mathbb{P}\left(\bar{\theta} \notin [s_L, e_L]\right) \leq 2L\exp\left(-\frac{1}{8}n\|F_{a_1} - F_{a_2}\|_1^2\right). \tag{26}$$

Proof. The lemma is a combination of Lemma 10 and Lemma 12.

We start by proving Equation (25). If $\bar{\theta} > \theta_{\text{lazy}}$, we know from Lemma 10, that $\{\omega_i^{(0)}\}_{i=1}^n$ is a stream of n i.i.d. samples from F_{a_2}. Thus, applying Equation (23) in Lemma 12 yields Equation (25).

Second, we continue by proving Equation (26). For each phase $l \in \{0, \ldots, L-1\}$, we denote by $\mathcal{E}_l$ the following event:

$$\mathcal{E}_l = \left\{\left(\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_1} \wedge Z_l > \tau\right\}\right) \vee \left(\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_2} \wedge Z_l \leq \tau\right)\right\}.$$

Now, observe that, under $\bigcap_{l=0}^{L-1}\mathcal{E}_l$, we are guaranteed that $\bar{\theta} \in [s_L, e_L]$. Indeed, if $\mathcal{E}_0$ holds, Algorithm 1 enters the exploration phase. Indeed, due to Lemma 10, $\bar{\theta} \leq \theta_{\text{lazy}}$ responds to $p^\star_{\theta_{\text{lazy}}}$ by

playing a_1 and $Z_0 > \tau$ holds from $\mathcal{E}_0$. Then, the exploration phase runs a binary search sub-routine. Let us analyze the first phase $l = 1$. Suppose that $\bar{\theta} \in \left(\frac{1}{2}\theta_{\text{lazy}}, \theta_{\text{lazy}}\right]$. Then, by Lemma 10, $\bar{\theta}$ responds to $p^\star_{\frac{1}{2}\theta_{\text{lazy}}}$ by playing a_2, and since $Z_1 \leq \tau$ holds, we have that the search continues by inspecting the interval $[s_2, b_2] = \left[\frac{1}{2}\theta_{\text{lazy}}, \theta_{\text{lazy}}\right]$ which is the one that contains $\bar{\theta}$. Similarly, if $\bar{\theta} \in [0, \theta_{\text{lazy}}]$, from Lemma 10, we have that $\bar{\theta}$ plays a_1, and since $Z_1 > \tau$, we have that the search continues by inspecting the interval $[s_2, b_2] = \left[0, \frac{1}{2}\theta_{\text{lazy}}\right]$ which is the one that contains $\bar{\theta}$. Then, by an inductive argument, we obtain that $\bar{\theta} \in [s_L, e_L]$ under $\bigcap_{l=0}^{L-1} \mathcal{E}_l$. Then, once we reach the last phase we have

$$
\begin{aligned}
\mathbb{P}\left(\bar{\theta} \notin [s_L, e_L]\right) &\leq \sum_{i=0}^{L-1} \mathbb{P}\left(\mathcal{E}_l^{\complement}\right) \\
&= \sum_{i=0}^{L-1} \mathbb{P}\left(\left\{\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_1} \wedge Z_l \leq \tau\right\} \vee \left\{\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_2} \wedge Z_l > \tau\right\}\right) \\
&\leq \sum_{i=0}^{L-1} \mathbb{P}\left(\left\{\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_1} \wedge Z_l \leq \tau\right\}\right) + \mathbb{P}\left(\left\{\{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_2} \wedge Z_l > \tau\right\}\right) \\
&\leq \sum_{i=0}^{L-1} \mathbb{P}\left(Z_l \leq \tau \middle| \{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_1}\right) + \mathbb{P}\left(Z_l \geq \tau \middle| \{\omega_i^{(l)}\}_{i=1}^n \sim F_{a_2}\right) \\
&\leq 2L \exp\left(-\frac{1}{8} n \|F_{a_1} - F_{a_2}\|_1^2\right), \qquad \text{(Lemma 12)}
\end{aligned}
$$

which concludes the proof. □

C.3 Regret Analysis

Finally, we analyze the regret of Algorithm 1. We split the analysis in three cases. First, in Section C.3.1 and Section C.3.2 we analyze the regret for $\bar{\theta} > \theta_{\text{lazy}}$ and $\bar{\theta} \leq \theta_{\text{lazy}}$ respectively. Then, in Section C.3.3 we analyze the regret for the case where $\|F_{a_1} - F_{a_2}\|_1 \leq \beta$. Finally, in Section C.3.4 we combine these results and prove Theorem 3 by properly choosing n, L, α and β.

C.3.1 Regret Analysis for Lazy Types

Lemma 14 (Regret lazy types)**.** *Suppose that $\bar{\theta} > \theta_{\text{lazy}}$ and $\|F_{a_1} - F_{a_2}\|_1 > \beta$. Then, Algorithm 1 guarantees that:*

$$
R_T \leq 2\|F_{a_1} - F_{a_2}\|_1 n + 2T(\|F_{a_1} - F_{a_2}\|_1 + \alpha) \exp\left(-\frac{1}{8} n \|F_{a_1} - F_{a_2}\|_1^2\right).
$$

Proof. Denote by $p^\star$ the null contract. Recall that, from Lemma 10 the null contract is optimal for types greater than θ_{lazy}. Furthermore, since $\theta > \theta_{\text{lazy}}$, it must be that $\theta_{\text{lazy}} < 1$. Hence, it holds that Algorithm 1 tests whether $\bar{\theta}$ is lazy. It follows that:

$$
\begin{aligned}
R_T &= \sum_{t=1}^{n} \left(U^P(p^\star, \bar{\theta}) - U^P(p^\star_{\theta_{\text{lazy}}}, \bar{\theta})\right) + \sum_{t=n+1}^{T} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})]\right) \\
&\leq 2\|F_{a_1} - F_{a_2}\|_1 n + \sum_{t=n+1}^{T} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})]\right) \qquad \text{(Lemma 11)}
\end{aligned}
$$

Next, fix any $t \geq n + 1$ and analyze $U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})]$ using the law of total expectation. Denote by $\mathcal{E}_0 = \{Z_0 \leq \tau\}$. Observe that, under $\mathcal{E}_0$, it holds

$$
U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) \mid \mathcal{E}_0] = 0,
$$

as Algorithm 1 plays the optimal contract until the end of the horizon. Instead, when $\mathcal{E}_0^{\complement}$ holds, we split the analysis in two cases. First, consider the case when $t \leq \{n + 1, \ldots, nL\}$. In this case,

Algorithm 1 is playing $p^\star_\theta$ for some $\theta \in [0, \theta_{\text{lazy}}]$ where $p^\star_\theta$ is as in Lemma 10. Thus, it holds that:

$$\begin{aligned} U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) | \mathcal{E}_0^{\complement}] &\leq \sup_{\theta \in [0, \theta_{\text{lazy}}]} U^P(p^\star, \bar{\theta}) - U^P(p^\star_\theta, \bar{\theta}) \\ &\leq 2\|F_{a_1} - F_{a_2}\|_1 \qquad \text{(Lemma 11)} \end{aligned}$$

If, instead, $t \geq nL + 1$, Algorithm 1 plays $\tilde{p} = p^\star_\theta + \alpha(r - p^\star_\theta)$ for some $\theta \in [0, \theta_{\text{lazy}}]$. It follows that:

$$\begin{aligned} U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) | \mathcal{E}_0^{\complement}] &\leq \sup_{\theta \in [0, \theta_{\text{lazy}}]} U^P(p^\star, \bar{\theta}) - U^P(p^\star_\theta + \alpha(r - p^\star_\theta), \bar{\theta}) \\ &\leq 2\|F_{a_1} - F_{a_2}\|_1 + \alpha. \qquad \text{(Lemma 11)} \end{aligned}$$

Combining these results, we have that:

$$\sum_{t=n+1}^{T} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})] \right) \leq 2T(\|F_{a_1} - F_{a_2}\|_1 + \alpha) \mathbb{P}\left(\mathcal{E}_0^{\complement}\right).$$

Using Lemma 13 we can upper bound $\mathbb{P}(\mathcal{E}_0^{\complement})$, obtaining:

$$R_T \leq 2\|F_{a_1} - F_{a_2}\|_1 n + 2T(\|F_{a_1} - F_{a_2}\|_1 + \alpha) \exp\left(-\frac{1}{8} n \|F_{a_1} - F_{a_2}\|_1^2\right),$$

which concludes the proof. □

C.3.2 Regret Analysis for Non-Lazy Types

Lemma 15 (Regret for non-lazy types)**.** *Suppose that* $\|F_{a_1} - F_{a_2}\|_1 > \beta$ *and* $\bar{\theta} \leq \theta_{\text{lazy}}$*. It holds that:*

$$R_T \leq 2nL\|F_{a_1} - F_{a_2}\|_1 + 2T\left(\left(\frac{2^{-L}}{\alpha} + \alpha\right) + 2(\|F_{a_1} - F_{a_2}\|_1 + \alpha) L \exp\left(-\frac{n\|F_{a_1} - F_{a_2}\|_1^2}{8}\right)\right).$$

Proof. Denote by $p^\star$ the optimal contract for the unknown true type $\bar{\theta}$. It holds that:

$$R_T = \sum_{t=1}^{nL} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})] \right) + \sum_{t=nL+1}^{T} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})] \right).$$

Focus on the first nL steps and observes that, by definition, Algorithm 1 plays an optimal contract for some type $\theta \in [0, 1]$. Thus, we can apply Lemma 11 to obtain:

$$\sum_{t=1}^{nL} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})] \right) \leq 2nL\|F_{a_1} - F_{a_2}\|_1.$$

Next, fix any $t \geq n + 1$ and analyze $U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})]$ using the law of total expectation. Precisely, we will use as event $\mathcal{E} = \{\bar{\theta} \in [s_L, e_L]\}$.

We first analyze the case when $\mathcal{E}$ holds. In this case, we first observe that $|\bar{\theta} - \hat{\theta}_L| \leq \frac{e_L - s_L}{2} \leq 2^{-L}$. Furthermore, Algorithm 1 plays $\tilde{p} = p^\star_{\hat{\theta}_L} + \alpha(r - p^\star_{\hat{\theta}_L})$ and we have that:

$$\begin{aligned} U^P(\tilde{p}, \bar{\theta}) &\geq U^P(p^\star_{\hat{\theta}_L}, b(\hat{\theta}_L, p^\star_{\hat{\theta}_L})) - \left(\frac{2^{-L} C_{a_1}}{\alpha} + \alpha\right) && \text{(Lemma 19)} \\ &= U^P(p^\star_{\hat{\theta}_L}, \hat{\theta}_L) - \left(\frac{2^{-L} C_{a_1}}{\alpha} + \alpha\right) \\ &\geq U^P(p^\star + \alpha(r - p^\star), \hat{\theta}_L) - \left(\frac{2^{-L} C_{a_1}}{\alpha} + \alpha\right) && (p^\star_{\hat{\theta}_L} \text{ is opt. for } \hat{\theta}_L) \\ &\geq U^P(p^\star, b(\bar{\theta}, p^\star)) - 2\left(\frac{2^{-L} C_{a_1}}{\alpha} + \alpha\right) && \text{(Lemma 19)} \end{aligned}$$

In particular, when we used Lemma 19, we have used that types θ_1, θ_2 for which $|\theta_1 - \theta_2| \leq \epsilon$ satisfy $b(\theta_1, p) \in \mathcal{B}_{C_{a_1} \cdot \epsilon}(\theta_2, p)$ for any contract p. Thus, under $\mathcal{E}$, we obtain that:

$$U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) | \mathcal{E}] \leq 2 \left(\frac{2^{-L} C_{a_1}}{\alpha} + \alpha \right) \leq 2 \left(\frac{2^{-L}}{\alpha} + \alpha \right).$$

Next, suppose that $\mathcal{E}$ does not hold. Then, we split the analysis under two events $\mathcal{E}_1^\complement$ and $\mathcal{E}_2^\complement$ which are disjoint and that are such that $\mathcal{E}^\complement = \mathcal{E}_1^\complement \cup \mathcal{E}_2^\complement$. The first one, $\mathcal{E}_1^\complement$ is when Algorithm 1 plays the null contract as the first test on lazy types failed. In this case, Lemma 11 yields:

$$U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) | \mathcal{E}_1^\complement] \leq 2 \|F_{a_1} - F_{a_2}\|_1.$$

In the second case, *i.e.*, on $\mathcal{E}_2^\complement$, instead, we only know that Algorithm 1 plays $\tilde{p} = p^\star_\theta + \alpha(r - p^\star_\theta)$ for some $\theta \in [0, \theta_{\text{lazy}}]$. Hence, we have that:

$$\begin{aligned} U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta}) | \mathcal{E}_2^\complement] &\leq \sup_{\theta \in [0, \theta_{\text{lazy}}]} U^P(p^\star, \bar{\theta}) - U^P(p^\star_\theta + \alpha(r - p^\star_\theta), \bar{\theta}) \\ &\leq 2 \|F_{a_1} - F_{a_2}\|_1 + \alpha. \qquad \text{(Lemma 11)} \end{aligned}$$

Combining these results, we have that, for any $t \geq nL + 1$:

$$\begin{aligned} \left(U^P(p^\star, \bar{\theta}) - \mathbb{E}[U^P(p_t, \bar{\theta})] \right) &\leq 2 \left(\frac{2^{-L}}{\alpha} + \alpha \right) + (2 \|F_{a_1} - F_{a_2}\|_1 + \alpha) \left(\mathbb{P}(\mathcal{E}_1^\complement) + \mathbb{P}(\mathcal{E}_2^\complement) \right) \\ &= 2 \left(\frac{2^{-L}}{\alpha} + \alpha \right) + (2 \|F_{a_1} - F_{a_2}\|_1 + \alpha) \mathbb{P}(\mathcal{E}^\complement) \end{aligned}$$

Combining these results with Lemma 13, we have that:

$$R_T \leq 2nL \|F_{a_1} - F_{a_2}\|_1 + 2T \left(\left(\frac{2^{-L}}{\alpha} + \alpha \right) + 2(\|F_{a_1} - F_{a_2}\|_1 + \alpha) L \exp\left(- \frac{n \|F_{a_1} - F_{a_2}\|_1^2}{8} \right) \right),$$

which concludes the proof. □

C.3.3 Regret when Actions are Similar

Lemma 16 (Regret when F's are close)**.** *Suppose* $\|F_{a_1} - F_{a_2}\|_1 \leq \beta$. *It holds that* $R_T \leq 2T\beta$.

Proof. When $\|F_{a_1} - F_{a_2}\|_1 \leq \beta$, Algorithm 1 simply plays the null contract until the end of the horizon. The result is then a straightforward application of Lemma 11. □

C.3.4 Combining the Results

We are now ready to combine all the results to prove Theorem 3.

Proof of Theorem 3. We start by combining Lemma 14, Lemma 15 and Lemma 16. Observe that the upper bound on the regret in Lemma 14 is always larger than the one of Lemma 15 for any choice of the parameters α, L and n. Hence, it is sufficient to optimize the parameters to minimize the upper bound in Lemma 15 and Lemma 16.

We choose β, α, L and n as follows:

$$\beta = \frac{1}{\sqrt{T}} \tag{27}$$

$$L = \min\{T, \lceil \log_2(4T) \rceil\} \tag{28}$$

$$n = \min\left\{ \left\lfloor \frac{T}{L} \right\rfloor, \max\left\{ 1, \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T \|F_{a_1} - F_{a_2}\|_1^2 \right) \right\rfloor \right\} \right\} \tag{29}$$

$$\alpha = \min\left\{ 1, \sqrt{2^{-L}} \right\} \tag{30}$$

With this choice, we first observe that when $\|F_{a_1} - F_{a_2}\|_1 \leq \frac{1}{\sqrt{T}}$, Lemma 16 directly leads to $R_T \leq \sqrt{T}$. Thus, we now focus on $\|F_{a_1} - F_{a_2}\|_1 > \frac{1}{\sqrt{T}}$ and we prove that Lemma 15 yields a $\widetilde{\mathcal{O}}(\sqrt{T})$ bound for the regret, thus concluding the proof. We report here the result of Lemma 15 for completeness. We have that $R_T \leq h_1 + h_2 + h_3 + h_4$, where the different terms are as follows

$$h_1 = 2nL\|F_{a_1} - F_{a_2}\|_1, \quad h_2 = 2T\left(\frac{2^{-L}}{\alpha} + \alpha\right)$$
$$h_3 = 4T\|F_{a_1} - F_{a_2}\|_1 L \exp\left(-\frac{n\|F_{a_1} - F_{a_2}\|_1^2}{8}\right)$$
$$h_4 = 4T\alpha L \exp\left(-\frac{n\|F_{a_1} - F_{a_2}\|_1^2}{8}\right)$$

We proceed by showing that each h_i belongs to $\widetilde{\mathcal{O}}(\sqrt{T})$.

Upper bound on h_1 Suppose that $n = \lfloor T/L \rfloor$, which implies that

$$T \leq L + L\max\left\{1, \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor\right\}.$$

Furthermore, observe that this choice of n ensures that $h_1 \leq 2T\|F_{a_1} - F_{a_2}\|_1$. Since $L \in \mathcal{O}(\log(T))$ and $\|F_{a_1} - F_{a_2}\|_1 \in [T^{-1/2}, 1]$, we have that $h_1 \in \widetilde{\mathcal{O}}(\sqrt{T})$. Next, suppose that $n = 1$. Here, from $L \in \mathcal{O}(\log(T))$ we obtain $h_1 \leq L \in \widetilde{O}(\sqrt{T})$. Finally, suppose that $n = \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor$. In this case,

$$h_1 \in \mathcal{O}\left(L\frac{1}{\|F_{a_1} - F_{a_2}\|_1}\log(T)\right) \leq \mathcal{O}\left(L\sqrt{T}\log(T)\right)$$

Since $L \in \mathcal{O}(\log(T))$, $h_1 \in \widetilde{O}(\sqrt{T})$.

Upper bound on h_2 We proceed by cases. Suppose that $\alpha = 1$ and $L = T$. In this case, $\alpha \leq \sqrt{2^{-L}}$, and we have that $h_2 \leq 4T2^{-T} \in \widetilde{\mathcal{O}}(\sqrt{T})$. Next, suppose that $\alpha = 1$ and $L = \lceil \log_2(4T) \rceil$. Here, we have that $h_2 \leq 4T2^{-\lceil \log_2(4T) \rceil} \in \widetilde{\mathcal{O}}(\sqrt{T})$. Next, when $\alpha = \sqrt{2^{-L}}$ and $L = T$, $h_2 = 4T\sqrt{2^{-T}} \in \widetilde{\mathcal{O}}(\sqrt{T})$. Finally, when $\alpha = \sqrt{2^{-L}}$ and $L = \lceil \log_2(4T) \rceil$, $h_2 = 4T\sqrt{2^{-\lceil \log_2(4T) \rceil}} \in \widetilde{\mathcal{O}}(\sqrt{T})$.

Upper bound on h_3 First, suppose that $n = \lfloor T/L \rfloor$, which implies that

$$T \leq L + L\max\left\{1, \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor\right\}.$$

We observe that h_3 reduces to:

$$\begin{aligned}
h_3 &= 2T\|F_{a_1} - F_{a_2}\|_1 L \exp\left(-\frac{T\|F_{a_1} - F_{a_2}\|_1^2}{8L}\right) \\
&\leq 2T\|F_{a_1} - F_{a_2}\|_1 L \exp\left(-\frac{1}{8L}\right) && (\|F_{a_1} - F_{a_2}\|_1 \geq 1/\sqrt{T}) \\
&\leq 2T\|F_{a_1} - F_{a_2}\|_1 L \\
&\leq 2\|F_{a_1} - F_{a_2}\|_1 L\left(L + L\max\left\{1, \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor\right\}\right) \\
&\leq 2L^2 + 2L^2\max\left\{1, \left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor\right\} \\
&\leq 2L^2 + 2L^2\max\left\{1, \left\lfloor 8\sqrt{T}\log\left(T\right) \right\rfloor\right\} && (\|F_{a_1} - F_{a_2}\|_1 \geq 1/\sqrt{T})
\end{aligned}$$

Since $L \in \mathcal{O}(\log(T))$, we obtained that $h_3 \in \widetilde{\mathcal{O}}(\sqrt{T})$.

Next, suppose that $n = 1$. In this case, $1 \geq \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right)$, which implies that

$$T\|F_{a_1} - F_{a_2}\|_1 \leq \frac{\exp(\|F_{a_1} - F_{a_2}\|_1^2/8)}{\|F_{a_1} - F_{a_2}\|_1} \leq \frac{1}{\|F_{a_1} - F_{a_2}\|_1}.$$

Thus, we have that:

$$\begin{aligned}
h_3 &= 2T\|F_{a_1} - F_{a_2}\|_1 L \exp\left(-\frac{n\|F_{a_1} - F_{a_2}\|_1^2}{8}\right) \\
&\leq 2T\|F_{a_1} - F_{a_2}\|_1 L \\
&\leq \frac{2L}{\|F_{a_1} - F_{a_2}\|_1} \\
&\leq 2L\sqrt{T} && (\|F_{a_1} - F_{a_2}\|_1 \geq 1/\sqrt{T})
\end{aligned}$$

Since $L \in \mathcal{O}(\log(T))$, we have $h_3 \in \widetilde{\mathcal{O}}(\sqrt{T})$.

Finally, consider the last case where n is attained at $\left\lfloor \frac{8}{\|F_{a_1} - F_{a_2}\|_1^2} \log\left(T\|F_{a_1} - F_{a_2}\|_1^2\right) \right\rfloor$. In this case:

$$\begin{aligned}
h_3 &= 2T\|F_{a_1} - F_{a_2}\|_1 L \exp\left(-\log(T\|F_{a_1} - F_{a_2}\|_1^2)\right) \\
&= 2T\|F_{a_1} - F_{a_2}\|_1 L \frac{1}{T\|F_{a_1} - F_{a_2}\|_1^2} \\
&= \frac{2L}{\|F_{a_1} - F_{a_2}\|_1} \\
&\leq 2L\sqrt{T}. && (\|F_{a_1} - F_{a_2}\|_1 \geq 1/\sqrt{T})
\end{aligned}$$

Since $L \in \mathcal{O}(\log(T))$, we have $h_3 \in \widetilde{\mathcal{O}}(\sqrt{T})$.

Upper bound on h_4 Here, we aim to upper bound α with $\|F_{a_1} - F_{a_2}\|_1$. This allows to directly exploit the result for h_3 to show that this term belongs to $\widetilde{\mathcal{O}}(\sqrt{T})$. To this end, observe that $\alpha^2 \leq 2^{-L} \leq \|F_{a_1} - F_{a_2}\|_1$ holds if and only if $L \geq \log_2(1/\|F_{a_1} - F_{a_2}\|_1^2)$. Since $\|F_{a_1} - F_{a_2}\|_1 \geq 1/\sqrt{T}$, however, we have that this is equivalent to $L \geq \log_2(T)$, which is always true for our choice of L.

This concludes the proof. □

D Proof of Theorem 4

Proof of Theorem 4. We build two instances $\mathfrak{I}_0$ and $\mathfrak{I}_1$ as follows. Both instances have two actions with $c_{a_1} = 1$ and $c_{a_2} = 0$. Moreover, in both instances, we have two outcomes with rewards $r_{\omega_1} = 1$ and $r_{\omega_2} = 0$. The outcome distribution for a_1 is $F_{a_1,\omega_1} = \frac{1}{2} + 4\epsilon$ and $F_{a_1,\omega_2} = \frac{1}{2} - 4\epsilon$, while for a_2 are $F_{a_2,\omega_1} = F_{a_2,\omega_2} = \frac{1}{2}$. The two instances only differ by the type θ and in particular, in the first instance $\theta_0 = \frac{40\epsilon}{\frac{1}{\epsilon}+8}$ and $\theta_1 = \frac{24\epsilon}{\frac{1}{\epsilon}+8}$. Note that in both instances it is optimal to put a zero payment on the second outcome ω_2. Let $x = p_{\omega_1}$ be the payment on the first outcome.

Note that an agent of type θ plays action a_1 if and only if $(\frac{1}{2} + 4\epsilon)x - \theta \geq \frac{x}{2}$, and thus if and only if $x \geq \frac{\theta}{4\epsilon}$. Thus, in instance $\mathfrak{I}_0$, the principal utility and mean of the Bernoulli describing the feedback distribution $\tau(x)$ (of observing outcome ω_1) are:

$$U_0^P(x) = \begin{cases} \frac{1}{2}(1-x) & \text{if} \quad x < \frac{10}{\frac{1}{\epsilon}+8} \quad \text{and} \\ (\frac{1}{2} + 4\epsilon)(1-x) & \text{otherwise} \end{cases} \qquad \tau_0(x) = \begin{cases} \frac{1}{2} & \text{if} \quad x < \frac{10}{\frac{1}{\epsilon}+8} \\ \frac{1}{2} + 4\epsilon & \text{otherwise} \end{cases},$$

while for instance $\mathfrak{I}_1$ are:

$$U_1^P(x) = \begin{cases} \frac{1}{2}(1-x) & \text{if} \quad x < \frac{6}{\frac{1}{\epsilon}+8} \quad \text{and} \\ (\frac{1}{2} + 4\epsilon)(1-x) & \text{otherwise} \end{cases} \qquad \tau_1(x) = \begin{cases} \frac{1}{2} & \text{if} \quad x < \frac{6}{\frac{1}{\epsilon}+8} \\ \frac{1}{2} + 4\epsilon & \text{otherwise} \end{cases}.$$

It is clear that the $x \in \{0, \frac{6}{\frac{1}{\epsilon}+8}, \frac{10}{\frac{1}{\epsilon}+8}\}$ are dominating actions in all instances in terms of both feedback and rewards. Then, a standard change-of-measure argument allows us to conclude the lower bound. Define $N_0(T)$ the number of times $x_0 = 0$ was played up to time T, and $N_1(T)$ and $N_2(T)$ the number of times $x_1 = \frac{6}{\frac{1}{\epsilon}+8}$ and $x_2 = \frac{10}{\frac{1}{\epsilon}+8}$ was played. Then, there exists an absolute constant $C > 0$ such that:

$$\mathrm{KL}(\nu_0, \nu_1) \leq C\mathbb{E}_{\nu_0}[N_1(T)]\epsilon^2,$$

moreover, direct computations show that for some constant $c > 0$

$$U_0^P(x_0) - U_0^P(x_1) \geq c\epsilon, \quad U_1^P(x_1) - U_1^P(x_0) \geq c\epsilon, \quad \text{and} \quad U_1^P(x_1) - U_1^P(x_2) \geq c\epsilon.$$

Hence

$$\mathbb{E}_{\nu_0}[R_T] \geq c\epsilon\mathbb{E}_{\nu_0}[N_1(T)], \qquad \mathbb{E}_{\nu_1}[R_T] \geq c\epsilon\mathbb{E}_{\nu_1}[T - N_1(T)].$$

Now, if $\mathbb{E}_{\nu_0}[N_1(T)] \geq \frac{T}{4}$ then $\mathbb{E}_{\nu_0}[R_T] \geq \Omega(\epsilon T)$. On the other hand, if $\mathbb{E}_{\nu_0}[N_1(T)] < \frac{T}{4}$ then Markov's inequality gives:

$$\mathbb{P}_{\nu_0}\left(N_1(T) > \frac{T}{2}\right) \leq \frac{1}{2},$$

and (with $\epsilon = \Theta(1/\sqrt{T})$), Pinsker's inequality gives

$$\begin{aligned}\mathbb{P}_{\nu_1}\left(N_1(T) > \frac{T}{2}\right) &\leq \mathbb{P}_{\nu_0}\left(N_1(T) > \frac{T}{2}\right) + TV(\nu_0, \nu_1) \\ &\leq \frac{3}{4}\end{aligned}$$

which can be rewritten as $\mathbb{P}_{\nu_1}(N_1(T) \leq T/2) \geq 1/4$. Finally, again by Markov's inequality:

$$\begin{aligned}\mathbb{E}_{\nu_1}[R_T] &\geq c\epsilon\mathbb{E}_{\nu_1}[T - N_1(T)] \\ &\geq c\epsilon\frac{T}{2}\mathbb{P}_{\nu_1}\left[T - N_1(T) \geq \frac{T}{2}\right] \\ &= \Omega\Big(\epsilon T\mathbb{P}_{\nu_1}\left[N_1(T) \leq \frac{T}{2}\right]\Big) \\ &\geq \Omega(\epsilon T),\end{aligned}$$

concluding the proof. □

E Useful Results

E.1 The $F_{a_2}^\top r \geq F_{a_1}^\top r$ case

In this section, we prove that when $F_{a_2}^\top r \geq F_{a_1}^\top r$ holds, there exists a trivial algorithm that yields $R_T = 0$ for any sequence $\{\theta_t\}_{t=1}^T$.

Lemma 17 (The $F_{a_2}^\top r \geq F_{a_1}^\top r$ case)**.** *Suppose that $F_{a_2}^\top r \geq F_{a_1}^\top r$. Then, there exists an algorithm for which $R_T = 0$.*

Proof. Let $\bar{p} = (0, \ldots, 0)$ and observe that, for any $\theta \in [0, 1]$, we have that

$$\operatorname*{argmax}_{a\in\{1,2\}} F_a^\top \bar{p} - \theta \cdot C_a = -\operatorname*{argmin}_{a\in\{1,2\}} \theta \cdot C_a.$$

Hence, since $C_{a_2} \leq C_{a_1}$, $a_2 \in \mathcal{B}(\theta, \bar{p})$ for all $\theta \in [0, 1]$. Furthermore, since $F_{a_2}^\top r \geq F_{a_1}^\top r$, it also holds that $a_2 \in b(\theta, \bar{p})$ for all types θ. As a consequence, we have that $U^P(\bar{p}, b(\theta, \bar{p})) = F_{a_2}^\top r$ for all

$\theta \in [0,1]$. Next, we observe that:

$$\begin{aligned}
\textsc{Opt} &= \sup_{p\in[0,1]^m} \left(\sum_{t=1}^{T} \mathbb{1}\{a_1 \in b(\theta_t, p)\} F_{a_1}^\top (r-p) + \mathbb{1}\{b(\theta_t,p) = a_2\} F_{a_2}^\top (r-p) \right) \\
&\leq \sup_{p\in[0,1]^m} \left(\sum_{t=1}^{T} \mathbb{1}\{a_1 \in b(\theta_t, p)\} F_{a_1}^\top r + \mathbb{1}\{b(\theta_t,p) = a_2\} F_{a_2}^\top r \right) && (p \geq 0) \\
&\leq \sum_{t=1}^{T} F_{a_2}^\top r && (F_{a_2}^\top r \geq F_{a_1}^\top r) \\
&= \sum_{t=1}^{T} U^P(\theta_t, b(\theta_t, \bar{p})). && (U^P(\bar{p}, b(\theta, \bar{p})) = F_{a_2}^\top r \text{ for all } \theta \in [0,1])
\end{aligned}$$

Hence, the constant strategy $\bar{p}$ is optimal for any sequence $\{\theta_t\}_{t=1}^T$ thus concluding the proof. □

E.2 Concentration Inequality

We report for completeness a statement of the Hoeffding's inequality for the concentration of i.i.d. random variables with values in $[0,1]$.

Lemma 18 (Hoeffding's inequality [17])**.** *Consider $X_1, \ldots, X_n$ be n i.i.d. random variables in $[0,1]$ whose expected value is μ. Then, for $\delta > 0$, it holds that:*

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu \geq \delta \right) \leq \exp\left(-2n\delta^2\right).$$

E.3 Contract Linearization

Lemma 19 (Proposition 2.4. in [30])**.** *Let $\epsilon > 0$ and $\rho : [0,1] \to \{a_1, a_2\}$ be a generic mappings from types to actions, and let $p \in [0,1]^m$. Suppose that $\rho(\theta) \in \mathcal{B}_\epsilon(\theta, p)$ holds for all θ. Then, let $\tilde{p} = p + \alpha(r-p)$ for some $\alpha > 0$. It holds that:*

$$U^P(\tilde{p}, b(\theta, \tilde{p})) \geq U^P(p, \rho(\theta)) - \left(\frac{\epsilon}{\alpha} + \alpha \right).$$

E.4 Optimization of Upper Semicontinuous Functions

We recall the definition of upper semicontinuity. Assume that (X, d) is a metric space and consider $f : X \to \mathbb{R}$. Then, f is upper semicontinuous at $x_0 \in X$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $f(x) < f(x_0) + \epsilon$ for all x such that $d(x, x_0) < \delta$. We now state a well-known result for optimization of upper semicontinuous functions.

Theorem 6. *If C is a compact set, and $f : C \to \mathbb{R}$ is upper semicontinuous, then f attains a maximum on C.*

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The statements of the abstract and intro correctly identify the paper's contribution.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have a final section discussing limitations and possible extensions.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The paper describes in full detail the assumptions of the paper.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: The paper does not include experiments.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The paper is theoretical, and does not pose immediate, concrete societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: This research has not been conducted by employing any non-trivial use of LLMs.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.