

Learning Approximately Optimal Contracts

Alon Cohen¹, Argyrios Deligkas², and Moran Koren³

¹ Tel-Aviv University and Google Research alonco@tauex.tau.ac.il

² Royal Holloway, University of London Argyrios.Deligkas@rhul.ac.uk

³ Tel Aviv University me@mkoren.org

Abstract. In principal-agent models, a principal offers a contract to an agent to preform a certain task. The agent exerts a level of effort that maximizes her utility. The principal is oblivious to the agent’s chosen level of effort, and conditions her wage only on possible outcomes. In this work, we consider a model in which the principal is unaware of the agent’s utility and action space: she sequentially offers contracts to identical agents, and observes the resulting outcomes. We present an algorithm for learning the optimal contract under mild assumptions. We bound the number of samples needed for the principal obtain a contract that is within ϵ of her optimal net profit for every $\epsilon > 0$. Our results are robust even when considering risk averse agents. Furthermore, we show that when there only two possible outcomes, or the agent is risk neutral, the algorithm’s outcome approximates the optimal contract described in the classical theory.

1 Introduction

Recent technological advances have changed the relationship between firms and employees dramatically. The rapid change in employees’ required skill set combined with the availability of off-shore, qualified, cheap hiring alternatives drove employers to adopt employees on a short-term, task specific bases. This hiring model, dubbed *dynamic workforce* or *gig-economy*, has been growing exponentially over the decade. In [8] it was found that 20-30% of US workers are employed independently, at least partly, and predicts this trend will continue with popularity of the “Lean Startup” business model, and the appearance of platforms such as “Amazon Mechanical Turk”.

The aforementioned labor market changes have a profound effect on the information available to firms when it comes to making hiring decisions. In the traditional workspace, hiring is often a long-term, expensive procedure. Nowadays hiring is cheap and short-termed. As a result, an employer has less information on the character and quality of her employees, thus lacking the knowledge on how to best align their incentives with those of the firm. In this work we study a model more suitable to the new “dynamic workforce”, and, using this model, we try to learn the best wages an employer should offer to her employees.

Economists use the term *agency problems* to describe models like the above. In such models one party, the *principal*, offers a *contract* to another party, the

agent, to perform a task [25]. In his seminal paper, Ross [30] introduced the term *principal-agent* to capture these models. In the basic model, the outcome of the task is chosen randomly from a distribution determined by the *level of effort* invested by the agent. A higher effort level induces a distribution in which the probability of a better outcome is higher than in lower effort levels. On the other hand, a higher level causes the agent greater disutility. Ross [30] assumes that the principal is *risk-neutral* while the agent is *risk-averse*, therefore the incentives of both parties are *not* fully aligned.* Hence, the effort level that is optimal from the perspective of the principal, is not necessarily the optimal one for the agent. To bridge this gap, the principal offers the agent a contract in which the agent is rewarded for any additional effort. The lion's share of previous work assume that the principal, while oblivious to the agent's choice, has full information about the agent's utility structure *and* the set of levels of effort she can choose from *and* the probability distribution associated with every level of effort [30,18,21,17,22]. These assumptions seem confining, especially when considering the motivating scenario of a dynamic workplace.

Holmstrom, in his Nobel Prize lecture [22], identified this gap in theory as one of the main challenges in current research. A first attempt to bridge this gap was [9]. Carroll [9] presented a model where a principal has only partial knowledge on the agent's action space. He assumed that the agent is risk neutral and thus, if the principal have known the agent's complete action space, the optimal contract would have been linear. In this setting, the principal can guarantee the net profit from best linear contract pertaining the actions she knows. That is her actual net profit will never be lower. An early attempt to approach the agency problem from the lens of theoretical computer science was made in [20]. In the model of [20], there are several types of agents, each one with her own utility function, set of effort levels and effort costs, and probability distributions over the outcomes; all of which unknown to the principal. They considered a repeated setting with T rounds, where in every round the type of the agent is chosen i.i.d. from an unknown probability distribution. Their goal was to find, before round T , the optimal contract from a predefined finite set S . They followed a multi-armed bandit approach [29] and derived an algorithm that finds an approximately-optimal contract. However, the contract they find is approximately-optimal *only* with respect to the best contract in S . They provide no theoretical guarantees on approximating the optimal contract overall, but only for the case of one effort level for the agent, i.e. reject the contract or accept it, known as *posted-price auctions*.* In [14], the authors study a similar scenario in which the principal

* When given the choice between participating in some lottery X or receiving $E(X)$ with probability one, a risk-averse agent will strictly prefer the latter while a risk-neutral agent is indifferent. In the Von Neumann Morgenstern utility theory, if an agent is risk-averse, then she has a concave utility function; if she is risk-neutral, then her utility is linear.

* As they highlight, it is not generally clear whether the best contract from S can provide a good theoretical guarantees for the general problem of dynamic contract design.

knows only the expected output generated by each contract, that is, the agent’s action set is unknown to the principal. As in [9], they assume that the agent’s utility is linear and that limited liability holds (i.e., no negative payments) and show that the worst-case expected profit for the principal is guaranteed by a linear contract. In addition, they provide tight approximation guarantees for linear contracts. In this paper we follow a similar route to [20] and [14]. We assume the principal has *zero* information about the agent; we assume unknown utility for the agent, unknown set of effort levels and their associated costs, and unknown probability distribution for any effort level. The only knowledge the principal has is the set of outcomes and their corresponding profits. Our model extends the setting of [20] as we assume that the contract set is continuous. We extend the setting of [9] and [14] in two folds: (1) Our set of *monotone-smooth* contracts is richer than the set of linear contracts. (2) We do not assume any functional form of agent utilities, we only assume that agents are risk averse (concave utility). In the next section we illustrate the importance of this generalization in this context.

1.1 The challenge of learning contracts with risk averse agents.

In contract theory, we study the tension between a principal and an agent. This tension is the result of misalignment between the incentives of both players. On the one hand, the principal wishes to maximize her profit, and is indifferent towards the cost the agent endures during effort; on the other hand, the agent is assumed to be risk averse, and wishes to minimize the volatility of her expected wage (*ceteris paribus*). In the classical PA model, the principal knows the full structure of the agent’s decision problem, and can offer a contract that serves her best yet minimizing the aforementioned misalignment with the agent. To illustrate this, we show an example where the agent may choose between three levels of effort and there are two possible outcomes. As we wish to illustrate a general point, we will keep the exact utility function and cost structure implicit and just mention that the utility is concave and higher costs induce “better” output distributions. The following figure depicts the space of contracts from which the principal may offer, and the resulting effort level the agent chooses, given each possible contract. Note that unlike [20], the space of contracts is continuous. By [27], the optimal contract can be found in any of the interaction points $\{A, B, C\}$. To see why this is the case, note, for example, that when offered the contract in point A , the agent is indifferent between accepting and rejecting the contract, and between investing medium effort or high effort. Thus, based on the model assumptions, she will choose to invest the high level of effort. Not only that, but this is the contract where the highest level of effort is achieved for the lowest possible expected pay. Similarly, the contract B , is the “cheapest” contract under which the agent chooses the middle effort level. Note that contract B is both cheaper, and yields lower expected output than contract A . When examining the principal-optimal contract, it is unclear which of the effects dominates. Therefore, in the general case, the optimal contract can be in any of A, B, C . When the

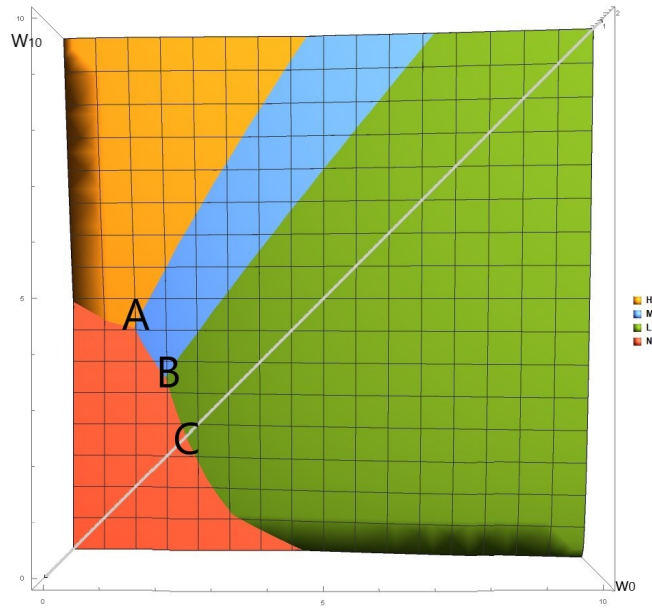


Fig. 1. The agent's optimal choice in a PA model with two possible outputs, three levels of effort: e_H, e_M, e_L , N is the rejection zone, and a concave utility.

principal is fully aware of the agent's incentive structure (i.e., her utility and effort-cost pairs), finding the optimal contract is straightforward.

If we consider an uninformed the principal, she can attempt to learn an optimal contract using the following pseudo-algorithm: (1) create a discretized version of the contract space (2) Find the contract on the grid which yields the optimal profit. Naively, using this approach, we will find a contract that is at most ε away either A, B or C . Note that this does not preclude the possibility that the contract that we find is ε away from point A , but the absolute optimal contract is actually contract B . In extreme cases, the difference in the principal's profit between the algorithms' result and the theoretically optimal one can be arbitrarily high (see for example $u(w) = 1 - \frac{1}{w}$). Furthermore, using a finer grid may lead to different contract which is arbitrarily away, thus convergence of the process is not guaranteed. In this paper, we provide a set of contracts, and algorithm, and an appropriate discretization for which the process converges to a single maximal contract. In addition, we detail two cases, for which the resulting contract is the absolute possible best.

Further related work. As we have already explained, the majority of works so far has focused on the full information setting and its variants: [6] studies a model with many agents; [5] studies a setting with externalities for the agent; [13,15] study principal-agent models with combinatorial structures over the actions of the agent, or the possible outcomes. Additionally, in [15], the authors also

discuss a black-box model and present sample-complexity results. More recently, a new line of work on Bayesian settings for the principal-agent model has emerged [2,10,11,19] where the agent type comes from a known probability distribution.

Our work lies in the intersection of principal-agent models and multi-armed bandit theory; [20] provide an excellent overview of literature in the field. [32,31,33] study a repeated setting where the principal interacts with the agent for multiple periods. [12] empirically compare several learning algorithms in a setting similar to [20].

The Lipschitz Bandit problem [1] is a generalization of the multi-armed bandit problem in which the set of arms comes from some compact space, and the expected reward of each arm is a Lipschitz function of the arm. It has received much attention from the bandit theory community over the years [24,4,7,23,26].

Our contributions. In this work we make several contributions to both economic and computer science theory.

- We contribute to the study of agency problems by extending the literature to a setting where the principal is oblivious to the agent decision problem structure.
- We study this setting in a the context of risk averse agents. An extension which is the driving force of the canonical theoretical principal-agent model. To the best of our knowledge, our paper is first in the introduction of agent risk aversion into the theoretical study of data driven decision problems.
- We introduce a novel set of contracts we call *monotone-smooth* contracts; a large subset of monotone contracts set studied in [20].
- We complement this definition with a suitable discretization of the contract space. Unlike [20], we show that for any monotone-smooth contract there exists a contract in the discretized space for which the principal’s expected net profit is ϵ -approximated. As far as we are aware, this is the first work to do so.
- Moreover, our result does not assume any specific agent utility function, but rather only mild assumptions. This allows to apply machineries from multi-armed bandit theory to find a contract ϵ -optimal against any monotone-smooth contract.
- Finally, we present two fundamental cases in which economic theory suggests the learned contract of our algorithm is ϵ -optimal against *any* contract the principal may offer. The first case is when there are only two possible outcomes. The second case is when there are many outcomes and the agent is risk neutral.

2 Preliminaries

In what follows, $[m^*] := \{0, 1, \dots, m\}$ and $[m] := \{1, \dots, m\}$ for every natural m .

We study principal-agent problems with k outcomes and n effort levels. Let $0 < \pi(1) < \pi(2) < \dots < \pi(k)$ denote the value the principal gets under outcome $i \in [k]$. We assume that $\pi(k) = H$, hence the values are bounded.

A *contract* $w = (w(1), \dots, w(k))$ specifies a positive payment to the agent for every outcome; namely, $w(i)$ is the wage the principal pays the agent for outcome i .

Upon receiving a contract, the agent chooses an effort level $e \in [n^*]$. Every effort level is associated with a probability distribution f_e over the set of outcomes and a cost $c(e)$; $f_e(j)$ is the probability of realizing outcome j when the agent chooses the effort level e . In effort level 0, the agent rejects the contract, her utility is zero under any contract, and by convention the value for the principal is zero. We assume that the effort levels are ordered, i.e., $1 \prec \dots \prec n$ (the order will be formally defined in Assumption 1). We follow the literature and assume that the agent has a von Neumann-Morgenstern utility. For a contract $w = (w(1), \dots, w(k))$ the agent chooses effort level of $\hat{e}(w)$ as to maximize her utility, defined as

$$U(w, e) = \sum_{j=1}^k f_e(j) \cdot u(w(j)) - c(e),$$

where u is a monotonically-increasing concave function. Hence, $\hat{e}(w) = \arg \max_{e \in [n^*]} U(w, e)$.

The principal is risk-neutral; when she offers contract w to the agent, her expected net profit from the contract is,

$$V(w) = \sum_{j=1}^k f_{\hat{e}(w)}(j) \cdot (\pi(j) - w(j)) .$$

To ensure that higher effort levels yield higher expected profit for the principal, the literature commonly lays down some assumptions about the outcome distributions.

Assumption 1 (First-order Stochastic Dominance (FOSD)) *A probability distribution associated with higher effort first order stochastically dominates a probability distribution associated with lower effort. Formally, if $e \succ e'$, then for every $j \in [k]$ it holds that $\sum_{i=j}^k f_e(i) \geq \sum_{i=j}^k f_{e'}(i)$.*

Note that the assumption is equivalent to the following. For every pair of effort levels $e \succ e'$ and for every sequence of real numbers $a(1) \leq \dots \leq a(k)$,

$$\sum_{i=1}^k f_e(i) \cdot a(i) \geq \sum_{i=1}^k f_{e'}(i) \cdot a(i) . \quad (1)$$

Additionally, to break ties between effort levels we assume the following.

Assumption 2 *The agent will choose the higher effort when indifferent between two or more levels of effort.*

In this work, the principal is faced with a stream of agents. The agents are all different but identical—they share a common utility function, effort levels,

costs from effort, and outcome distributions associated with each effort level. The principal proceeds in rounds $t = 1, 2, \dots$. On round t , the principal offers a contract w_t to the agent associated with this round. The agent privately chooses effort level $\hat{e}(w_t)$ unknown to the principal. The principal observes only the outcome i_t independently drawn from $f_{\hat{e}(w_t)}$, and consequently gets a net profit of $\pi(i_t) - w_t(i_t)$.

In what follows, for $\epsilon > 0$, the goal of the principal is to find an ϵ -optimal contract in the minimum number of rounds. A contract w is ϵ -optimal if $V(w) \geq V(w') - \epsilon$ for every $w' \in W$, for a set of contracts W to be defined in the sequel.

2.1 Multi-armed Bandit

In the multi-armed bandit problem [29], a decision maker sequentially collects rewards from a given set of arms. In each round, the decision maker chooses a single arm, and observes an independent sample from a reward distribution associated with that arm. In our case, the goal of the decision maker is, after a predetermined number of rounds, to select an ϵ -optimal arm; that is, an arm whose expected reward is at most ϵ less than the expected reward of any arm.

When the set of arms is finite, of size N , and the rewards are bounded in $[0, B]$, the seminal work of [16] presents an algorithm called MEDIANELIMINATION with the following guarantee.

Theorem 1 ([16]). *The MEDIANELIMINATION(ϵ, δ) returns an ϵ -optimal arm with probability at least $1 - \delta$ after $O((NB^2/\epsilon^2) \cdot \log(1/\delta))$ rounds.*

In our problem, each contract can be seen as an arm. The expected reward of each arm is exactly the principal's utility associated with this contract. It is then expected that the principal would simply execute MEDIANELIMINATION on the space of contracts to obtain an ϵ -optimal one. However, the space of contracts is not finite which is crucial for MEDIANELIMINATION to run. In the sequel we show how to overcome this difficulty by discretizing the space of contracts, and running MEDIANELIMINATION over the discretization.

3 Main Technical Result

In this section we present our algorithm and analyze its sample complexity, but before doing so let us first define the space of contracts W that we can learn. The algorithm is presented in Section 3.2.

3.1 Learnable Contracts

Let $w_0 > 0$ be a minimum wage for any outcome.

Definition 1 (B -bounded contract). *A contract w is B -bounded if $w_0 \leq w(i) \leq B$ for every $i \in [k]$.*

For a bounded contract, together with the assumption that the principal's profits are bounded, ensures that the principal's expected net profit $V(w)$ can be estimated statistically.

Definition 2 (Monotone-smooth contract). *A contract w is monotone-smooth if for every $i \in [k - 1]$ it holds that $0 \leq w(i + 1) - w(i) \leq \pi(i + 1) - \pi(i)$.*

For a monotone-smooth contract, Eq. (1) ensures that, keeping the contract fixed, the principal's utility cannot decrease if the agent increases her effort level. In the sequel, this property allows us to bound the difference in the principal's utility between two similar contracts.

We define W as follows.

$$W = \{w : w \text{ is monotone-smooth and } H\text{-bounded}\} .$$

We are aware that this set seems restrictive at first glance, yet we argue that in some important special cases, the principal's optimal net profit is achieved by a contract from this set. For example, when there are only two outcomes or the utility of the agent is linear (see Section 4).

3.2 Algorithm

Let $w^* \in W$ be an optimal contract in W , that is $V(w^*) \geq V(w)$ for all $w \in W$. The goal of our algorithm is to find an ϵ -optimal contract w , namely a contract for which $V(w^*) \leq V(w) + \epsilon$ within a predetermined number of rounds. We conjecture that it cannot be done in general therefore we make the following simplifying assumption.

Assumption 3 (Bounded Risk-Aversion) *The agent's utility from wage u is twice continuously-differentiable. Moreover, there exists a reference utility function r that is a monotonically increasing, twice-differentiable, concave function such that $u''(w)/u'(w) \geq r''(w)/r'(w)$ for all $w > 0$. This is equivalent to $w \mapsto u'(w)/r'(w)$ being monotone-nondecreasing in w .*

Intuitively, this assumption ensures that making small changes to a contract does not produce behavior by the agent that is drastically different than if the agent's utility would have been r instead of u .*

Our algorithm works as follows. The principal initially constructs a cover W_η of W , and then run MEDIANELIMINATION on W_η . Indeed, the main technical difficulty in this paper is in defining W_η properly so that the following result holds.

* An alternative, slightly less general, version of Assumption 3 is: Assume there exists a finite $\eta > 1$ such that $-xu''(w)/u'(w) > \eta$ for all $w > 0$. Note that the element on the left is the canonical Arrow-Pratt relative risk aversion measure of the agent [3,28], and the element of the right corresponds with the Arrow-Pratt relative risk aversion measure of the Isoelastic utility function $r(w) = \frac{w^{\eta-1}-1}{\eta-1}$.

Theorem 2. *Suppose that Assumptions 1 to 3 hold. Let $\eta < r'(2H) \cdot H/k$. There exists a contract space W_η such that for every contract $w \in W$, there is a contract $w' \in W_\eta$ for which $V(w) \leq V(w') + k\eta/r'(2H)$. Moreover, the size of W_η is at most $M = ((r(2H) - r(w_0))/\eta)^k$, and W_η can be constructed in time $O(M)$.*

The proof of the theorem is found in Section 3.3. Finally, we have our main result.

Theorem 3. *Suppose that Assumptions 1 to 3 hold. Let $\eta = \epsilon \cdot r'(2H)/2k$. Executing `MEDIANELIMINATION`($\epsilon/2, \delta$) on the set W_η produces the following guarantee. With probability at least $1 - \delta$, the algorithm outputs an ϵ -optimal contract after*

$$O\left(\left(\frac{4kH(r(2H) - r(w_0))}{\epsilon}\right)^{k+2} \cdot \log(1/\delta)\right)$$

rounds.

Proof. By Theorem 2 and by the choice of η , there is a $w' \in W_\eta$ for which $V(w') \leq V(w^*) + \epsilon/2$. By Theorem 1, with probability $1 - \delta$, `MEDIANELIMINATION` returns a contract $\hat{w} \in W_\eta$ such that $V(\hat{w}) \leq V(w') + \epsilon/2$. Combining both results we get

$$V(\hat{w}) \leq V(w') + \epsilon/2 \leq V(w^*) + \epsilon/2 + \epsilon/2 = V(w^*) + \epsilon ,$$

as required. Moreover, `MEDIANELIMINATION` is done in the following number of rounds:

$$O\left(\frac{|W_\eta|H^2}{(\epsilon/2)^2} \log(1/\delta)\right) = O\left(\left(\frac{4kH(r(2H) - r(w_0))}{\epsilon}\right)^{k+2} \cdot \log(1/\delta)\right) .\square$$

3.3 Discretization of the Contract Space

In this section we prove Theorem 2. We start by defining the notion of a coarse contract. To that end, we utilize the inverse function of r (that exists everywhere since r is increasing) which we denote by r^{-1} .

Definition 3 (η -coarse contract). *A contract W is η -coarse if there exists natural numbers l_0, l_1, \dots, l_{k-1} such that $w(1) = r^{-1}(r(w_0) + \eta \cdot l_0)$, and for $i \in [k - 1]$, $w(i + 1) = r^{-1}(r(w(i)) + \eta \cdot l_i)$.*

That is, in a coarse contract the ratios between wages of consecutive outcomes come from a discrete set of options. We define:

$$W_\eta = \{w : w \text{ is } \eta\text{-coarse and } 2H\text{-bounded}\} .$$

We prove the following.

Lemma 1. *The size of W_η is at most $M = ((r(2H) - r(w_0))/\eta)^k$. Moreover, W_η can be constructed in time $O(kM)$.*

Proof. The wage of outcome i has the form $r^{-1}(r(w_0) + \eta \cdot l)$ for a natural number l , and satisfies $w(i) \leq 2H$. Therefore, the number of choices for $w(i)$ is at most $(r(2H) - r(w_0))/\eta$. Since there are k outcomes, there must be at most M such contracts. To construct W_η we can go over all of its elements one-by-one, which takes $O(M)$ time. \square

Finally let $w \in W$, we need to show that there is $w' \in W_\eta$ such that $V(w) \leq V(w') + 2kH\eta$. We construct $w' \in W_\eta$ as follows. We let $l_0 = \lceil (r(w(1) - r(w_0))/\eta) \rceil$, and for $i \in [k-1]$, $l_i = \lceil (r(w(i+1)) - r(w(i)))/\eta \rceil$. Since r^{-1} is also monotonically-increasing, we have that $w'(1) \geq w(1)$ as well as $r(w'(i+1)) - r(w'(i)) \geq r(w(i+1)) - r(w(i))$ for $i \in [k-1]$ by construction. Clearly, w' is η -coarse and $w'(i) \geq w_0$, yet it remains to show that $w'(i) \leq 2H$. For that, we have the following lemma.

Lemma 2. *We have for all $i \in [k-1]$, $r(w'(i)) \leq r(w(i)) + \eta i$.*

Proof. By construction, for each $i \in [k-1]$ we have $r(w'(i+1)) - r(w'(i)) \leq r(w(i+1)) - r(w(i)) + \eta$. From this we entail that $r(w'(i)) - r(w'(1)) \leq r(w(i)) - r(w(1)) + (i-1) \cdot \eta$. Since also by construction $r(w'(1)) \leq r(w(1)) + \eta$, we get that $r(w'(i)) \leq r(w(i)) + i \cdot \eta$ as required. \square

With the lemma at hand, and by assumption that $\eta < r'(2H) \cdot H/k$ we obtain

$$w'(k) \leq r^{-1}(r(w(k)) + \eta \cdot k) \leq r^{-1}(r(H) + r'(2H) \cdot H) \leq 2H ,$$

using the concavity of r . Therefore we have $w' \in W_\eta$.

We now show that, compared to w , under w' the agent's effort cannot decrease. Then, we use this fact to bound the difference in the principal's utility between w and w' .

In order to prove our first claim, we will make use of the following lemma.

Lemma 3 (Grossman and Hart [18]). *Let w^1 and w^2 be contracts. Then,*

$$\sum_{i=1}^k (f_{\hat{e}(w^1)}(i) - f_{\hat{e}(w^2)}(i)) \cdot (u(w^1(i)) - u(w^2(i))) \geq 0.$$

Lemma 4. *The effort level the agent chooses can only increase from w to w' .*

Proof. First, notice that since the wages only increase, had the agent accepted the contract w , i.e., chose an effort level different than 0, she would also accept contract w' . So, let $e' = \hat{e}(w')$ and $e = \hat{e}(w)$ be the effort levels the agent chooses under contracts w' and w respectively.

If we apply Lemma 3 with w and w' , we obtain

$$\sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot (u(w'(i)) - u(w(i))) \geq 0 .$$

Assume for now that $u(w'(i)) - u(w(i))$ is monotone nondecreasing in i . Given this, we will show by contradiction that $e' \succ e$. So, for the sake of contradiction assume that $e' \prec e$. From the fact that f_e dominates $f_{e'}$, Eq. (1) implies that

$$\sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot (u(w'(i)) - u(w(i))) \leq 0 .$$

Thus,

$$\sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot u(w(i)) = \sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot u(w'(i)). \quad (2)$$

Therefore, by optimality of e and e' under contracts w and w' respectively, we obtain

$$\begin{aligned} \sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot u(w(i)) &\leq c(e') - c(e), \text{ and} \\ \sum_{i=1}^k (f_{e'}(i) - f_e(i)) \cdot u(w'(i)) &\geq c(e') - c(e). \end{aligned}$$

Combining Eq. (2) with the two inequalities above, we obtain that

$$U(w', e') = \sum_{i=1}^k f_{e'}(i) \cdot u(w'(i)) - c(e') = \sum_{i=1}^k f_e(i) \cdot u(w'(i)) - c(e) = U(w', e) .$$

This means that the agent is indifferent between effort levels e and e' under contract w' . Since, by Assumption 2, the agent chooses the highest effort in this case, we must have $e \prec e'$ — a contradiction.

Hence, in order to prove the lemma it suffices to prove that $u(w'(i)) - u(w(i))$ is monotone nondecreasing in i . This is equivalent to showing that $u(w'(i+1)) - u(w'(i)) \geq u(w(i+1)) - u(w(i))$. Denote $c = r(w'(i)) - r(w(i))$. By construction, $r(w'(i+1)) - r(w(i+1)) \geq c$. Consequently, since u is monotone nondecreasing, it suffices to show

$$u(r^{-1}(r(w(i+1)) + c)) - u(r^{-1}(r(w(i)) + c)) \geq u(w(i+1)) - u(w(i)).$$

Thus, the proof boils down to showing $c \mapsto u(r^{-1}(r(w(i+1)) + c)) - u(r^{-1}(r(w(i)) + c))$ is monotone nondecreasing in c . Taking the derivative with respect to c , we need to show

$$\frac{u'(r^{-1}(r(w(i+1)) + c))}{r'(r^{-1}(r(w(i+1)) + c))} - \frac{u'(r^{-1}(r(w(i)) + c))}{r'(r^{-1}(r(w(i)) + c))} \geq 0.$$

However, the above holds since the agent is BRA (Assumption 3), and as $r^{-1}(r(w(i+1)) + c) \geq r^{-1}(r(w(i)) + c)$ due to both r and r^{-1} being monotone-increasing. \square

We can now bound the loss the principal suffers when she offers w' instead of w .

Lemma 5. *It holds that $V(w) \leq V(w') + k\eta/r'(2H)$.*

Proof. Since we focus on scenarios where the optimal contract is monotone-smooth, we get that the net profit of the principal at the optimal contract, $\pi(i) - w(i)$, is nondecreasing in i . Furthermore, from Eq. (1), keeping w fixed, the principal only benefits from an increase of the agent's effort level. Denote $e = \hat{e}(w)$ and $e' = \hat{e}(w')$. From Lemma 4 we know that $e' \succ e$. We have,

$$\begin{aligned} V(w) &= \sum_{i=1}^k f_e(i) \cdot (\pi(i) - w(i)) \\ &\leq \sum_{i=1}^k f_{e'}(i) \cdot (\pi(i) - w(i)) \\ &= V(w') + \sum_{i=1}^k f_{e'}(i) \cdot (w'(i) - w(i)). \end{aligned} \quad (3)$$

Now by Lemma 2, $w'(i) \leq r^{-1}(r(w(i)) + \eta \cdot i)$ for all $i \in [k]$. Note that r^{-1} is convex. Then, as $\eta < r'(2H) \cdot H/k$, we obtain

$$\begin{aligned} w'(i) - w(i) &\leq r^{-1}(r(w(i)) + \eta \cdot i) - w(i) \\ &\leq \frac{\eta \cdot i}{r'(2H) \cdot H} \left(r^{-1}(r(w(i)) + r'(2H) \cdot H) - w(i) \right) \\ &\leq \frac{\eta \cdot i}{r'(2H) \cdot H} \left(r^{-1}(r(H) + r'(2H) \cdot H) - H \right) \\ &\leq \frac{\eta \cdot i}{r'(2H) \cdot H} (2H - H) \\ &\leq \frac{\eta \cdot k}{r'(2H)}, \end{aligned}$$

where the second and third inequalities are by the convexity of r^{-1} , and the fourth inequality is by the concavity of r . Combining the latter with Eq. (3), we get $V(w) \leq V(w') + \eta \cdot k/r'(2H)$. \square

4 Applications

In this section we highlight two cases that received attention in the past. For each of them, when Assumptions 1 to 3 hold, the optimal contract will be in the set W , and thus by learning an ε -optimal contract in W , we approximate the best contract the principle could have offered the agent had she known her utility function, effort levels and costs, and the distributions they induce over outcomes.

Two outcomes. Firstly, we focus on the case where there are only two outcomes and show that the optimal contract is in W .

Lemma 6. *When there are only two outcomes, the optimal contract is monotone-smooth and $2H$ -bounded.*

Proof. By [18], the optimal contract in the two outcome case is of the shape: $w(2) = w(1) + a(\pi(2) - \pi(1))$ for some $a \in [0, 1]$. By plugging this expression into the inequality in Definition 2, we get that the optimal contract is monotone-smooth. To see that the optimal contract is $2H$ -bounded, let e denote her chosen effort level. Since the principal's utility at the optimal contract is nonnegative, $f_e(1)w(1) + f_e(2)w(2) \leq f_e(1)\pi(1) + f_e(2)\pi(2) \leq H$, and in particular $w(1) \leq H$. Now, $w(2) = w(1) + a(\pi(2) - \pi(1)) \leq H + 1 \cdot H = 2H$. \square

Thus, by Theorem 3, applied to our discretized contract space, MEDIANE-LIMINATION finds an ϵ -optimal contract under Assumptions 1 to 3. Note, the FOSD assumption (Assumption 1) is standard in the literature, hence our result essentially requires only the bounded risk-averse assumption (Assumption 3).

Risk neutral agent. [9] studies a setting in the agent is risk-neutral, and the principal has only partial knowledge of the agent's action space. Had the principal known the complete action space of the agent, the optimal contract would have been linear. In this setting, the principal can derive the optimal *linear* contract with respect to only the known actions of the agent. [9] show that her profit from the actual action taken by the agent (which can be one that the principal is unaware of) can only be higher than the principal's expectations.

In the following lemma we show that when the agent is risk-neutral, the optimal theoretical contract is in W , and thus our algorithm learns a contract that approximates it.

Lemma 7. *If the agent is risk-neutral then the optimal contract is H -bounded and monotone-smooth.*

Proof. When the agent is risk neutral, the optimal contract is of the shape $w(i) = \pi(i) - \alpha$ for some constant $\alpha \in \mathbb{R}_+$ (Proposition 14.B.2 on page 482 in [27]). As $\pi(i+1) > \pi(i)$, this contract is monotone-smooth. And as $\pi(k) \leq H$ it is also H -bounded. \square

5 Conclusions

In this paper we studied the principal-agent problem when the principal has zero information about the the agent. We introduced the class of monotone-smooth contracts and showed that when the optimal contract is monotone-smooth and the agent is bounded risk-averse, then we can learn an approximately optimal contract. We complemented this result with a multi-armed bandit algorithm that finds an approximately optimal contract and we provided bounds on the number of samples it needs. Then, we applied our algorithm to two fundamental scenarios.

The first one is when the output space of the task is binary and the second one is when the agent is risk-neutral. Economic theory suggests that the optimal contract *is* monotone-smooth. Thus the net profit of the principal generated by the resulting contract approximates the optimal net profit she can achieve in general. To the best of our knowledge these are the first positive results even regarding approximately optimal contracts.

Several intriguing questions remain. It is interesting to understand whether the assumption of bounded risk-aversion is needed to guarantee learning of monotone-smooth contracts. The answer to this question is not obvious even when there are only two outcomes. Furthermore, we wish to find other conditions and assumptions that allow learning. In our model, we assume that the agents are identical. Can we learn ϵ -optimal contracts if there are many different types of agents? We conjecture that we *can* under the suitable assumptions, like the setting of [20]. On the other hand, we furthermore conjecture that there exist cases in which learning is *not* possible at all. Lower bounds, or even partial characterizations of such cases would be of great interest.

References

1. Agrawal, R.: The continuum-armed bandit problem. *SIAM journal on control and optimization* **33**(6), 1926–1951 (1995)
2. Alon, T., Dütting, P., Talgam-Cohen, I.: Contracts with private cost per unit-of-effort. In: *Proceedings of the 22nd ACM Conference on Economics and Computation*. pp. 52–69 (2021)
3. Arrow, K.J.: *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnsson Foundation, Helsinki (1965)
4. Auer, P., Ortner, R., Szepesvári, C.: Improved rates for the stochastic continuum-armed bandit problem. In: *International Conference on Computational Learning Theory*. pp. 454–468. Springer (2007)
5. Babaioff, M., Feldman, M., Nisan, N.: Mixed strategies in combinatorial agency. *Journal of Artificial Intelligence Research* **38**, 339–369 (2010)
6. Babaioff, M., Feldman, M., Nisan, N., Winter, E.: Combinatorial agency. *Journal of Economic Theory* **147**(3), 999–1034 (2012)
7. Bubeck, S., Stoltz, G., Yu, J.Y.: Lipschitz bandits without the lipschitz constant. In: *International Conference on Algorithmic Learning Theory*. pp. 144–158. Springer (2011)
8. Bughin, J., Manyika, J., Woetzel, J., Labaye, E., chairman Michael Chui, Lund, S., Cadena, A., Dobbs, R., George, K., Gupta, R., Hazan, E., Leke, A., Nyquist, S., Pinkus, G., Sankhe, S.: Independent work: Choice, necessity, and the gig economy. Tech. rep., McKinsey & Company 2016 (2016), <https://www.mckinsey.com/{~}/media/McKinsey/GlobalThemes/EmploymentandGrowth/IndependentworkChoiceneccessityandthegigeconomy/Independent-Work-Choice-necessity-and-the-gig-economy-Full-report.ashx>
9. Carroll, G.: Robustness and linear contracts. *American Economic Review* **105**(2), 536–563 (2015). <https://doi.org/10.1257/aer.20131159>
10. Castiglioni, M., Marchesi, A., Gatti, N.: Bayesian agency: Linear versus tractable contracts. *Artificial Intelligence* **307**, 103684 (2022)

11. Castiglioni, M., Marchesi, A., Gatti, N.: Designing menus of contracts efficiently: The power of randomization. arXiv preprint arXiv:2202.10966 (2022)
12. Conitzer, V., Garera, N.: Learning algorithms for online principal-agent problems (and selling goods online). In: Proceedings of the 23rd international conference on Machine learning. pp. 209–216. ACM (2006)
13. Dütting, P., Ezra, T., Feldman, M., Kesselheim, T.: Combinatorial contracts. In: 62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022. pp. 815–826. IEEE (2021). <https://doi.org/10.1109/FOCS52979.2021.00084>, <https://doi.org/10.1109/FOCS52979.2021.00084>
14. Dütting, P., Roughgarden, T., Talgam-Cohen, I.: Simple versus optimal contracts. In: Karlin, A., Immorlica, N., Johari, R. (eds.) Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019. pp. 369–387. ACM (2019). <https://doi.org/10.1145/3328526.3329591>, <https://doi.org/10.1145/3328526.3329591>
15. Dütting, P., Roughgarden, T., Talgam-Cohen, I.: The complexity of contracts. *SIAM J. Comput.* **50**(1), 211–254 (2021). <https://doi.org/10.1137/20M132153X>, <https://doi.org/10.1137/20M132153X>
16. Even-Dar, E., Mannor, S., Mansour, Y.: Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research* **7**(Jun), 1079–1105 (2006)
17. Gershkov, A., Perry, M.: Dynamic Contracts with Moral Hazard and Adverse Selection. *The Review of Economic Studies* **79**(1), 268–306 (jan 2012). <https://doi.org/10.1093/restud/rdr026>, <https://academic.oup.com/restud/article-lookup/doi/10.1093/restud/rdr026>
18. Grossman, S.J., Hart, O.D.: An Analysis of the Principal-Agent Problem. *Econometrica* **51**(1), 7–45 (1983). <https://doi.org/10.2307/1912246>
19. Guruganesh, G., Schneider, J., Wang, J.R.: Contracts under moral hazard and adverse selection. In: Proceedings of the 22nd ACM Conference on Economics and Computation. pp. 563–582 (2021)
20. Ho, C.J., Slivkins, A., Vaughan, J.W.: Adaptive contract design for crowd-sourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research* **55**, 317–359 (2016). <https://doi.org/10.1145/2600057.2602880>
21. Holmstrom, B.: Moral Hazard and Observability. *The Bell Journal of Economics* **10**(1), 74 (1979). <https://doi.org/10.2307/3003320>, <http://www.jstor.org/stable/3003320?origin=crossref>
22. Holmström, B.: Pay for performance and beyond. *American Economic Review* **107**(7), 1753–1777 (2017). <https://doi.org/10.1586/14737167.7.4.351>, <http://www.tandfonline.com/doi/full/10.1586/14737167.7.4.351>
23. Kleinberg, R., Slivkins, A., Upfal, E.: Bandits and experts in metric spaces. arXiv preprint arXiv:1312.1277 (2013)
24. Kleinberg, R.D.: Nearly tight bounds for the continuum-armed bandit problem. In: Advances in Neural Information Processing Systems. pp. 697–704 (2005)
25. Laffont, J.J., Martimort, D.: The theory of incentives: the principal-agent model. Princeton university press (2009)
26. Magureanu, S., Combes, R., Proutiere, A.: Lipschitz bandits: Regret lower bounds and optimal algorithms. arXiv preprint arXiv:1405.4758 (2014)
27. Mas-Colell, A., Whinston, M.D., Green, J.R.: Microeconomic theory. Oxford University Press (1995)

28. Pratt, J.W.: Risk Aversion in the Small and in the Large. *Econometrica* **32**(1/2), 122 (jan 1964). <https://doi.org/10.2307/1913738>, <https://www.jstor.org/stable/1913738?origin=crossref>
29. Robbins, H.: Some aspects of the sequential design of experiments. In: Herbert Robbins Selected Papers, pp. 169–177. Springer (1985)
30. Ross, S.A.: The Economic Theory of Agency : The Principal’s Problem. *American Economic Review* **63**(2), 134–139 (1973)
31. Sannikov, Y.: Contracts: The theory of dynamic principal–agent relationships and the continuous-time approach **1**, 89–124 (01 2011)
32. Sannikov, Y.: A continuous- time version of the principal: Agent problem. *The Review of Economic Studies* **75**(3), 957–984 (2008), <http://www.jstor.org/stable/20185061>
33. Williams, N.: On dynamic principal-agent problems in continuous time. Tech. rep., UCLA Department of Economics (2004)