Contextual Pandora's Box

Alexia Atsidakou[1*], **Constantine Caramanis**[1], **Evangelia Gergatsouli**[2*], **Orestis Papadigenopoulos**[3*], **Christos Tzamos**[2,4]

[1]University of Texas - Austin,
[2]University of Wisconsin - Madison,
[3] Columbia University,
[4] University of Athens

Abstract

PANDORA'S BOX is a fundamental stochastic optimization problem, where the decision-maker must find a good alternative, while minimizing the search cost of exploring the value of each alternative. In the original formulation, it is assumed that accurate distributions are given for the values of all the alternatives, while recent work studies the online variant of PANDORA'S BOX where the distributions are originally unknown. In this work, we study PANDORA'S BOX in the online setting, while incorporating context. At each round, we are presented with a number of alternatives each having a context, an exploration cost and an unknown value drawn from an unknown distribution that may change at every round. Our main result is a no-regret algorithm that performs comparably well against the optimal algorithm which knows all prior distributions exactly. Our algorithm works even in the bandit setting where the algorithm never learns the values of the alternatives that were not explored. The key technique that enables our result is a novel modification of the realizability condition in contextual bandits that connects a context to a sufficient statistic of each alternative's distribution (its *reservation value*) rather than its mean.

1 Introduction

PANDORA'S BOX is a fundamental stochastic optimization framework, which models the trade-off between exploring a set of alternatives and exploiting the already collected information, in environments where data acquisition comes at a cost. In the original formulation of the problem – introduced by Weitzman in (Weitzman 1979) – a decision-maker is presented with a set of alternatives (called "boxes") each containing an unknown value drawn independently from some known box-specific distribution. In addition, each box is associated with a known cost, namely, the price that needs to be paid in order to observe its realized value. At each step, the decision-maker can either open a box of her choice, paying the associated cost and observing its value, or stop and collect the minimum value contained in the already opened boxes. The objective is to minimize the sum of the smallest observed value plus the total cost incurred by opening the boxes.

The model captures a variety of different settings where the decision-maker needs to balance the value of the selected alternative and the effort devoted to find it. We include some examples below.

- Consider an online shopping environment where a search engine needs to present users with results on a product they want to buy. Visiting all potential e-shops that sell this product to find the cheapest option would be prohibitive in terms of time needed to present the search results to the user. The search engine needs to explore the different options only up to the extent that would make the marginal improvement in the best price found worthwhile.

- Consider a path planning service provider like Google Maps. Upon request, the provider must search its database for a good path to recommend to the user, but the higher the time spent searching the higher is the server cost. The provider must trade off computation cost with the quality of the result.

Rather surprisingly, despite the richness of the setting, Weitzman shows that the optimal policy for any instance of PANDORA'S BOX admits a particularly simple characterization: for each box, one can compute a *reservation value* as a function of its cost and value distribution. Then, the boxes are inspected in increasing order of these values, until a simple termination criterion is fulfilled. This characterization is revealing: obtaining an optimal algorithm for PANDORA'S BOX does not require complete knowledge of the distributions or the costs, yet only access to a *single statistic* for each box.

This raises the important question of how easy it is to learn a near-optimal search strategy, especially in environments where the distributions may not remain fixed across time but can change according to the characteristics of the instance at hand. In the case of online shopping, depending on the type of product we are searching, the e-shops have different product-specific distributions on the prices. We would be interested in a searching strategy that is not tied to a specific product, but is able to minimize the expected cost for any product we may be interested in. Similarly, in the case of path recommendations, the optimal search strategy may depend on the time of day or the day of the year.

Motivated by the above question, we extend the PAN-

DORA'S BOX model to a contextual online learning setting, where the learner faces a new instance of the problem at each round. At the beginning of each round, the learner observes a context and must choose a search strategy for opening boxes depending on the observation. While the context and the associated opening costs of the boxes of each round are observed, the learner has no access to the value distributions, which may be arbitrary in each round.

Realizability. In the above description, the context may be irrelevant to the realized values. Such an adversarial setting is impossible to solve as it is related to the problem of learning thresholds online. In fact, even in the offline version of the problem, the task would still be computationally hard as it corresponds to agnostic learning (see **??** for more details). This naturally raises the following question: *What are the least possible strong assumptions, under which the problem becomes tractable?*

One of the main contributions of this paper is identifying a minimal realizability assumption under which the problem remains tractable. This assumption is parallel to the realizability assumption in contextual bandits (see chapter 19 of (Lattimore and Szepesvári 2020)). We describe below how the difficulty of problem increases as we move from the strongest (1) to the weakest (4) assumptions possible:

1. **Contexts directly related to values**: in this case any learning algorithm is able to fit the contexts and predict exactly the realized values of the boxes. Such a setting is trivial yet unrealistic.

2. **Contexts directly related to distributions**: this is the case where there exists a learnable mapping from the context to the distribution of values. This is a more realistic setting, but it is still relatively constrained; it requires being able to perfectly determine the distribution family, which would need to be parametric.

3. **Contexts related to sufficient statistic**: in the more general case, instead of the whole value distributions, the contexts give us information about only a sufficient statistic of the problem. This is one of the main contribution of this work; we show that the problem remains tractable when the contexts give us information on the *reservation values* of the boxes. Observe that in this case, the value distributions on the boxes can be arbitrarily different at each round, as long as they "implement" the correct reservation value based on the context. This model naturally extends the standard realizability assumption made in bandit settings, according to which, the mean of the distributions is predictable from the context. In that case, the sufficient statistic needed in order to select good arms is, indeed, the mean reward of each arm (Lattimore and Szepesvári 2020).

4. **No assumptions**: in this case, as explained before, the problem becomes intractable (see **??**).

1.1 Our Contribution

We introduce a novel contextual online learning variant of the PANDORA'S BOX problem, namely the CONTEXTUAL PANDORA'S BOX, that captures the problem of learning near-optimal search strategies in a variety of settings.

Our main technical result shows that even when the sequence of contexts and distributions is adversarial, we can find a search strategy with sublinear average regret (compared to an optimal one) as long as no-regret algorithms exist for a much simpler online regression problem with a linear-quadratic loss functions.

Main Theorem (Informal). *Given an oracle that achieves expected regret $r(T)$ after T rounds for Linear-Quadratic Online Regression, there is an algorithm that obtains $O(\sqrt{Tr(T)})$ regret for the* CONTEXTUAL PANDORA'S BOX *problem.*

The main technical challenge in obtaining the result is that the class of search strategies can be very rich. Even restricting to greedy policies based on reservation values for each box, the cost of the policies is a non-convex function of the reservation values. We manage to overcome this issue by considering a "proxy" function for the expected cost of the search policy which bounds the difference from the optimal cost, based on a novel sensitivity analysis of the original Weitzman's algorithm. The proxy function has a simple linear-quadratic form and thus optimizing it reduces our setting to an instance of *linear-quadratic online regression*. This allows us to leverage existing methods for minimizing regret in online regression problems in a black box manner.

Using the above reduction, we design algorithms with sublinear regret guarantees for two different variants of our problem: the *full information*, where the decision-maker observes the realized values of all boxes at the end of each round, and the *bandit* version, where only the realized values of the opened boxes can be observed. We achieve both results by constructing oracles based on the *Follow the Regularized Leader* family of algorithms.

Beyond the results shown in this paper, an important conceptual contribution of our work is extending the traditional bandit model in the context of stochastic optimization. Instead of trying to learn simple decision rules, in stochastic optimization we are interested in learning complex algorithms tailored to a distribution. Our model can be extended to a variety of such problems beyond the PANDORA'S BOX setting: one concrete such example is the case of designing revenue optimal auctions for selling a single item to multiple buyers given distributional information about their values. Modeling these settings through an online contextual bandit framework allows obtaining results without knowledge of the prior distributions which may change based on the context. Our novel realizability assumption allows one to focus on predicting only the sufficient statistics required for running a specific algorithm. In the case of designing revenue optimal auctions, contexts may refer to the attributes of the item for sale and a sufficient statistic for the bidder value distributions are Myerson's reserve prices (Myerson 1981).

Due to space constraints, any omitted proofs have been moved to the Supplementary Material.

1.2 Related Work

We model our search problem using PANDORA'S BOX, which was first introduced by Weitzman in the Economics literature (Weitzman 1979). Since then, there has been a long

line of research studying PANDORA'S BOX and its variants e.g. where boxes can be selected without inspection (Doval 2018; Beyhaghi and Kleinberg 2019), there is correlation between the boxes (Chawla et al. 2020, 2023), the boxes have to be inspected in a specific order (Boodaghians et al. 2020) or boxes are inspected in an online manner (Esfandiari et al. 2019) or over T rounds (Gergatsouli and Tzamos 2022; Gatmiry et al. 2022). Some work is also done in the generalized setting where more information can be obtained for a price (Charikar et al. 2000; Gupta and Kumar 2001; Chen et al. 2015b,a). Finally a long line of research considers more complex combinatorial constraints like budget constraints (Goel, Guha, and Munagala 2006), packing constraints (Gupta and Nagarajan 2013), matroid constraints (Adamczyk, Sviridenko, and Ward 2016), maximizing a submodular function (Gupta, Nagarajan, and Singla 2016, 2017), an approach via Markov chains (Gupta et al. 2019) and various packing and covering constraints for both minimization and maximization problems (Singla 2018).

A more recent line of work, that is very closely related to our setting, studies PANDORA'S BOX and other stochastic optimization problems like min-sum set cover in online settings (Gergatsouli and Tzamos 2022; Fotakis et al. 2020). Similar to our work, these papers do not require specific knowledge of distributions and provide efficient algorithms with low regret. However, these settings are not contextual, therefore they can only capture much simpler practical applications. Moreover, they only obtain multiplicative regret guarantees as, there, the problems are NP-hard to solve exactly.

Our problem is closely related to contextual multi-armed bandits, where the contexts provide additional information on the quality of the actions at each round. In particular, in the case of *stochastic linear bandits* (Abe and Long 1999), the reward of each round is given by a (noisy) a linear function of the context drawn at each round. Optimistic algorithms proposed for this setting rely on maintaining a confidence ellipsoid for estimating the unknown vector (Dani, Hayes, and Kakade 2008; Rusmevichientong and Tsitsiklis 2010; Abbasi-yadkori, Pál, and Szepesvári 2011; Valko et al. 2014). On the other hand, in *adversarial linear bandits*, a context vector is adversarially selected at each round. The loss is characterized by the inner product of the context and the selected action of the round. Common approaches for this setting include variants of the multiplicative-weights algorithm (Hazan, Karnin, and Meka 2014; van der Hoeven, van Erven, and Kotłowski 2018), as well as, tools from online linear optimization (Blair 1985; Cesa-Bianchi and Lugosi 2006) such as follow-the-regularised-leader and mirror descent (see (Bubeck and Eldan 2015; Abernethy, Hazan, and Rakhlin 2008; Shalev-Shwartz and Singer 2007; Bubeck, Cohen, and Li 2018) and references therein).

We note that our model generalizes the contextual bandits setting, since any instance of contextual bandits can be reduced to CONTEXTUAL PANDORA'S BOX for box costs selected to be large enough. One work from the contextual bandits literature that is more closely related to ours is the recent work of Foster and Rakhlin (Foster and Rakhlin 2020). Similarly to our work, they provide a generic reduc-

tion from contextual multi-armed bandits to online regression, by showing that any oracle for online regression can be used to obtain a contextual bandits algorithm.

Our work also fits in the recent direction of learning algorithms from data and algorithm configuration, initiated by Gupta and Roughgarden (Gupta and Roughgarden 2017), and continued by (Balcan et al. 2017, 2018; Balcan, Dick, and Vitercik 2018; Kleinberg, Leyton-Brown, and Lucier 2017; Weisz, Gyorgy, and Szepesvari 2018; Alabi et al. 2019). Similar work was done before in self-improving algorithms by (Ailon et al. 2006; Clarkson, Mulzer, and Seshadhri 2010). A related branch of work initiated by Munoz and Vassilvitski (Medina and Vassilvitskii 2017) and Lykouris and Vassilvitski (Lykouris and Vassilvitskii 2018) combines Machine Learning predictions to improve algorithm design studying various online algorithm questions like ski rental (Purohit, Svitkina, and Kumar 2018; Gollapudi and Panigrahi 2019; Wang, Li, and Wang 2020), caching (Lykouris and Vassilvitskii 2018; Rohatgi 2020; Wei 2020), scheduling (Lattanzi et al. 2020; Mitzenmacher 2020), online primal-dual method (Bamas, Maggiori, and Svensson 2020), energy minimization (Bamas et al. 2020) and secretary problem/matching (Antoniadis et al. 2023). While most of these works assume that the predictions are given a priori, recent works in the area (Diakonikolas et al. 2021; Lavastida et al. 2021) also focus on the task of learning the predictions. Our work can also be seen as learning to predict for the reservation values of the boxes in CONTEXTUAL PANDORA'S BOX.

2   Problem Definition and Notation

We begin by describing the original PANDORA'S BOX formulation. Then, in Section 2.2 we describe our online extension, which solves a contextual instance of PANDORA'S BOX at each round.

2.1   Original Pandora's Box Formulation

In PANDORA'S BOX we are given a set $\mathcal{B}$ of n boxes each with cost c_i and value $v_i \sim \mathcal{D}_i$, where the distributions $\mathcal{D}_i$ and the costs c_i for each box are known and the distributions are independent. The goal is to adaptively choose a box of small cost while spending as little as possible in opening costs. When opening box i, the algorithm pays c_i and observes the value $v_i \sim \mathcal{D}_i$ instantiated inside the box. The formal definition follows.

Definition 1 (PANDORA'S BOX cost). *Let $\mathcal{P}$ and c_i be the set of boxes opened and the cost of the box selected, respectively. The cost of the algorithm is*

$$\mathbb{E}_{\mathcal{P}, v_i \sim \mathcal{D}_i \forall i \in \mathcal{B}} \left[\min_{i \in \mathcal{P}} v_i + \sum_{i \in \mathcal{P}} c_i \right] \qquad (1)$$

where the expectation is taken over the distributions of the values v_i and the (potentially random) choice of $\mathcal{P}$ by the algorithm.

Observe that an adaptive algorithm for this problem has to decide on: (a) a (potentially adaptive) order according to

Algorithm 1: Weitzman's algorithm.

Input: n boxes, costs c_i, reservation values σ_i for $i \in \mathcal{B}$

1: $\pi \leftarrow$ sort boxes by increasing σ_i
2: $v_{\min} \leftarrow \infty$
3: **for** every box π_i **do**
4:     Pay c_i and open box π_i and observe v_i
5:     $v_{\min} \leftarrow \min(v_{\min}, v_i)$
6:     **if** $v_{\min} < \sigma_{\pi_{i+1}}$ **then**
7:         Stop and collect $v_{\min}$
8:     **end if**
9: **end for**

which it examines the boxes and, (b) a stopping rule. Weitzman's algorithm, first introduced in (Weitzman 1979), and formally presented in Algorithm 1, gives a solution to PANDORA'S BOX. The algorithm uses the order induced by the *reservation values* to open the boxes.

We denote by $\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}; \boldsymbol{c})$ the expected cost of running Weitzman's algorithm using reservation values $\boldsymbol{\sigma}$ on an instance with distribution $\mathcal{D}$ and costs $\boldsymbol{c}$. Weitzman showed that this algorithm achieves the optimal expected cost of Eq. (1) for the following selection of reservation values:

Theorem 2.1 (Weitzman (Weitzman 1979)). *Weitzman's algorithm is optimal for* PANDORA'S BOX *when run with reservation values* $\boldsymbol{\sigma}^*$ *that satisfy* $\mathbb{E}_{v_i \sim \mathcal{D}_i}[\text{ReLU}(\sigma_i^* - v)] = c_i$ *for every box* $i \in \mathcal{B}$, *where* $\text{ReLU}(x) = \max\{x, 0\}$.

2.2 Online Contextual Pandora's Box

We now describe an online contextual extension of PANDORA'S BOX. In CONTEXTUAL PANDORA'S BOX there is a set $\mathcal{B}$ of n boxes with costs $\boldsymbol{c} = (c_1, \ldots, c_n)$[1]. At each round $t \in [T]$:

1. An (unknown) product distribution $\mathcal{D}_t = (\mathcal{D}_{t,1}, \ldots, \mathcal{D}_{t,n})$ is chosen and for every box i, a value $v_{t,i} \sim \mathcal{D}_{t,i}$ is independently realized.

2. A vector of contexts $\boldsymbol{x}_t = (\boldsymbol{x}_{t,1}, \ldots \boldsymbol{x}_{t,n})$ is given to the learner, where $\boldsymbol{x}_{t,i} \in \mathbb{R}^d$ and $\|\boldsymbol{x}_{t,i}\|_2 \leq 1$ for each box $i \in \mathcal{B}$.

3. The learner decides on a (potentially adaptive) algorithm $\mathcal{A}$ based on past observations.

4. The learner opens the boxes according to $\mathcal{A}$ and chooses the lowest value found.

5. At the end of the round, the learner observes all the realized values of all boxes (*full-information* model), or observes only the values of boxes opened in the round (*bandit* model).

Assumption 2.2 (Realizability). *There exist vectors* $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_n^* \in \mathbb{R}^d$ *and a function* h, *such that for every time* $t \in [T]$ *and every box* $i \in \mathcal{B}$, *the optimal reservation value* $\sigma_{t,i}^*$ *for the distribution* $\mathcal{D}_{t,i}$ *is equal to* $h(\boldsymbol{w}_i^*, \boldsymbol{x}_{t,i})$, *i.e.*

$$\mathbb{E}_{v_{t,i} \sim \mathcal{D}_{t,i}}[\text{ReLU}(h(\boldsymbol{w}_i^*, \boldsymbol{x}_{t,i}) - v_{t,i})] = c_i.$$

[1]Every result in the paper holds even if the costs change for each $t \in [T]$ and can be adversarially selected.

The goal is to achieve low expected regret over T rounds compared to an optimal algorithm that has prior knowledge of the vectors $\boldsymbol{w}_i^*$ and, thus, can compute the exact reservation values of the boxes in each round and run Weitzman's optimal policy. The regret of an algorithm is defined as the difference between the cumulative PANDORA'S BOX cost achieved by the algorithm compared to the cumulative cost achieved by running Weitzman's optimal policy at every round. That is:

Definition 2 (Expected Regret). *The expected regret of an algorithm* $\mathcal{A}$ *that opens boxes* $\mathcal{P}_t$ *at round* t *over a time horizon* T *is*

$$\text{Regret}(\mathcal{A}, T) =$$
$$\mathbb{E}\left[\sum_{t=1}^{T}\left(\min_{i \in \mathcal{P}_t} v_{t,i} + \sum_{i \in \mathcal{P}_t} c_i - \text{WEITZ}_{\boldsymbol{D}_t}(h(\boldsymbol{w}^*, \boldsymbol{x}_t); \boldsymbol{c})\right)\right].$$
(2)

The expectation is taken over the randomness of the algorithm, the contexts $\boldsymbol{x}_t$, *distributions* $\mathcal{D}_t$ *and realized values* $\boldsymbol{v}_t \sim \mathcal{D}_t$, *over all rounds* $t \in [T]$.

Remark. If the learner uses Weitzman's algorithm at every time step t, with reservation values $\sigma_{t,i} = h(\boldsymbol{w}_{t,i}, \boldsymbol{x}_{t,i})$ for some chosen parameter $\boldsymbol{w}_{t,i} \in \mathbb{R}^d$ for every box $i \in \mathcal{B}$, the regret can be written as

$$\text{Regret}(\mathcal{A}, T) =$$
$$\mathbb{E}\Big[\sum_{t=1}^{T}\Big(\text{WEITZ}_{\boldsymbol{D}_t}(h(\boldsymbol{w}_t, \boldsymbol{x}_t); \boldsymbol{c})$$
$$- \text{WEITZ}_{\boldsymbol{D}_t}(h(\boldsymbol{w}^*, \boldsymbol{x}_t); \boldsymbol{c})\Big)\Big],$$

where $\boldsymbol{w}_t = (\boldsymbol{w}_{t,1}, \ldots, \boldsymbol{w}_{t,n})$.

3 Reduction to Online Regression

In this section we give a reduction from CONTEXTUAL PANDORA'S BOX problem to an instance of online regression, while maintaining the regret guarantees given by online regression. We begin by formally defining the regression problem:

Definition 3 (Linear-Quadratic Online Regression). *Online regression with loss* ℓ *is defined as follows: at every round* t, *the learner first chooses a prediction* $\boldsymbol{w}_t$, *then an adversary chooses an input-output pair* $(\boldsymbol{x}_t, y_t)$ *and the learner incurs loss* $\ell(h(\boldsymbol{w}_t, \boldsymbol{x}_t) - y_t)$.

*In the **costly feedback** setting, the learner may observe the input-output pair at the end of round* t, *if they choose to pay an information acquisition cost* a. *The **full information** setting, corresponds to the case where* $a = 0$, *in which case the input-output pair is always visible. The regret of the learner after* T *rounds, when the learner has acquired information* k *times, is equal to*

$$\sum_{t=1}^{T} \ell(h(\boldsymbol{w}_t, \boldsymbol{x}_t) - y_t) - \min_{\boldsymbol{w}} \sum_{t=1}^{T} \ell(h(\boldsymbol{w}_t, \boldsymbol{x}_t) - y_t) + ak.$$

Algorithm 2: $\mathcal{CPB}$: Contextual Pandora's Box

Input: Oracle $\mathcal{O}$ for Linear-Quadratic Online Regression.

1: For every box $i \in \mathcal{B}$, instantiate a copy $\mathcal{O}_i$ of the oracle with linear-quadratic loss H_{c_i}
2: **for** each round $t \in T$ **do**
3:    #Prediction Phase
4:    Call oracle $\mathcal{O}_i$ to get a prediction $\boldsymbol{w}_{t,i}$ for each box i
5:    Obtain context $\boldsymbol{x}_t^i$ for each box $i \in \mathcal{B}$
6:    Run Weitzman's algorithm 1 with reservation values $\sigma_{t,i} = h(\boldsymbol{w}_{t,i}, \boldsymbol{x}_{t,i})$ for each box $i \in \mathcal{B}$

7:    #Update Phase
8:    **for** every box $i \in \mathcal{B}$ **do**
9:      **if** the oracle $\mathcal{O}_i$ requests an input-output pair at this round **then**
10:       Observe value $v_{t,i}$ and give the input-output pair $(\boldsymbol{x}_{t,i}, v_{t,i})$
11:      **end if**
12:    **end for**
13: **end for**

Linear-Quadratic Online Regression *is the special case of online regression where the loss function $\ell(z)$ is chosen to be a linear-quadratic function of the form*

$$H_c(z) = \frac{1}{2}\text{ReLU}(z)^2 - cz \qquad (3)$$

for some parameter $c > 0$.

The reduction presented in Algorithm 2 shows how we can use an oracle for linear-quadratic online regression, to obtain an algorithm for CONTEXTUAL PANDORA'S BOX problem. We show that our algorithm achieves $O(\sqrt{Tr(T)})$ regret when the given oracle has a regret guarantee of $r(T)$.

Theorem 3.1. *Given an oracle that achieves expected regret $r(T)$ for Linear-Quadratic Online Regression, Algorithm 2 achieves $2n\sqrt{Tr(T)}$ regret for the CONTEXTUAL PANDORA'S BOX problem. In particular, if the regret $r(T)$ is sublinear in T, Algorithm 2 achieves sublinear regret.*

Our algorithm works by maintaining a regression oracle for each box, and using it at each round to obtain a prediction on $\boldsymbol{w}_{t,i}$. Specifically, in the prediction phase of the round the algorithm obtains a prediction and then uses the context $\boldsymbol{x}_{i,t}$ to calculate an estimated reservation value for each box. Then, based on the estimated reservation values, it uses Weitzman's algorithm 1 to decide which boxes to open. Finally, it accumulates the PANDORA'S BOX cost acquired by Weitzman's play at this round and the cost of any extra boxes opened by the oracle in the update phase (in the bandit setting). The update step is used to model the full-information setting (where the value of each box is always revealed at the end of the round) vs the costly feedback setting (where the value inside each box is only revealed if we paid the opening cost).

In the rest of this section we outline the proof of Theorem 3.1. The proof is based on obtaining robustness guar-

antees for Weitzman's algorithm when it is run with estimates instead of the true reservation values. In this case, we show that the cost incurred by Weitzman's algorithm is proportional to the error of the *approximate costs* of the boxes (Definition 4). This analysis is found on Section 3.1. Then, in Section 3.2 we exploit the form of the Linear-Quadratic loss functions to connect the robustness result with the regret of the Linear-Quadratic Online Regression problem and conclude our main Theorem 3.1 of this section.

An empirical evaluation of Algorithm 2 can be found in **??**.

3.1 Weitzman's Robustness

We provide guarantees on Weitzman's algorithm 1 performance when instead of the optimal reservation values $\boldsymbol{\sigma}^*$ of the boxes, the algorithm uses estimates $\boldsymbol{\sigma} \neq \boldsymbol{\sigma}^*$. We first define the following:

Definition 4 (Approximate Cost)**.** *Given a distribution $\mathcal{D}$ such that $v \sim \mathcal{D}$ and a value σ, the approximate cost with respect to σ and $\mathcal{D}$ is defined as:*

$$c_{\mathcal{D}}(\sigma) = \mathbb{E}_{v \sim \mathcal{D}}\left[\text{ReLU}(\sigma - v)\right]. \qquad (4)$$

Moreover, given n boxes with estimated reservation values $\boldsymbol{\sigma}$ and distributions $\boldsymbol{\mathcal{D}}$ we denote the vector of approximate costs as $\boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma}) = (c_{\mathcal{D}_1}(\sigma_1), \ldots, c_{\mathcal{D}_n}(\sigma_n))$.

Remark. Observe that, in the PANDORA'S BOX setting, if box i has value distribution $\mathcal{D}_i$ opening cost c_i and optimal reservation value σ_i^*, then, by definition, the quantity $c_{\mathcal{D}_i}(\sigma_i^*)$ corresponds to the true cost, c_i, of the box. This also holds for the vector of approximate costs, i.e. $\boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma}^*) = \boldsymbol{c}$.

We now state our robustness guarantee for Weitzman's algorithm. In particular, we show that the extra cost incurred due to the absence of initial knowledge of vectors $\boldsymbol{w}_i^*$ is proportional to the error of the approximate costs the boxes, as follows:

Proposition 3.2. *For a PANDORA'S BOX instance with n boxes with distributions $\boldsymbol{\mathcal{D}}$, costs $\boldsymbol{c}$ and corresponding optimal reservation values $\boldsymbol{\sigma}^*$ so that $\boldsymbol{c} = \boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma}^*)$, Weitzman's Algorithm 1, run with reservation values $\boldsymbol{\sigma}$ incurs cost at most*

$$\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}; \boldsymbol{c}) \leq \text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}^*; \boldsymbol{c}) + \|\boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma}) - \boldsymbol{c}\|_1.$$

Before showing Proposition 3.2, we prove the following lemma, that connects the optimal PANDORA'S BOX cost of an instance with optimal reservation values $\boldsymbol{\sigma}^*$ to the optimal cost of the instance with optimal reservation values $\boldsymbol{\sigma}$.

Lemma 3.3. *Let $\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}^*; \boldsymbol{c})$ and $\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}; \boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma}))$ be the optimal PANDORA'S BOX costs corresponding to instances with optimal reservation values $\boldsymbol{\sigma}^*$ and $\boldsymbol{\sigma}$ respectively. Then*

$$\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}^*; \boldsymbol{c}) \geq$$
$$\text{WEITZ}_{\mathcal{D}}(\boldsymbol{\sigma}; \boldsymbol{c}_{\mathcal{D}}(\boldsymbol{\sigma})) - \sum_{i \in \mathcal{B}} \text{ReLU}(c_{\mathcal{D}_i}(\sigma_i) - c_i).$$

The proof of the above Lemma, together with that of Proposition 3.2, are deferred to **??**.

3.2 Proof of Theorem 3.1

Moving on to show our main theorem, we connect the robustness Proposition 3.2 with the performance guarantee of the Linear-Quadratic Online Regression problem. The robustness guarantee of Weitzman's algorithm is expressed in terms of the error of the approximate costs of the boxes, while the regret of the Online Regression problem is measured in terms of the cumulative difference of the linear-quadratic loss functions $H_c(\cdot)$. Thus, we begin with the following lemma:

Lemma 3.4. *For any distribution $\mathcal{D}$ with $c_{\mathcal{D}}(\sigma^*) = c$, it holds that*

$$\mathbb{E}_{v \sim \mathcal{D}}\left[H_c(\sigma - v) - H_c(\sigma^* - v)\right] \geq \frac{1}{2}(c_{\mathcal{D}}(\sigma) - c_{\mathcal{D}}(\sigma^*))^2$$

The proof of the lemma is deferred to the Appendix.

Proof of Theorem 3.1. Recall that at every step $t \in [T]$, Algorithm 2 runs Weitzman's algorithm as a subroutine, using an estimate $\boldsymbol{\sigma}_t$ for the optimal reservation values of the round, $\boldsymbol{\sigma}_t^*$. From the robustness analysis of Weitzman's algorithm we obtain that the regret of Algorithm 2 can be bounded as follows:

$$\text{Regret}(\mathcal{CPB}, T)$$
$$= \mathbb{E}\left[\sum_{t \in [T]} \text{WEITZ}_{\boldsymbol{\mathcal{D}}_t}(\boldsymbol{\sigma}_t; \boldsymbol{c}_t) - \text{WEITZ}_{\boldsymbol{\mathcal{D}}_t}(\boldsymbol{\sigma}_t^*; \boldsymbol{c}_t)\right]$$
$$\leq \sum_{t \in [T], i \in \mathcal{B}} |c_{\mathcal{D}_{t,i}}(\sigma_{t,i}) - c_{\mathcal{D}_{t,i}}(\sigma_{t,i}^*)|$$
$$\leq \sqrt{nT}\sqrt{\sum_{t \in [T], i \in \mathcal{B}}(c_{\mathcal{D}_{t,i}}(\sigma_{t,i}) - c_{\mathcal{D}_{t,i}}(\sigma_{t,i}^*))^2},$$

where the first inequality follows by Proposition 3.2 and for the last inequality we used that for any k-dimensional vector z we have that $\|z\|_1 \leq \sqrt{k}\|z\|_2$ and the fact that the above sum over $T, \mathcal{B}$ is equivalent to ℓ_1 norm on nT dimensions. Moreover, we have that

$$\sum_{t \in [T], i \in \mathcal{B}}(c_{\mathcal{D}_{t,i}}(\sigma_{t,i}) - c_{\mathcal{D}_{t,i}}(\sigma_{t,i}^*))^2$$
$$\leq 2\,\mathbb{E}\left[\sum_{t \in [T], i \in \mathcal{B}}\left(H_{c_{t,i}}(\sigma_{t,i} - v_{t,i}) - H_{c_{t,i}}(\sigma_{t,i}^* - v_{t,i})\right)\right]$$
$$\leq 2 \cdot n \cdot r(T),$$

where for the first inequality we used Lemma 3.4, and then the guarantee of the oracle. Thus, we conclude that the total expected regret is at most $2n\sqrt{Tr(T)}$. □

4 Linear Contextual Pandora's Box

Using the reduction we developed in Section 3, we design efficient no-regret algorithms for CONTEXTUAL PANDORA'S BOX in the case where the mapping from contexts to reservation values is linear. That is, we assume that $h(\boldsymbol{w}, \boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$.

4.1 Full Information Setting

In this section we study the full-information version of the CONTEXTUAL PANDORA'S BOX problem, where the algorithm observes the realized values of all boxes at the end of each round, irrespectively of which boxes were opened. Initially we show that there exists an online regression oracle, that achieves sublinear regret for the full information version of the Linear-Quadratic Online Regression problem. Then, in Theorem 4.2 we combine our reduction of Theorem 3.1 with the online regression oracle guarantee, to conclude that Algorithm 2 using this oracle is no-regret for CONTEXTUAL PANDORA'S BOX. The lemma and the theorem follow.

Lemma 4.1. *When $h(\boldsymbol{w}, \boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$, $\|\boldsymbol{w}\|_2 \leq M$ and $\|\boldsymbol{x}\|_2 \leq 1$, there exists an oracle for Online Regression with Linear-Quadratic loss H_c under **full information** that achieves regret at most $\max\{M, c\}\sqrt{2MT}$.*

To show Lemma 4.1, we view Linear-Quadratic Online Regression as an instance of Online Convex Optimization and apply the *Follow The Regularized Leader* (FTRL) family of algorithms to obtain the regret guarantees.

Theorem 4.2. *In the full information setting, using the oracle of Lemma 4.1, Algorithm 2 for CONTEXTUAL PANDORA'S BOX achieves a regret of*

$$\text{Regret}(\mathcal{CPB}, T) \leq 3n\sqrt{\max\{M, c_{\max}\}}M^{1/4}T^{\frac{3}{4}},$$

assuming that for all times t and boxes $i \in \mathcal{B}$, $\|\boldsymbol{w}_{t,i}^*\|_2 \leq M$ and $c_{\max} = \max_{i \in \mathcal{B}} c_i$.

The process of using FTRL as an oracle is described in detail in **??**, alongside the proofs of Lemma 4.1 and Theorem 4.2.

4.2 Bandit Setting

We move on to extend the results of the previous section to the bandit setting, and show how to obtain a no-regret algorithm for this setting by designing a regression oracle with costly feedback. In this case, the oracle $\mathcal{O}_i$ of Algorithm 2 of each box $i \in \mathcal{B}$ does not necessarily receive information on the value of the box after each round. However, in each round it chooses whether to obtain the information for the box by paying the opening cost c.

We initially show that we can use any regression oracle given for the full-information setting, in the costly feedback setting without losing much in terms of regret guarantees. This is formalized in the following theorem.

Lemma 4.3. *Given an oracle that achieves expected regret $r(T)$ for Online Regression with Linear-Quadratic loss H_c under **full information**, Algorithm 3 is an oracle for Linear-Quadratic Online regression **with costly feedback**, that achieves regret at most $kr(T/k) + cT/k$.*

Algorithm 3 obtains an oracle with costly feedback from a full information oracle. It achieves this by splitting the time interval $[T]$ in intervals of size k, and choosing a uniformly random time per interval to acquire the costly information about the input-output pair. The proof of Lemma 4.3 is included in the Appendix.

Algorithm 3: Costly Feedback oracle from Full Information

Input: Parameter k, full information oracle $\mathcal{O}$

1: Split the times $[T]$ into T/k intervals $\mathcal{I}_1 \ldots, \mathcal{I}_{T/k}$
2: **for** Every interval $\mathcal{I}_\tau$ **do**
3:     Pick a t_p uniformly at random from $\mathcal{I}_\tau$
    #Prediction Phase
4:     Call $\mathcal{O}$ to get a vector $\boldsymbol{w}_\tau$.
5:     For each $t \in \mathcal{I}_\tau$ predict $\boldsymbol{w}_\tau$

    #Update Phase
6:     Obtain feedback for time $t_p \in \mathcal{I}_\tau$ and give input-output pair $(\boldsymbol{x}_{t_p}, y_{t_p})$ to $\mathcal{O}$.
7: **end for**

Given that we can convert an oracle for full-information to one with costly feedback using Algorithm 3, we can now present the main theorem of this section (see **??** for the proof):

Theorem 4.4. *In the bandit setting, using the oracle of Lemma 4.3 together with the oracle of Lemma 4.1, Algorithm 2 for* CONTEXTUAL PANDORA'S BOX *achieves a regret of*

$$\text{Regret}(\mathcal{CPB}, T) \leq 2n(2c_{\max}M \max\{M, c_{\max}\}^2)^{1/6}T^{5/6},$$

assuming that for all times t and boxes $i \in \mathcal{B}$, $\|\boldsymbol{w}_{t,i}^*\|_2 \leq M$ and $c_{\max} = \max_{i \in \mathcal{B}} c_i$.

Conclusion and Further Directions

We introduce and study an extension of the PANDORA'S BOX model to an online contextual regime, in which the decision-maker faces a different instance of the problem at each round. We identify the minimally restrictive assumptions that need to be imposed for the problem to become tractable, both computationally and information-theoretically. In particular, we formulate a natural realizability assumption – parallel to the one used widely in contextual bandits – which enables leveraging contextual information to recover a sufficient statistic for the instance of each round. Via a reduction to Linear-Quadratic Online Regression, we are able to provide a no-regret algorithm for the problem assuming either full or bandit feedback on the realized values. We believe that the framework we develop in this work could be extended to other stochastic optimization problems beyond PANDORA'S BOX. As a example, an interesting future direction would be its application to the case of revenue optimal online auctions; there Myerson's reserve prices can serve as a natural sufficient statistic (similarly to reservation prices) for obtaining optimality. In addition to applying our framework to different stochastic optimization problems, an interesting future direction would be the empirical evaluation of our methods on real data. Finally, we leave as an open question the derivation of non-trivial regret lower bounds for the considered problem. We discuss lower bounds guarantees for related settings and their implications to our problem in the Appendix.

References

Abbasi-yadkori, Y.; Pál, D.; and Szepesvári, C. 2011. Improved Algorithms for Linear Stochastic Bandits. In Shawe-Taylor, J.; Zemel, R.; Bartlett, P.; Pereira, F.; and Weinberger, K., eds., *NeurIPS*, volume 24. Curran Associates, Inc.

Abe, N.; and Long, P. M. 1999. Associative Reinforcement Learning using Linear Probabilistic Concepts.

Abernethy, J. D.; Hazan, E.; and Rakhlin, A. 2008. Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization. In *COLT*.

Adamczyk, M.; Sviridenko, M.; and Ward, J. 2016. Submodular Stochastic Probing on Matroids. *Math. Oper. Res.*, 41(3): 1022–1038.

Ailon, N.; Chazelle, B.; Comandur, S.; and Liu, D. 2006. Self-improving algorithms. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*, 261–270.

Alabi, D.; Kalai, A. T.; Ligett, K.; Musco, C.; Tzamos, C.; and Vitercik, E. 2019. Learning to Prune: Speeding up Repeated Computations. In Beygelzimer, A.; and Hsu, D., eds., *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, 30–33. PMLR.

Antoniadis, A.; Gouleakis, T.; Kleer, P.; and Kolev, P. 2023. Secretary and online matching problems with machine learned advice. *Discret. Optim.*, 48(Part 2): 100778.

Balcan, M.; Dick, T.; Sandholm, T.; and Vitercik, E. 2018. Learning to Branch. In *Proceedings of the 35th ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, 353–362.

Balcan, M.; Dick, T.; and Vitercik, E. 2018. Dispersion for Data-Driven Algorithm Design, Online Learning, and Private Optimization. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, 603–614.

Balcan, M.; Nagarajan, V.; Vitercik, E.; and White, C. 2017. Learning-Theoretic Foundations of Algorithm Configuration for Combinatorial Partitioning Problems. In *Proceedings of the 30thCOLT, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, 213–274.

Bamas, É.; Maggiori, A.; Rohwedder, L.; and Svensson, O. 2020. Learning Augmented Energy Minimization via Speed Scaling. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Bamas, É.; Maggiori, A.; and Svensson, O. 2020. The Primal-Dual method for Learning Augmented Algorithms. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Beyhaghi, H.; and Kleinberg, R. 2019. Pandora's Problem with Nonobligatory Inspection. In Karlin, A.; Immorlica,

N.; and Johari, R., eds., *Proceedings of the 2019 ACM Conference on Economics and Computation, EC 2019, Phoenix, AZ, USA, June 24-28, 2019*, 131–132. ACM.

Blair, C. E. 1985. Problem Complexity and Method Efficiency in Optimization (A. S. Nemirovsky and D. B. Yudin). *Siam Review*, 27: 264–265.

Boodaghians, S.; Fusco, F.; Lazos, P.; and Leonardi, S. 2020. Pandora's Box Problem with Order Constraints. In Biró, P.; Hartline, J. D.; Ostrovsky, M.; and Procaccia, A. D., eds., *EC '20: The 21st ACM Conference on Economics and Computation, Virtual Event, Hungary, July 13-17, 2020*, 439–458. ACM.

Bubeck, S.; Cohen, M. B.; and Li, Y. 2018. Sparsity, variance and curvature in multi-armed bandits. *ArXiv*, abs/1711.01037.

Bubeck, S.; and Eldan, R. 2015. The entropic barrier: a simple and optimal universal self-concordant barrier. In Grünwald, P.; Hazan, E.; and Kale, S., eds., *Proceedings of The 28thCOLT*, volume 40 of *Proceedings of Machine Learning Research*, 279–279. Paris, France: PMLR.

Cesa-Bianchi, N.; and Lugosi, G. 2006. *Prediction, Learning, and Games*. ISBN 978-0-521-84108-5.

Charikar, M.; Fagin, R.; Guruswami, V.; Kleinberg, J. M.; Raghavan, P.; and Sahai, A. 2000. Query strategies for priced information (extended abstract). In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, 582–591.

Chawla, S.; Gergatsouli, E.; McMahan, J.; and Tzamos, C. 2023. Approximating Pandora's Box with Correlations. In Megow, N.; and Smith, A. D., eds., *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2023, September 11-13, 2023, Atlanta, Georgia, USA*, volume 275 of *LIPIcs*, 26:1–26:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Chawla, S.; Gergatsouli, E.; Teng, Y.; Tzamos, C.; and Zhang, R. 2020. Pandora's Box with Correlations: Learning and Approximation. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, 1214–1225. IEEE.

Chen, Y.; Hassani, S. H.; Karbasi, A.; and Krause, A. 2015a. Sequential Information Maximization: When is Greedy Near-optimal? In *Proceedings of The 28thCOLT, COLT 2015, Paris, France, July 3-6, 2015*, 338–363.

Chen, Y.; Javdani, S.; Karbasi, A.; Bagnell, J. A.; Srinivasa, S. S.; and Krause, A. 2015b. Submodular Surrogates for Value of Information. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, 3511–3518.

Clarkson, K. L.; Mulzer, W.; and Seshadhri, C. 2010. Self-improving Algorithms for Convex Hulls. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, 1546–1565.

Dani, V.; Hayes, T. P.; and Kakade, S. M. 2008. Stochastic Linear Optimization under Bandit Feedback. In *COLT*.

Diakonikolas, I.; Kontonis, V.; Tzamos, C.; Vakilian, A.; and Zarifis, N. 2021. Learning Online Algorithms with Distributional Advice. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, 2687–2696. PMLR.

Doval, L. 2018. Whether or not to open Pandora's box. *J. Econ. Theory*, 175: 127–158.

Esfandiari, H.; Hajiaghayi, M. T.; Lucier, B.; and Mitzenmacher, M. 2019. Online Pandora's Boxes and Bandits. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 1885–1892. AAAI Press.

Foster, D. J.; and Rakhlin, A. 2020. Beyond UCB: Optimal and Efficient Contextual Bandits with Regression Oracles. In *Proceedings of the 37th ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, 3199–3210. PMLR.

Fotakis, D.; Lianeas, T.; Piliouras, G.; and Skoulakis, S. 2020. Efficient Online Learning of Optimal Rankings: Dimensionality Reduction via Gradient Descent. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Gatmiry, K.; Kesselheim, T.; Singla, S.; and Wang, Y. 2022. Bandit Algorithms for Prophet Inequality and Pandora's Box. *CoRR*, abs/2211.08586.

Gergatsouli, E.; and Tzamos, C. 2022. Online Learning for Min Sum Set Cover and Pandora's Box. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvári, C.; Niu, G.; and Sabato, S., eds., *ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, 7382–7403. PMLR.

Goel, A.; Guha, S.; and Munagala, K. 2006. Asking the right questions: model-driven optimization using probes. In *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 26-28, 2006, Chicago, Illinois, USA*, 203–212.

Gollapudi, S.; and Panigrahi, D. 2019. Online Algorithms for Rent-Or-Buy with Expert Advice. In *Proceedings of the 36th ICML 2019, 9-15 June 2019, Long Beach, California, USA*, 2319–2327.

Gupta, A.; Jiang, H.; Scully, Z.; and Singla, S. 2019. The Markovian Price of Information. In *Integer Programming and Combinatorial Optimization - 20th International Conference, IPCO 2019, Ann Arbor, MI, USA, May 22-24, 2019, Proceedings*, 233–246.

Gupta, A.; and Kumar, A. 2001. Sorting and Selection with Structured Costs. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, 416–425.

Gupta, A.; and Nagarajan, V. 2013. A Stochastic Probing Problem with Applications. In *IPCO 2013, Valparaíso, Chile, March 18-20, 2013. Proceedings*, 205–216.

Gupta, A.; Nagarajan, V.; and Singla, S. 2016. Algorithms and Adaptivity Gaps for Stochastic Probing. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, 1731–1747.

Gupta, A.; Nagarajan, V.; and Singla, S. 2017. Adaptivity Gaps for Stochastic Probing: Submodular and XOS Functions. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, 1688–1702.

Gupta, R.; and Roughgarden, T. 2017. A PAC Approach to Application-Specific Algorithm Selection. *SIAM J. Comput.*, 46(3): 992–1017.

Hazan, E.; Karnin, Z.; and Meka, R. 2014. Volumetric Spanners: an Efficient Exploration Basis for Learning. In Balcan, M. F.; Feldman, V.; and Szepesvári, C., eds., *Proceedings of The 27thCOLT*, volume 35 of *Proceedings of Machine Learning Research*, 408–422. Barcelona, Spain: PMLR.

Kleinberg, R.; Leyton-Brown, K.; and Lucier, B. 2017. Efficiency Through Procrastination: Approximately Optimal Algorithm Configuration with Runtime Guarantees. In *Proceedings of IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 2023–2031.

Lattanzi, S.; Lavastida, T.; Moseley, B.; and Vassilvitskii, S. 2020. Online Scheduling via Learned Weights. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, 1859–1877.

Lattimore, T.; and Szepesvári, C. 2020. *Bandit Algorithms*. Cambridge University Press.

Lavastida, T.; Moseley, B.; Ravi, R.; and Xu, C. 2021. Learnable and Instance-Robust Predictions for Online Matching, Flows and Load Balancing. In Mutzel, P.; Pagh, R.; and Herman, G., eds., *29th ESA 2021, September 6-8, 2021, Lisbon, Portugal (Virtual Conference)*, volume 204 of *LIPIcs*, 59:1–59:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Lykouris, T.; and Vassilvitskii, S. 2018. Competitive Caching with Machine Learned Advice. In *Proceedings of the 35th ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, 3302–3311.

Medina, A. M.; and Vassilvitskii, S. 2017. Revenue Optimization with Approximate Bid Predictions. In *NeurIPS 2017, 4-9 December 2017, Long Beach, CA, USA*, 1858–1866.

Mitzenmacher, M. 2020. Scheduling with Predictions and the Price of Misprediction. In Vidick, T., ed., *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA*, volume 151 of *LIPIcs*, 14:1–14:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Myerson, R. B. 1981. Optimal Auction Design. *Math. Oper. Res.*, 6(1): 58–73.

Purohit, M.; Svitkina, Z.; and Kumar, R. 2018. Improving Online Algorithms via ML Predictions. In *NeurIPS 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, 9684–9693.

Rohatgi, D. 2020. Near-Optimal Bounds for Online Caching with Machine Learned Advice. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, 1834–1845.

Rusmevichientong, P.; and Tsitsiklis, J. N. 2010. Linearly Parameterized Bandits. *Math. Oper. Res.*, 35(2): 395–411.

Shalev-Shwartz, S.; and Singer, Y. 2007. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69: 115–142.

Singla, S. 2018. The Price of Information in Combinatorial Optimization. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, 2523–2532.

Valko, M.; Munos, R.; Kveton, B.; and Kocák, T. 2014. Spectral Bandits for Smooth Graph Functions. In *ICML*.

van der Hoeven, D.; van Erven, T.; and Kotłowski, W. 2018. The Many Faces of Exponential Weights in Online Learning. In Bubeck, S.; Perchet, V.; and Rigollet, P., eds., *Proceedings of the 31st COLT*, volume 75 of *Proceedings of Machine Learning Research*, 2067–2092. PMLR.

Wang, S.; Li, J.; and Wang, S. 2020. Online Algorithms for Multi-shop Ski Rental with Machine Learned Advice. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *NeurIPS 2020, December 6-12, 2020, virtual*.

Wei, A. 2020. Better and Simpler Learning-Augmented Online Caching. In Byrka, J.; and Meka, R., eds., *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2020, August 17-19, 2020, Virtual Conference*, volume 176 of *LIPIcs*, 60:1–60:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Weisz, G.; Gyorgy, A.; and Szepesvari, C. 2018. LeapsAndBounds: A Method for Approximately Optimal Algorithm Configuration. In *International Conference on Machine Learning*, 5257–5265.

Weitzman, M. L. 1979. Optimal Search for the Best Alternative. *Econometrica*, 47(3): 641–654.