2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Online Learning for Equilibrium Pricing in Markets under Incomplete Information

Devansh Jalota*, Haoyuan Sun*, and Navid Azizan

Abstract— **The study of market *equilibria* is central to economic theory, particularly in efficiently allocating scarce resources. However, the computation of *equilibrium* prices at which the supply of goods matches their demand typically relies on having access to complete information on private attributes of agents, e.g., suppliers' cost functions, which are often unavailable in practice. Motivated by this practical consideration, we consider the problem of setting equilibrium prices in the incomplete information setting wherein a market operator seeks to satisfy the customer demand for a commodity by purchasing the required amount from competing suppliers with privately known cost functions unknown to the market operator. In this incomplete information setting, we consider the online learning problem of learning equilibrium prices over time while jointly optimizing three performance metrics—*unmet demand*, *cost regret*, and *payment regret*—pertinent in the context of equilibrium pricing over a horizon of T periods. In the general setting when suppliers' cost functions are time-varying, we show that no online algorithm can achieve sublinear regret on all three metrics. Thus, we consider the setting when suppliers' cost functions are fixed and develop algorithms that achieve a regret of (i) $O(\log\log T)$ when the customer demand is constant over time and (ii) $O(\sqrt{T}\log\log T)$ when the demand is variable over time.**

I. Introduction

The study of market mechanisms for efficiently allocating scarce resources traces back to the seminal work of Walras [2]. In his work, Walras investigated the design of pricing schemes to mediate the allocation of scarce resources such that the economy operates at an *equilibrium*, i.e., the supply of each good matches its demand. Market equilibria exist under mild conditions on agents' preferences [3] and, under convexity assumptions, can often be computed via a centralized optimization problem. As a case in point, in electricity markets with convex supplier cost functions, the equilibrium prices correspond to the shadow prices of a convex optimization problem that minimizes the sum of the supplier costs subject to a market clearing constraint [4].

While methods such as convex programming provide computationally tractable approaches for computing equilibria, the efficacy of such approaches suffers from several limitations. First, centralized optimization methods rely on complete information on agents' utilities and cost functions that are often unavailable to a market operator. For instance, with the deregulation of electricity markets, suppliers' cost functions are private information, which has led to strategic bidding [5], [6] and has been associated with millions of dollars of over-payments to suppliers [7]. Moreover, even if a market operator has some information on agents' utilities and cost functions, such information often only provides a noisy or imperfect estimate of their actual preferences [8]. In the context of electricity markets, the advent of renewables and distributed energy resources has accompanied a high degree of uncertainty in the supply of energy at different times of the day and year, as these energy sources are sensitive to weather conditions. To further compound these challenges, agents' preferences may also be time-varying, e.g., in electricity markets, customer demands may change over time and suppliers' cost functions may depend on fluctuating weather conditions. Thus, a market operator may need to periodically collect agents' preferences and solve a large-scale centralized optimization at each period to set equilibrium prices, which may be computationally challenging.

Motivated by the limitations of centralized optimization approaches for computing market equilibria, we study the problem of setting equilibrium prices in the incomplete information setting where a market operator seeks to satisfy the customer demand for a commodity by purchasing the required amount from competing suppliers with privately known cost functions. We investigate this problem under several informational settings regarding the time-varying nature of the customer demands and supplier cost functions and develop online learning algorithms that iteratively adjust the market prices over time for each of these settings. We employ the observation that a market operator can effectively learn information on suppliers' costs and equilibrium prices through observations of their production given different prices. To analyze our algorithms, we combine techniques from online learning and parametric optimization as we seek to jointly optimize multiple, often competing, performance metrics pertinent in the context of equilibrium pricing.

Contributions: In this work, we study the problem of setting equilibrium prices faced by a market operator that seeks to satisfy a customer demand for a commodity by purchasing the required amount from competing suppliers. We study this problem in the incomplete information setting when the cost functions of suppliers are private information and thus unknown to the market operator. In this setting, we consider the problem of learning equilibrium prices over T periods to achieve sub-linear regret, in the number of periods T, across three performance (regret) metrics: (i) *unmet demand*, (ii) *cost regret*, and (iii) *payment regret*. Here unmet demand refers to the cumulative difference between the demand and the total production of the commodity corresponding to an online pricing policy. Further, cost regret (payment regret) refers to the difference between the total cost of all suppliers (payment made to all suppliers) corre-

*Equal Contribution.

Please refer to the extended version [1] for omitted proofs and additional details on our results.

D. Jalota is with Stanford University, Stanford, CA 94305, USA djalota@stanford.edu

H. Sun and N. Azizan are with the Massachusetts Institute of Technology, Cambridge, MA 02139, USA {haoyuans,azizan}@mit.edu

This work was supported in part by MathWorks. D. Jalota acknowledges support from the Stanford Interdisciplinary Graduate Fellowship.

sponding to the online allocation and that of the offline oracle with complete information on suppliers' cost functions. For a detailed discussion of these regret metrics, see Section III-B.

We first consider the general setting when suppliers' cost functions can vary across time (Section IV) and show that if the operator does not know the process that governs the variation in the cost functions, no online algorithm can achieve sub-linear regret on all three regret metrics.

Given this impossibility result, we then consider the setting when suppliers' cost functions are fixed over time and develop algorithms with a regret of $O(\log \log T)$ when the customer demand is constant over time (Section V), and $O(\sqrt{T} \log \log T)$ when the demand is time-varying (Section VI), for strongly convex costs. To establish these regret guarantees for the three metrics, we leverage and combine techniques from parametric optimization and online learning. And in the extended version of our paper [1], we further show through an example that if the cost functions are not strongly convex, no online algorithm can achieve a sub-linear regret on all metrics.

Finally, to extend the sub-linear regret guarantees obtained in the fixed cost setting to the time-varying cost setting, we introduce an augmented setting in Section VII where the market operator has access to hints (contexts) that, without revealing the complete specification of the cost functions, reflect the change in the cost functions over time. In this augmented problem setting, we propose an algorithm that achieves sub-linear regret on all three regret metrics in the extended version of our paper [1].

II. Literature Review

The design of mechanisms to efficiently allocate resources under incomplete information on agents' preferences and costs has received considerable attention. For instance, mechanism design has enabled the optimal allocation of resources even in settings when certain information is privately known to agents [9]. Further, inverse game theory [10] and revealed preference approaches [11] have enabled learning agents' underlying preferences given past observations of their actions. While we also consider an incomplete information setting wherein suppliers' cost functions are private information, we instead study the problem of learning equilibrium prices as an online decision-making problem.

The paradigm of online-decision making has enabled the allocation of resources in settings with incomplete information and includes the well-studied problem classes of online linear programming (OLP) and online convex optimization (OCO). As in the works on OCO, we also consider convex objectives; however, as opposed to the resource constraints that need to be satisfied over the entire time horizon in the literature on OCO [12], [13], we adopt a stronger performance metric where we accumulate regret at each period when the customer demand is not satisfied (see Section III-B for more details on our performance metrics). Thus, our algorithms are considerably different from those in [12], [13] that involve applying dual sub-gradient descent.

Our algorithms are inspired by the multi-armed bandit (MAB) literature and involve a trade-off between exploration and exploitation [14]. In a typical MAB setting, a decision-maker performs sequential trials on a set of actions, observes the outcome of the actions, and maximizes a single reward function. However, we consider a setting of jointly optimizing multiple regret metrics where suppliers' cost functions are not revealed to the market operator.

III. Model

In this section, we present the offline market model to set equilibrium prices to satisfy the customer demand for a commodity (Section III-A) and regret metrics to evaluate the efficacy of an online pricing policy (Section III-B).

A. Market Model and Equilibrium Pricing

We study a market run by an operator seeking to meet the customer demand d for a commodity, e.g., electricity, by purchasing the required amount from n competing suppliers. Each supplier $i \in [n]$ has a cost function $c_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, where $c_i(x_i)$ is the cost incurred by supplier i for producing x_i units of the commodity. To meet the customer demand, the market operator posts a price p as the payment made for each unit of the commodity produced by a given supplier. In particular, for producing x_i units of the commodity, a supplier i receives a payment of $p\, x_i$ from the operator. Then, given a posted price p and a cost function $c_i(\cdot)$, each supplier individually decides on the optimal production quantity $x_i^*(p)$ to maximize their total profit, as described through the following optimization problem

$$\max_{x_i \geq 0} \quad p x_i - c_i(x_i). \tag{1}$$

The posted price that suppliers best respond to is set by a market operator that seeks to determine an equilibrium price p^* that satisfies the following three desirable properties:

1) Market Clearing: The total supply equals the total demand, i.e., $\sum_{i=1}^n x_i^*(p^*) = d$.
2) Minimal Supplier Cost: The total production cost $\sum_{i=1}^n c_i(x_i^*(p^*))$ of all suppliers is minimal among all feasible production quantities $x_i \geq 0$ for all $i \in [n]$ satisfying the customer demand, i.e., $\sum_{i=1}^n x_i = d$.
3) Minimal Payment: The total payment $\sum_{i=1}^n p^* x_i^*(p^*)$ made to all suppliers is minimal among all feasible production quantities satisfying the customer demand.

While these properties are, in general, not possible to achieve simultaneously, e.g., in markets where the supplier cost functions are non-convex [4], in markets where the cost functions $c_i(\cdot)$ of all suppliers are convex, there exists an equilibrium price p^* that satisfies the above properties. Moreover, in markets with convex costs, the equilibrium price can be computed through the dual variables of the market clearing constraint of the following convex optimization problem

$$\min_{x_i \geq 0, \forall i \in [n]} \sum_{i \in [n]} c_i(x_i) \quad \text{s.t.} \quad \sum_{i \in [n]} x_i = d, \tag{2}$$

where the objective is to minimize the total supplier production cost. While the equilibrium price p^* has several desirable properties, it typically cannot be directly computed by solving Problem (2) as suppliers' cost functions are, in general, unknown to the market operator. Further, suppliers' cost functions and customer demands tend to be time-varying and thus would involve the operator periodically re-solving Problem (2) to determine equilibrium prices at short time intervals, which may be computationally prohibitive. To overcome these challenges, we propose online learning algorithms to learn equilibrium prices over time in the incomplete information setting when the suppliers' cost functions are unknown (or only partially known) to the market operator.

B. Performance Metrics

We now introduce the online learning setting, wherein the market operator sets prices for the commodity over multiple periods, and present the performance metrics to evaluate the efficacy of an online pricing policy. In particular, we consider the setting when the market operator seeks to satisfy the customer demand over multiple periods $t = 1, \ldots, T$. At each period $t \in [T]$, the customer demand for the commodity is given by d_t and each supplier $i \in [n]$ has a private cost function $c_{it}(\cdot)$ that is increasing, continuously differentiable, strongly convex, and normalized to satisfy $c_{it}(0) = 0$. We assume that the demand at each period t lies in a bounded interval, i.e., $d_t \in [\underline{d}, \bar{d}]$ for all t for some $\underline{d}, \bar{d} > 0$. Further, for ease of exposition, we normalize the set of feasible prices corresponding to any customer demand and realization of supplier cost functions to be such that the optimal price of the commodity belongs to the normalized interval $[0, 1]$. In addition, we note that we consider strongly convex supplier cost functions, as opposed to general convex costs, due to the performance limitations of any online algorithm under the incomplete information setting studied in this work for non-strongly convex costs (see Section V for further details).

In this work, upon observing the demand d_t, the market operator sets a price p_t that depends on the realized customer demands and past observations of supplier productions, i.e., revealed preference feedback [15] in response to set prices. Over the T periods, the market operator sets prices given by the pricing policy $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)$, where $p_t = \pi_t(\{(x_{it'}^*)_{1=1}^n, d_{t'}\}_{t'=1}^{t-1}, d_t)$, where x_{it}^* is the optimal solution of Problem (1) for supplier i at period t. When the pricing policy is clear from the context, we will overload the notation and simply write $\boldsymbol{\pi} = (p_1, p_2, \ldots, p_T)$.

We evaluate the efficacy of an online pricing policy $\boldsymbol{\pi}$ using three regret metrics: (i) unmet demand, (ii) cost regret, and (iii) payment regret. These regret metrics represent the performance loss of the policy $\boldsymbol{\pi}$ relative to the optimal offline algorithm with complete information on the three desirable properties of equilibrium prices elucidated in Section III-A. We also note that these performance metrics naturally generalize to the augmented problem setting we consider when suppliers' cost functions are time-varying and present the corresponding generalizations in the extended version of our paper [1] for completeness.

a) Unmet Demand: We evaluate the unmet demand through the sum of the differences between the demand and the total supplier productions corresponding to $\boldsymbol{\pi}$ at each period. Letting $x_{it}^*(p_t)$ be the optimal solution of Problem (1) for supplier i at period t, the unmet demand for an online pricing policy $\boldsymbol{\pi}$ is $U_T(\boldsymbol{\pi}) = \sum_{t \in [T]} \left(d_t - \sum_{i \in [n]} x_{it}^*(p_t)\right)_+$.

b) Cost Regret: We evaluate the cost regret of a pricing policy $\boldsymbol{\pi}$ through the difference between its total supplier production cost and the minimum total production cost, given complete information on the supplier cost functions. In particular, the cost regret $C_T(\boldsymbol{\pi})$ of an algorithm $\boldsymbol{\pi}$ is $C_T(\boldsymbol{\pi}) = \sum_{t \in [T]} \sum_{i \in [n]} \left(c_{it}(x_{it}^*(p_t)) - c_{it}(x_{it}^*(p_t^*))\right)$, where the price p_t^* for each period $t \in [T]$ is the optimal dual variable of the market clearing constraint of Problem (2) given the demand d_t and cost functions c_{it} for all $i \in [n]$.

c) Payment Regret: We evaluate the payment regret through the difference between the total payment made to all suppliers corresponding to $\boldsymbol{\pi}$ and the minimum total payment, given complete information on the cost functions. In particular, the payment regret of an algorithm $\boldsymbol{\pi}$ is $P_T(\boldsymbol{\pi}) = \sum_{t \in [T]} \sum_{i \in [n]} \left(p_t x_{it}^*(p_t) - p_t^* x_{it}^*(p_t^*)\right)$.

In this work, we develop algorithms that jointly optimize and achieve a sub-linear regret, in the number of periods T, on these three regret metrics. Note that achieving good performance on one of these metrics is typically easy as setting low prices will lead to low cost and payment regrets while setting high prices will lead to no unmet demand. Thus, the challenge is to find the equilibrium price at which all these regret metrics are kept small.

A few comments about our regret metrics are in order. First, our unmet demand metric aligns with real-world markets, e.g., electricity markets, where the demand needs to be satisfied at each period, and over-production at particular periods cannot compensate for unmet demand at subsequent periods. Therefore, we define our unmet demand metric as a stronger benchmark than the typical constraint violation metrics in the literature of jointly optimizing multiple regret metrics [16], [17], [18], where resource constraints only need to be approximately satisfied in the long run. Formally, $U_T(\boldsymbol{\pi}) = \sum_{t=1}^T \left(d_t - \sum_{i=1}^n x_{it}^*(p_t)\right)_+ \geq \left[\sum_{t=1}^T \left(d_t - \sum_{i=1}^n x_{it}^*(p_t)\right)\right]_+$, where the latter term corresponds to the setting when the customer demand only needs to be satisfied in the long-run. Further, since we obtain regret guarantees for the above unmet demand metric using techniques from parametric optimization, our regret guarantees naturally extend for the corresponding stronger notions of the payment and cost regret metrics as well. However, we present our payment and cost regret metrics in alignment with the classical regret metrics in the literature, wherein lower payments (costs) at particular periods can compensate for excess payments (costs) at other periods.

IV. Impossibility With Time-varying Cost Functions

In this section, we consider the general setting, where both suppliers' cost functions and customer demands are time-varying and show that if the operator does not know the process governing the variation of the cost functions, then it is impossible to simultaneously achieve sub-linear regret on all three regret metrics. In particular, Proposition 1 presents a counterexample establishing that even if the cost functions are drawn i.i.d. from a known distribution, no online algorithm can achieve sub-linear regret on all three metrics if the operator is not informed about the outcome of the random draws from the distribution.

Proposition 1 (Impossibility of Sub-linear Regret): There exists an instance with time-invariant demand and a single supplier whose cost functions are drawn i.i.d. from a known distribution such that no online algorithm can achieve sub-linear regret on all three regret metrics.

Proof: (Sketch). Consider a setting with a fixed demand of $d = 1$ and a single supplier whose cost functions are drawn from a distribution such that its cost function is either $c_1(x) = \frac{1}{8}x^2$ or $c_2(x) = \frac{1}{16}x^2$, each with probability 0.5, at each period t. We suppose that the market operator has knowledge of the distribution from which the cost function is sampled i.i.d. but does not know the outcome of the random draw at any period. Then, we analyze the total regret, i.e., the sum of the unmet demand, payment regret, and cost regret for three price ranges - (i) $p < \frac{1}{8}$, (ii) $1/8 \leq p \leq 1/4$, and (iii) $p > 1/4$ - and show that irrespective of the set price at any period, the expected total regret at any period is at least $\frac{7}{64}$, i.e., the total regret is at least $\frac{7}{64}T$. Finally, since the sum of the three regret metrics is linear in T, at least one of them must be linear in T, establishing our claim. ■

For a complete proof of Proposition 1, see the extended version of our paper [1].

Proposition 1 establishes that sub-linear regret on all three metrics is not possible for a general sequence of time-varying cost functions. The setting with time-varying costs is challenging because the equilibrium prices may change over time even if the customer demand remains constant. In contrast to the settings in [16], [13], where online gradient descent approaches can simultaneously achieve sub-linear regret for multiple performance metrics, we note that our definition of unmet demand is considerably stronger as over-production at particular periods cannot compensate for unmet demand at other periods (see Section III-B for more details). Thus, Proposition 1 shows that, with the stronger unmet demand metric, it is impossible to jointly optimize the three regret metrics, where decreasing the payment or cost regret must lead to an increase in the unmet demand and vice versa.

While it is not possible to achieve sub-linear regret in the setting when suppliers' cost functions are time-varying, we develop algorithms with sub-linear regret in the fixed cost function setting (see Sections V and VI). Furthermore, in the extended version of our paper [1], we consider an augmented setting when the market operator has additional context on how a suppliers' cost function varies over time (see Section VII) and design an algorithm in this augmented setting with sub-linear regret guarantees.

V. Fixed Cost Functions and Demand

Given the impossibility of achieving sub-linear regret in the setting with time-varying costs, we now investigate the design of online pricing policies with sub-linear regret in the setting with fixed costs. To this end, as a warm-up, we first consider the setting when suppliers' cost functions and customer demand are fixed over time, i.e., the cost functions satisfy $c_{it}(\cdot) = c_{it'}(\cdot)$ for all $t, t' \in [T]$ and $i \in [n]$ and the demands satisfy $d_t = d_{t'}$ for all $t, t' \in [T]$. For ease of exposition, in this section, we drop the subscript t in the notation for the demand (and cost functions) and denote $d_t = d$ (and $c_{it}(\cdot) = c_i(\cdot)$ for all i) for all periods t. In this setting, we develop an algorithm that achieves a regret of $O(\log \log T)$ on the three regret metrics when the suppliers' cost functions are strongly convex.

To motivate our algorithm, we first note that since the demand and cost functions are fixed, the optimal price $p^* \in [0,1]$ is also fixed for all periods. Further, the cumulative production $x_t^*(p) = \sum_{i=1}^n x_{it}^*(p)$ is monotonically non-decreasing in the price p as suppliers' cost functions are increasing. Using this monotonicity property, note that if $x_t^*(p_1) > d$ and $x_t^*(p_2) < d$ for two prices p_1, p_2, then p_1 and p_2, respectively, are upper and lower bounds on the optimal price.

Following these observations, we present Algorithm 1, which maintains a feasible interval for the optimal price p^* and sets a sequence of prices for each arriving user to continuously shrink this feasible price set. In particular, the feasible price interval $[a, b]$ is initialized to $\mathcal{S}_p = [0, 1]$ and a precision parameter ε is set to 0.5. Then, for a given algorithm sub-phase associated with feasible price interval $[a, b]$, the operator posts prices $a, a+\varepsilon, a+2\varepsilon, \ldots$ (up to b) at each period until the total supply exceeds the demand at the offered price. If $a + k\varepsilon$ for some $k \in \mathbb{N}$ was the last price such that $x_t^*(a+k\varepsilon) \leq d$, then $[a+k\varepsilon, a+(k+1)\varepsilon]$ is set as the new feasible interval for the optimal price, and the precision parameter is re-set to ε^2. This process of shrinking the feasible interval and updating the precision parameter is repeated until the length of the feasible interval is smaller than $\frac{1}{T}$, following which the market operator posts the price at the lower end of the feasible interval for the remaining periods. This process is presented formally in Algorithm 1.

Algorithm 1: Feasible Price Set Tracking

Initialize feasible price set $\mathcal{S}_p = [a, b] \leftarrow [0, 1]$ and precision parameter $\varepsilon = \frac{1}{2}$;
while *length of feasible price set is more than* $\frac{1}{T}$ **do**
  Offer prices $a, a+\varepsilon, \ldots, a+(k+1)\varepsilon$ (all of which are $\leq b$) to each subsequent request where $a + k\varepsilon$ is the last price where $x_t^*(a+k\varepsilon) < d$;
  Set feasible interval $\mathcal{S}_p$ to $[a+k\varepsilon, a+(k+1)\varepsilon]$ and reduce the precision parameter to ε^2 ;
end
Offer price $p_t = a$ for the remaining periods ;

While Algorithm 1 is similar to the corresponding algorithm in [19] for the setting of fixed user valuations, our market setting is considerably different than the revenue maximization setting in [19]. First, in this work, suppliers have a continuous rather than a binary action space as in [19]. Further, as opposed to the single regret metric in [19], we consider three, often competing, regret metrics.

We now show that Algorithm 1 achieves an $O(\log \log T)$ regret on the three regret metrics in Section III-B.

Theorem 2 (Sub-Linear Regret Fixed Demand): The unmet demand, cost regret, and payment regret of Algorithm 1 are $O(\log \log T)$ if the cost functions are strongly convex.

The proof of Theorem 2 relies on the following Lipschitzness condition between the optimal supplier production and the prices set by the market operator.

Lemma 3 (Lipschitzness of Production in Prices): If the functions $c_i(\cdot)$ are μ_i-strongly convex, then, at any period t, the optimal production of supplier i given by the solution of Problem (1) is Lipschitz in the price, i.e., for all $p_1, p_2 \in [0, 1]$, $|x_{it}^*(p_1) - x_{it}^*(p_2)| \leq L|p_1 - p_2|$ for some $L > 0$.

Lemma 3 follows from a direct application of the inverse function theorem. Using Lemma 3, we now present a proof sketch of Theorem 2 and present its complete proof with that of Lemma 3 in the extended version of our paper [1].

Proof: (Sketch). To establish this result, we first note that we need $O(\log\log T)$ sub-phases of repeated squaring of the parameter ε to reduce ε from 0.5 to $\frac{1}{T}$. Due to the monotonicity of the optimal supplier production in the prices, both payment and cost regret are only accumulated when $p_t > p^*$. Further, since $p_t > p^*$ occurs at most once in each sub-phase, the total payment and cost regret are $O(\log\log T)$. Next, to bound the unmet demand, we use Lemma 3 to map prices to productions and show that there is a constant unmet demand accumulated in each sub-phase in Algorithm 1, resulting in an $O(\log\log T)$ unmet demand when the length of the feasible price interval is more than $\frac{1}{T}$, as there are $O(\log\log T)$ sub-phases. In the final phase, when the length of the feasible price interval is less than $\frac{1}{T}$ and the price is fixed, we again use Lemma 3 to show that the unmet demand through this phase is constant. Thus, the unmet demand is $O(\log\log T)$, establishing our claim. ∎

We reiterate that the proof of Theorem 2 crucially relies on the strong convexity of the cost functions of the suppliers, which was necessary to establish Lemma 3. We also note that the $O(\log\log T)$ regret obtained in Theorem 2 indicates that in the setting with fixed demands and supplier cost functions, Algorithm 1's prices converge to the equilibrium price p^* super-exponentially and the pricing policy corresponding to Algorithm 1 incurs little performance loss on all three regret metrics as compared to when equilibrium prices are set with complete knowledge of suppliers' cost functions. Furthermore, the obtained regret guarantee of Algorithm 1 compares favorably to the $\Omega(\log\log T)$ regret lower bound for any online algorithm in the revenue maximization setting with fixed user valuations studied in [19].

While Algorithm 1 achieved sub-linear regret on all three metrics for strongly convex costs, we note that this result does not generalize to the setting of general convex costs. In particular, in the extended version of our paper [1], we show that sub-linear regret cannot be achieved simultaneously on all three regret metrics for linear cost functions.

VI. Fixed Costs and Time-varying Demand

In this section, we investigate a more general setting where the suppliers' cost functions are static over time while customer demands can vary in a continuous but bounded interval, i.e., $d_t \in [\underline{d}, \bar{d}]$. In this setting, we extend the algorithm developed for fixed supplier cost functions and customer demands (Algorithm 1) and show that it achieves a regret of $O(\sqrt{T}\log\log T)$ on all three regret metrics for strongly convex cost functions.

To motivate our algorithmic approach and address the challenge that the demands can vary between the interval $[\underline{d}, \bar{d}]$, we first consider a direct extension of Algorithm 1, wherein a feasible price set is maintained for each realized demand. However, as there may be up to $O(T)$ different demand realizations, the worst-case regret of such an algorithm is $O(T)$. To resolve this issue, we leverage the intuition that customer demands that are close to each other correspond to equilibrium prices that are also close together. Thus, we uniformly partition the demand interval $[\underline{d}, \bar{d}]$ into sub-intervals of width γ and consider any demand in the same sub-interval the same. In particular, any demand lying in a given sub-interval, i.e., $d_t \in [\underline{d}+k\gamma, \underline{d}+(k+1)\gamma]$ for some $k \in \mathbb{N}$, is considered as a demand equal to the lower bound of that interval. Then, for these $O(\frac{1}{\gamma})$ distinct demands, corresponding to the lower bounds of the $O(\frac{1}{\gamma})$ sub-intervals, we apply the aforementioned direct extension of Algorithm 1. Our approach is formally presented in Algorithm 2.

Algorithm 2: Feasible Price Set Tracking for Time-Varying Demands

---

Input : Discretized demand intervals $I_1, \ldots, I_K$ where $K = \lceil (\bar{d} - \underline{d})/\gamma \rceil$ and $I_k = \{\underline{d} + (k-1)\gamma, \underline{d} + k\gamma\}$

Initialize feasible price set $\mathcal{S}_k = (0, 1]$, current price $p_k = 0$, and price precision $\varepsilon_k = 1/2$ for each I_k;

for $t = 1, \ldots, T$ **do**

  Determine k_t such that $d_t \in I_{k_t} =: [a_{k_t}, b_{k_t}]$;

  Offer price p_{k_t} to the supplier;

  if *width of feasible price set* $|\mathcal{S}_{k_t}| \geq \frac{1}{\sqrt{T}}$ **then**

    if $\sum_{i=1}^{n} x_{it}^*(p_{k_t}) \geq a_{k_t}$ **then**

      Set $\mathcal{S}_{k_t} \leftarrow (p_{k_t} - \varepsilon_{k_t}, p_{k_t}]$;

      Set next price $p_{k_t} \leftarrow p_{k_t} - \varepsilon_{k_t}$;

      Re-set the precision to $\varepsilon_{k_t} \leftarrow \varepsilon_{k_t}^2$;

    else

      Set next price $p_{k_t} \leftarrow p_{k_t} + \varepsilon_{k_t}$

  end

end

We now show that Algorithm 2 achieves a regret of $O(\sqrt{T}\log\log T)$ if the sub-interval width $\gamma = \frac{1}{\sqrt{T}}$ for strongly convex cost functions of suppliers. We note that choosing $\gamma = \frac{1}{\sqrt{T}}$ optimally balances between two different sources of regret in the time-varying demand setting, as is elucidated through the proof sketch of the following theorem.

Theorem 4 (Sub-Linear Regret Varying Demand): Let $\gamma = \frac{1}{\sqrt{T}}$. Then, the unmet demand, cost regret, and payment regret of Algorithm 2 are $O(\sqrt{T}\log\log T)$ if the cost functions of the suppliers are strongly convex.

Proof: (Sketch) For each regret metric, the regret incurred by Algorithm 2 can be divided into two parts: (i) the regret incurred by the Algorithm 1 sub-routine for each demand sub-interval, and (ii) the inaccuracies of considering all demands in a given sub-interval to be the lower bound of that sub-interval. By Theorem 2, the first part is $O(K\log\log T)$ for each regret metric, where $K := \lceil (\bar{d} - \underline{d})/\gamma \rceil$. Next, since all demands in a given sub-interval are treated as a demand equal to the lower bound of that sub-interval and the suppliers' optimal production is monotonic in the price, every price p_t in Algorithm 2 is an under-estimate to the equilibrium price for demand d_t. Thus, the second part of the regret is only positive for the unmet demand and is at most $O(\gamma T)$, as the width of each sub-interval is γ, and there are T periods. Finally, choosing $\gamma = \frac{1}{\sqrt{T}}$ achieves an optimal balance (up to logarithmic terms) between the two quantities, i.e., $O(\gamma^{-1}\log\log T)$ and $O(\gamma T)$, which establishes the $O(\sqrt{T}\log\log T)$ regret bound. ∎

For a complete proof of Theorem 4, see the extended version of our paper [1]. We reiterate that Theorem 4 applies to strongly convex costs as with Theorem 2 and that extending this result to general convex costs, e.g., linear costs, is, in general, not possible (see Section V). Further, compared to Theorem 2 for fixed demands, Theorem 4 shows that time-varying demands result in an additional factor of $O(\sqrt{T})$ in the regret. However, we do note that if the set of demand realizations D is known *a priori* to be $o(\sqrt{T})$, then the guarantee in Theorem 2 can be improved to $O(|D| \log\log T)$ by running the direct extension of Algorithm 1, wherein a feasible price interval is maintained for each demand. Finally, we note that the regret guarantee obtained in Theorem 4 compares favorably to classical $O(\sqrt{T})$ regret guarantees in the OCO or MAB literature [20].

VII. Augmented Setting for Time-Varying Costs

Thus far, we have developed algorithms with sub-linear regret in the settings when the suppliers' cost functions are fixed and demonstrated the impossibility of achieving sub-linear regret in the setting with time-varying supplier cost functions. Given the performance limitations in the setting with time-varying costs, we now propose a natural augmented problem setting wherein the market operator, without knowing the complete specification of cost functions of suppliers, additionally has access to a hint (i.e., context) that reflects the variation in cost functions of suppliers over time. Such a setting aligns with real-world markets, e.g., electricity markets, wherein the cost functions of suppliers are private information yet will typically vary over time based on observed quantities, such as the weather conditions.

In this augmented setting, we assume that each supplier's cost function has two parts: (i) an unknown time-invariant component, and (ii) a time-varying component revealed as a hint to the market operator. More precisely, the cost function of each supplier i is parameterized as $c_{it}(\cdot) = c_i(\cdot; \phi_i, \theta_{it})$, where ϕ_i is private information and θ_{it} is the time-varying component given to the operator as a *context*. Note that for any fixed ϕ_i, the context θ_{it} uniquely determines the cost function of supplier i at period t. We stress that we do not assume any structure on the parameterization of the cost functions and so the time-varying and time-invariant components of the cost functions need not be separable. Further, since ϕ_i's are unknown, the market operator cannot directly solve Problem (2) to obtain equilibrium prices.

In this augmented setting, at each period t, in addition to receiving the demand d_t, the market operator observes a context θ_t, which it can use along with the prior history of supplier production quantities, demands, and contexts, to set a price p_t. In particular, the market operator sets a sequence of prices given by the pricing policy $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)$, where $p_t = \pi_t(\{(x_{t'}^*)_{1=1}^n, d_{t'}, \theta_{t'}\}_{t'=1}^{t-1}, d_t, \theta_t)$, where x_t^* represents the sum of optimal production quantity given by the solution of Problem (1) for each supplier at period t. The performance of this class of pricing policies can be evaluated by the three regret metrics in Section III-B. Note that we can naturally extend these three metrics to the augmented setting with contexts by plugging in $c_{it}(\cdot) = c_i(\cdot; \phi_i, \theta_{it})$.

This augmented problem naturally leads to a formulation based on ideas from the contextual bandit literature. In the extended version of our paper [1], we present a more precise problem statement and propose an algorithm which can simultaneously achieve sublinear regret on all three performance metrics.

VIII. Conclusion and Future Work

We studied the problem of equilibrium pricing in markets where suppliers' cost functions are unknown to the operator. There are several directions for future research. First, it would be interesting to investigate the design of algorithms whose memory does not scale with T but still achieves similar guarantees to Algorithm 2. Next, it would be worthwhile to study the design of algorithms under a more relaxed unmet demand notion with the possibility of using excess supply at earlier periods to satisfy the demand at subsequent periods. Further, there is a scope to generalize the model to the setting when suppliers' cost functions are non-convex, in which case the operator may need different pricing strategies for each supplier [4].

References

[1] D. Jalota, H. Sun, and N. Azizan, "Online learning for equilibrium pricing in markets under incomplete information," *arXiv preprint arXiv:2303.11522*, 2023.

[2] L. Walras, *Elements of Pure Economics: Or, The Theory of Social Wealth*, ser. American Economic Association. Translation series. American Economic Association, 1954.

[3] K. J. Arrow and G. Debreu, "Existence of an equilibrium for a competitive economy," *Econometrica*, vol. 22, no. 3, 1954.

[4] N. Azizan, Y. Su, K. Dvijotham, and A. Wierman, "Optimal pricing in markets with nonconvex costs," *Operations Research*, vol. 68, no. 2, pp. 480–496, 2020.

[5] G. Liberopoulos and P. Andrianesis, "Critical review of pricing schemes in markets with non-convex costs," *Operations Research*, vol. 64, no. 1, pp. 17–31, 2016.

[6] A. K. David and F. Wen, "Strategic bidding in competitive electricity markets: a literature survey," in *Power Engineering Society Summer Meeting*, vol. 4. IEEE, 2000, pp. 2168–2173.

[7] B. Wingfield and D. Kopecki, "Jpmorgan to pay $410 million to settle u.s. energy probe," Bloomberg, July 2013. [Online]. Available: https://www.bloomberg.com/news/articles/2013-07-30/jpmorgan-to-pay-410-million-in-u-s-ferc-settlement?leadSource=uverify%20wall

[8] M. L. Weitzman, "Prices vs. Quantities12," *The Review of Economic Studies*, vol. 41, no. 4, pp. 477–491, 1974.

[9] N. Heydaribeni and A. Anastasopoulos, "Distributed mechanism design for multicast transmission," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4200–4205.

[10] D. Bertsimas, V. Gupta, and I. C. Paschalidis, "Data-driven estimation in equilibrium using inverse optimization," *Math. Program.*, vol. 153, no. 2, p. 595–633, Nov 2015.

[11] M.-F. Balcan, A. Daniely, R. Mehta, R. Urner, and V. V. Vazirani, "Learning economic parameters from revealed preferences," in *International Conference on Web and Internet Economics*. Springer, 2014.

[12] S. Agrawal and N. R. Devanur, "Fast algorithms for online stochastic convex programming," in *ACM-SIAM Symposium on Discrete Algorithms*, P. Indyk, Ed. SIAM, 2015, pp. 1405–1424.

[13] S. R. Balseiro, H. Lu, and V. Mirrokni, "The best of many worlds: Dual mirror descent for online allocation problems," *Operations Research*, 2022.

[14] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[15] A. Roth, J. Ullman, and Z. S. Wu, "Watch and learn: Optimizing from revealed preferences feedback," in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016, pp. 949–962.

[16] H. Yu, M. Neely, and X. Wei, "Online convex optimization with stochastic constraints," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[17] R. Jenatton, J. Huang, and C. Archambeau, "Adaptive algorithms for online convex optimization with long-term constraints," in *International Conference on Machine Learning*. PMLR, 2016, pp. 402–411.

[18] M. Mahdavi, R. Jin, and T. Yang, "Trading regret for efficiency: online convex optimization with long term constraints," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2503–2528, 2012.

[19] R. Kleinberg and T. Leighton, "The value of knowing a demand curve: bounds on regret for online posted-price auctions," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, 2003, pp. 594–605.

[20] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.