Optimal Posted Prices for Online Resource Allocation with Supply Costs

Xiaoqi Tan, Alberto Leon-Garcia
ECE, University of Toronto
firstname.lastname@utoronto.ca

Danny H.K. Tsang,
ECE, HKUST
eetsang@ust.hk

ABSTRACT

We study a general online resource allocation problem, where a service provider sells multiple types of capacity-limited resources to heterogeneous customers that arrive in a sequential manner. The provider charges payment from customers who purchase the resource but must pay an increasing marginal supply cost with respect to the total resource allocated (e.g., production costs and/or operational costs). The goal is to maximize the social welfare, namely, the total valuation of customers minus the total supply cost of the provider. We adopt the standard competitive analysis framework and provide an optimal online posted-pricing mechanism. Our online mechanism is *optimal* in the sense that no other online algorithms can achieve a better competitive ratio.

1. INTRODUCTION

In this paper, we study a general online resource allocation problem in network economics and online optimization. A service provider has multiple types of capacity-limited resources (e.g., computing cycles, storage, or electricity) with a set of prices at which they can be sold. Heterogeneous customers arrive in an arbitrary manner. Upon the arrival of a customer, she will be offered a price and adopt a take-it-or-leave-it policy. The provider charges payment from those customers who purchase the resource but must pay an increasing marginal supply cost (e.g., production cost and/or operational cost). The goal is to design a set of online posted prices to maximize the social welfare, namely, the total valuation of customers minus the cost of the provider.

Online resource allocation problems have been studied extensively in different research communities. However, most of the existing literature assume that the resource allocation can be performed without increasing the cost of the provider. This assumption is reasonable for the case of online allocation of digital goods [1], while for most other applications, this is usually not the case as either the production cost or the operational cost is an increasing function of the resources that have been allocated (i.e., diseconomies of scale). For example, in cloud computing, the operational costs of data centers are usually an increasing function of its computing resources that are currently occupied. Motivated by this, Blum et al. [2] pioneered the study of online combinatorial auctions with increasing production cost. In this setting, the seller can produce any number of copies of the items being sold (i.e., without capacity limit), but needs to pay a non-decreasing marginal production cost per copy. Blum et al. proposed a pricing scheme called *twice-the-index* for several reasonable marginal production cost functions such as linear, lower-degree polynomial and logarithmic functions. For each of them, a constant competitive ratio was derived. In particular, for polynomial cost function $f(\omega) = \rho\omega^\gamma$, the competitive ratio is $(1+\epsilon)4\gamma$, where $\epsilon > 0$ is an adjustable parameter to achieve a certain tradeoff between competitive ratio and additive loss. Huang et al. [3] later studied an almost identical problem and achieved a tighter competitive ratio with a unified pricing framework. In particular, for polynomial cost function $f(\omega) = \rho\omega^\gamma$ without capacity limit, Huang et al. [3] proposed a $\gamma^{\gamma/(\gamma-1)}$-competitive algorithm when each customer is fractional, and they also proved that no other online algorithms can be $(\gamma^{\gamma/(\gamma-1)} - \epsilon)$-competitive for any $\epsilon > 0$.

(**Our Contribution**) We develop optimal posted-pricing algorithms for capacity-limited online resource allocation problems under increasing marginal costs. We focus on social welfare maximization and make the following contributions. First, we derive an analytical lower bound $\underline{\alpha}(u)$ for the competitive ratio of any pricing function indexed by u, where u is the resource utilization level when the posted price equals the maximum marginal cost. We prove that there exists an optimal u_* such that for any $\epsilon > 0$, no other online algorithms can be $(\underline{\alpha}(u_*) - \epsilon)$-competitive. Our lower bound $\underline{\alpha}(\cdot)$ generalizes the optimal competitive ratio derived in [3] to the capacity-limited case. Second, we provide an optimal pricing framework to achieve the best-possible competitive ratios of any online algorithms. Our proposed optimal pricing functions improve the results of [2] and [3]. Interestingly, the pricing scheme proposed in [3] is a special case in our pricing framework when the capacity limit is relaxed, and directly applying the design in [3] is suboptimal in the capacity-limited case. Third, our optimal pricing function is designed based on the assumption of knowing the maximum valuation-to-demand ratio (say $\overline{p}$), as in most existing work. However, we obtain a counter-intuitive result, namely, when $\overline{p}$ is lower than a certain threshold (which can be quantified), the optimal competitive ratio can be achieved regardless of having the exact knowledge of $\overline{p}$. Therefore, we can relax the assumption of knowing the exact value of $\overline{p}$ under this scenario, leading to an optimal online algorithm that requires as fewer assumptions as possible.

2. PROBLEM STATEMENT

We consider the problem of a single service provider, who is selling a set $\mathcal{K} = \{1, \cdots, K\}$ of K types of resources to its customers $\mathcal{I} = \{1, \cdots, I\}$. Customers arrive one at a time in some arbitrary manner and want to purchase a bundle of resources $b \in \mathcal{B}_i$ based on their own private preferences, where $\mathcal{B}_i$ denotes all the possible bundles for customer i. We denote the demand of customer i for resource type k in bundle b by $r_{i,k}^b$. We consider limited-supply, and normalize the capacity limit to be 1, and thus the demand $r_{i,k}^b$ denotes the proportion of the capacity limit. Each customer i decides whether to buy a bundle of resources or not based on the posted price p_k for each resource type k, and picks the utility-maximizing bundle by solving $\arg\max_{b\in\mathcal{B}_i}\ v_i^b - \sum_{k\in\mathcal{K}} p_k r_{i,k}^b$, where v_i^b is the valuation for the resources in bundle b, and $\sum_{k\in\mathcal{K}} p_k r_{i,k}^b$ is the payment made by customer i if bundle b is chosen. Considering the individual rationality, customers will purchase the resource if and only if the payment is less than or equal to the valuation (i.e., non-negative utility).

Let us denote the choice of customer i by binary variable $x_i^b \in \{0,1\}$, where $x_i^b = 1$ denotes customer i chooses to buy the b-th bundle, and $x_i^b = 0$ otherwise. We use $x_i^b(\texttt{alg})$ to denote the dependency of x_i^b on a particular pricing algorithm $\texttt{alg}$. Obviously we have $\sum_{b\in\mathcal{B}_i} x_i^b(\texttt{alg}) \leq 1$. The provider collects payments from all the customers who purchase the resources and pays a total cost of $\sum_{k\in\mathcal{K}} f(w_k)$, where $w_k = \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}_i} r_{i,k}^b x_i^b(\texttt{alg})$ denotes the total resource allocated by algorithm $\texttt{alg}$, and $f(\cdot)$ denotes total supply cost of the service provider[1]. We focus on the power cost case and assume $f(\cdot)$ takes the following form

$$f(\omega) = \begin{cases} \rho\omega^\gamma & \text{if } 0 \leq \omega \leq 1, \\ +\infty & \text{if } \omega > 1, \end{cases} \tag{1}$$

where 1 denotes the normalized capacity limit and $\gamma \geq 2$.

Let $W_{\texttt{alg}}$ denote the social welfare achieved by online algorithm $\texttt{alg}$, which is given by summing over the utilities of all the customers and the sevice provider, i.e., $W_{\texttt{alg}} = \sum_{i\in\mathcal{I}} v_i^b x_i^b(\texttt{alg}) - \sum_{k\in\mathcal{K}} f(w_k)$, where the payment terms cancel out and thus the social welfare equals the total valuation of customers minus the total cost of the provider. If we assume a complete knowledge of customer arrival information, then the social welfare maximization in the offline setting can be written as follows:

$$W_{\texttt{opt}} = \max_{\{x_i^b\}\forall i,b} \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}_i} v_i^b x_i^b - \sum_{k\in\mathcal{K}} f(w_k)$$

$$s.t. \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}_i} r_{i,k}^b x_i^b = w_k, \forall k; \sum_{b\in\mathcal{B}_i} x_i^b \leq 1, x_i^b \in \{0,1\}, \forall i,$$

where $W_{\texttt{opt}}$ denotes the optimal social welfare. Our target is to design an online posted-pricing algorithm (i.e., the design of p_k) such that $W_{\texttt{alg}}$ is as close to $W_{\texttt{opt}}$ as possible. The performance of our pricing algorithm is quantified by the standard competitive analysis framework, namely, if $W_{\texttt{alg}} \geq \frac{1}{\alpha} W_{\texttt{opt}}$ for all possible instances, then we say $\texttt{alg}$ is α-competitive, where $\alpha \geq 1$ and the closer to 1 the better.

We make the following mild assumptions throughout the paper. First, each customer's demand is much smaller compared to the total resource capacity. That is, after normalization, the demand of each customer $r_{i,k}^b$ is much smaller than 1. This is a common assumption in online auctions and pricing algorithm design, e.g., [3], since it allows us to focus on the online nature of our problem with mathematical convenience. Meanwhile, in many real-world large-scale systems, this assumption naturally holds. Second, we assume that the customers' valuation-to-demand ratio is upper bounded by $\overline{p}$, i.e.,

$$\max_{\forall i,b,k,r_{i,k}^b\neq 0} \left\{ \frac{v_i^b}{r_{i,k}^b} \right\} \leq \overline{p}. \tag{3}$$

Note that $\overline{p}$ can be interpreted as the maximum price that customers are willing to accept for purchasing a single unit of resource. We make this assumption to eliminate those rare cases with extremely high valuations.

[1]The cost function $f(\cdot)$ is allowed to be different among different resource types (i.e., $f_k(\cdot)$). Here we consider a simplified setting due to space limitation.

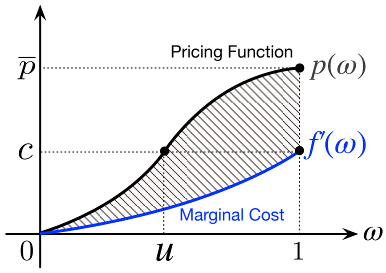

Figure 1: The pricing function $p(\omega)$ and the marginal supply cost $f'(\omega)$. The design variable $u \in (0,1)$ is a threshold such that $p(u) = c$, where $c \triangleq f'(1)$ is the maximum marginal cost when the resource utilization reaches the capacity limit 1.

3. MAJOR RESULTS

We propose an online posted-pricing mechanism PPM in Algorithm 1. PPM takes $(f, \overline{p})$ and pricing function $p(\cdot)$ as inputs. At each round when there is a new arrival of customer $i \in \mathcal{I}$, PPM offers her the current price $p_k = p(w_k)$ for resource type k based on the current total resource utilization w_k. If customer i decides to purchase a bundle of resources, say $b_* \in \mathcal{B}_i$, then PPM will charge this customer with the payment calculated in line 4 and update the total resource utilization level w_k in line 5.

Algorithm 1: Posted-Price Mechanism (PPM)

1: **Input:** $(f, \overline{p})$ and $p(\cdot)$, and initialize $w_k = 0$ for all k
2: **while** a new customer i arrives **do**
3: Offer resource type k at price $p_k = p(w_k)$.
4: Customer chooses bundle b_* and pays $\sum_{k\in\mathcal{K}} p_k r_{i,k}^{b_*}$.
5: Service provider updates the resource utilization by $w_k = w_k + r_{i,k}^{b_*}, \forall k \in \mathcal{K}$.
6: **end while**

To facilitate an online implementation, $p_k = p(w_k)$ is calculated based on the current resource consumption w_k only (i.e., casual information only). Fig. 1 illustrates the relationship between $p(\cdot)$ and the marginal cost function $f'(\cdot)$. As can be seen, $p(\cdot)$ is always lower bounded by $f'(\cdot)$ and is strictly increasing w.r.t. the total resource allocated. There is a special threshold $u \in (0,1)$, such that $p(u) = f'(1) \triangleq c$, i.e., *$u$ is the resource utilization level when the posted price equals the maximum marginal price.* Intuitively, a smaller u indicates a more aggressive pricing strategy, and a basic challenge is how to design $p(\cdot)$ in a strategic way such that a good balance between aggressiveness and conservativeness can be achieved. Below we present the major results of this paper.

Theorem 1 (Lower Bound). *Given the cost function f and $\overline{p} > c$, suppose the pricing function satisfies $p(u) = c$, where $u \in (0, 1)$, then the best-possible competitive ratio that* PPM *can achieve is $\underline{\alpha}(u)$, which is given by*

$$\underline{\alpha}(u) = \alpha_\gamma(u) \cdot \mathbb{I}_{\{u\in(0,u_\gamma)\}} + \alpha_\gamma^{\inf} \cdot \mathbb{I}_{\{u\in[u_\gamma,1]\}}, \tag{4}$$

where $\alpha_\gamma(u)$, $\alpha_\gamma^{\inf}$, and u_γ are given by

$$\alpha_\gamma(u) \triangleq \frac{\gamma - 1}{u - u^\gamma}, \alpha_\gamma^{\inf} \triangleq \gamma^{\gamma/(\gamma-1)}, u_\gamma \triangleq \left(\frac{1}{\gamma}\right)^{\frac{1}{\gamma-1}}, \tag{5}$$

and indicator function $\mathbb{I}_{\{A\}}$ equals 1 if A is true, and 0 otherwise.

Theorem 1 provides a lower bound of α for any pricing function indexed by u. Note that $\underline{\alpha}(u)$ is non-increasing in $u \in (0, 1)$ and achieves its infimum $\alpha_\gamma^{\inf}$ when $u \in [u_\gamma, 1)$, where $\alpha_\gamma^{\inf}$ depends on γ only. Based on Theorem 1, finding the optimal pricing function is equivalent to finding the optimal threshold u_* so that the minimum competitive ratio $\underline{\alpha}(u_*)$ can be achieved. Below we give a proposition to show the existence of such optimal threshold u_*.

Proposition 2. *Given the cost function f and $\overline{p} > c$, there exists a unique threshold $u_* \in (0, 1)$ that satisfies*

$$\frac{1 - Q_{\gamma-1}(u_*, \underline{\alpha}(u_*))}{\exp(u_* \cdot \underline{\alpha}(u_*))} = \frac{\overline{p}/c - Q_{\gamma-1}(1, \underline{\alpha}(u_*))}{\exp(\underline{\alpha}(u_*))}, \tag{6}$$

where $Q_{\gamma-1}(\omega, \alpha)$ is a polynomial in degree $\gamma - 1$, given by

$$Q_{\gamma-1}(\omega, \alpha) = \sum_{n=1}^{\gamma} z_n \omega^{\gamma-n}, \tag{7}$$

and z_n is given by

$$z_1 = 1, z_n = \frac{z_{n-1}(\gamma - n + 1)}{\alpha}, n = \{2, \cdots, \gamma\}. \tag{8}$$

Note that both sides of Eq. (6) are functions of the single variable u_*, and thus the unique u_* can be found by various numerical methods such as bisection search. We emphasize that both $Q_{\gamma-1}(\omega, \alpha)$ and Eq. (6) are derived by solving a group of first-order ordinary differential equations with boundary conditions. Meanwhile, based on whether u_* is within $(0, u_\gamma]$ or $(u_\gamma, 1)$, the optimal pricing functions to be designed in Theorem 3 have different forms.

Note that the calculation of u_* depends on the value of $\overline{p}$. For each given $\overline{p} \geq c$, let us define the unique u_* that satisfies Eq. (6) as a function of $\overline{p}$ as follows:

$$u_* \triangleq \Lambda(\overline{p}), \forall \overline{p} \in (c, +\infty). \tag{9}$$

Based on Proposition 2, below we give our optimal pricing function design in Theorem 3.

Theorem 3 (Optimal Design). *Let us define p_γ as follows*

$$p_\gamma \triangleq cQ_{\gamma-1}\big(1, \alpha_\gamma^{\inf}\big) + c\big(1 - Q_{\gamma-1}\big(u_\gamma, \alpha_\gamma^{\inf}\big)\big)e^{\alpha_\gamma^{\inf}(1-u_\gamma)}.$$

The optimal pricing function is determined as follows:

- *If $\overline{p} \in (0, c]$,* PPM *achieves an optimal competitive ratio of $\alpha_\gamma^{\inf}$ if $p(\omega) = \rho\gamma^2\omega^{\gamma-1}$. Meanwhile, there exist infinitely-many pricing functions such that* PPM *is $\alpha_\gamma^{\inf}$-competitive.*
- *If $\overline{p} \in (c, p_\gamma)$, for each $q \in [\overline{p}, p_\gamma)$, there exists a pricing function $p(\cdot)$ with threshold $u_* = \Lambda(q) \in (u_\gamma, 1)$ so that* PPM *achieves an optimal competitive ratio of $\alpha_\gamma^{\inf}$, where $p(\cdot)$ is given by*

$$p(\omega) = \begin{cases} c\varphi^{\gamma-1}, & \text{if } \omega \in [0, u_*], \\ \frac{q - cQ_{\gamma-1}\left(1,\alpha_\gamma^{\inf}\right)}{\exp\left(\alpha_\gamma^{\inf}(1-\omega)\right)} + cQ_{\gamma-1}\big(\omega, \alpha_\gamma^{\inf}\big), & \text{if } \omega \in (u_*, 1], \end{cases}$$

where φ is the unique root to the following equation in variable $\varphi \in (0, 1)$ for any given $\omega \in (0, u_*]$:

$$\int_{\frac{1}{u_*}}^{\frac{\varphi}{\omega}} \frac{\eta^{\gamma-1}}{\eta^\gamma - \frac{\alpha_\gamma^{\inf}}{\gamma-1}\eta^{\gamma-1} + \frac{\alpha_\gamma^{\inf}}{\gamma-1}} d\eta = \ln\left(\frac{u_*}{\omega}\right). \tag{10}$$

In this scenario, $p(1) = q$ and $u_* = \Lambda(q) \in (u_\gamma, 1)$ is calculated based on Eq. (6) by replacing $\overline{p}$ with q.

- *If $\overline{p} \geq p_\gamma$, there exists a unique pricing function p with threshold $u_* = \Lambda(\overline{p}) \in (0, u_\gamma]$ so that* PPM *achieves the optimal competitive ratio of $\alpha_\gamma(u_*)$, where $p(\cdot)$ is given by*

$$p(\omega) = \begin{cases} c\omega^{\gamma-1}/u_*^{\gamma-1}, & \text{if } \omega \in [0, u_*], \\ \frac{\overline{p} - cQ_{\gamma-1}(1, \underline{\alpha}(u_*))}{\exp(\underline{\alpha}(u_*)(1-\omega))} + cQ_{\gamma-1}(\omega, \underline{\alpha}(u_*)), & \text{if } \omega \in (u_*, 1]. \end{cases}$$

In this scenario, $p(1) = \overline{p}$. In particular, when $\overline{p} = p_\gamma$, we have $u_* = u_\gamma$.

Theorem 3 summarizes our optimal pricing function design in three different scenarios, where in the first two scenarios the optimal pricing functions are not unique, while in the last scenario there exists a unique pricing function so that PPM achieves the optimal competitive ratio of $\alpha_\gamma(u_*)$. An interesting results revealed by Theorem 3 is that, when $\overline{p}$ is below p_γ, it is not necessary to know the exact value of $\overline{p}$ to achieve the best-possible competitive ratio of $\alpha_\gamma^{\inf}$. Note that $\overline{p}$ represents the uncertainty or variance of users' valuation, and thus Theorem 3 indicates that, when the uncertainty level is below a certain threshold (i.e, $\overline{p} \leq p_\gamma$), the best-possible competitive ratio can be achieved by PPM regardless of having the exact knowledge of $\overline{p}$.

4. CONCLUSION

We studied a general online resource allocation problem in the presence of increasing marginal supply cost and capacity limit. We proposed an optimal online posted-pricing mechanism that achieves the best-possible competitive ratio under different scenarios. Our proposed online mechanism generalizes the design in [3] to the capacity-limited case, and improves the results in both [2] and [3].

5. REFERENCES

[1] A. Mehta, "Online matching and ad allocation," *Found. Trends Theor. Comput. Sci.*, vol. 8, no. 4, pp. 265–368, Oct. 2013.

[2] A. Blum, A. Gupta, Y. Mansour, and A. Sharma, "Welfare and profit maximization with production costs," in *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2011)*, Washington, DC, USA, 2011.

[3] Z. Huang and A. Kim, "Welfare maximization with production costs: A primal dual approach," in *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)*, Philadelphia, PA, USA, 2015, pp. 59–72.