

On the complexity of a bundle pricing problem

Alexander Grigoriev · Joyce van Loon · Marc Uetz

Received: 30 September 2010 / Revised: 24 January 2011 / Published online: 11 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We consider the problem of pricing items in order to maximize the revenue obtainable from a set of single minded customers. We relate the tractability of the problem to structural properties of customers' valuations: the problem admits an efficient approximation algorithm, parameterized along the inhomogeneity of the valuations.

Keywords Pricing problems · Computational complexity · Approximation algorithm

MSC classification (2000) 90B50 · 68Q25 · 68W25

1 Introduction

1.1 Problem definition

Let $I = \{1, \dots, m\}$ represent a set of items for sale and let $J = \{1, \dots, n\}$ represent a set of potential customers. Every customer $j \in J$ requests a subset of items, denoted $I_j \subseteq I$. We refer to these subsets as *bundles*. Customers are single minded, which refers to the fact that they are interested in their particular bundle only. The

A. Grigoriev (✉) · J. van Loon
Quantitative Economics, Maastricht University SBE, P.O. Box 616, 6200 MD, Maastricht,
The Netherlands
e-mail: a.grigoriev@maastrichtuniversity.nl

J. van Loon
e-mail: j.vanloon@maastrichtuniversity.nl

M. Uetz
Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands
e-mail: m.uetz@utwente.nl

valuation v_j for each bundle I_j , $j \in J$, is publicly known. This is reasonable when assuming customers' rationality and a competitive market environment: any customer can observe the publicly known prices for her bundle at all companies in the market, and then, behaving rationally, the customer defines her valuation being the cheapest market price for her bundle. We assume $v_j > 0$, $j \in J$, for otherwise the customers having non-positive valuations can be deleted from the instance. We assume the items are available in unlimited supply, that is to say, we deal with digital goods or services. Let p_i be the price for item $i \in I$. We refer to the set $W = W(p) = \{j \in J \mid \sum_{i \in I_j} p_i \leq v_j\}$ as the set of *winners*. The bundle pricing problem asks for a vector of item prices $p = (p_1, \dots, p_m)$ such that the total revenue $\Pi(p) = \sum_{j \in W(p)} \sum_{i \in I_j} p_i$ is maximal. Let us denote by Π the maximal revenue that can be extracted from the given set of customers.

1.2 Related work

The bundle pricing problem was introduced in combinatorial optimization literature by [Guruswami et al. \(2005\)](#). They show that the problem is APX-hard. Later, [Demaine et al. \(2008\)](#) prove that the problem is inapproximable by a semi-logarithmic factor in the number of customers n . On the positive side, [Guruswami et al. \(2005\)](#) present a polynomial time $O(\log n + \log m)$ -approximation algorithm. [Hartline and Koltun \(2005\)](#) design near-linear and near-cubic time approximation schemes under the assumption that the number of distinct items m is constant. Under the monotonicity condition that the total price of any bundle does not exceed the total price of any bigger bundle, [Grigoriev et al. \(2008\)](#) show that the problem is still NP-hard but admits a polynomial time approximation scheme.

1.3 Our result

In this note we interpret customers' valuations in such a way that we come a step closer towards understanding the complexity of the problem. To start with, let us make the following definition.

Definition 1 For any instance of the bundle pricing problem, define $\bar{b}_j = v_j/|I_j|$ as the average (per item) valuation of customer j . Let $\bar{b}_{\min} = \min_{j \in J} \{\bar{b}_j\}$ and $\bar{b}_{\max} = \max_{j \in J} \{\bar{b}_j\}$. Define the *inhomogeneity* of valuations as $\alpha = \bar{b}_{\max}/\bar{b}_{\min}$.

Notice that $\alpha \geq 1$, and that the problem becomes trivial as soon as the valuations are homogeneous (that is, $\alpha = 1$ and $\bar{b}_j =: \bar{b}$ for all j). In this case, setting the price for each item $i \in I$ uniformly at $p_i = \bar{b}$, we obtain the optimal solution.

In contrast to the trivially solvable homogeneous case, the problem with inhomogeneity of valuations is NP-hard. While this does not sound very surprising, the main point is that the NP-hardness holds even if the inhomogeneity α is bounded from above by any constant $1 + \varepsilon$. In some sense, we thereby delineate the borderline between triviality and NP-hardness for the bundle pricing problem.

For the fact that the bundle pricing problem is NP-hard even for inhomogeneity arbitrarily close to 1, consider the NP-hardness reduction from INDEPENDENT SET to

the bundle pricing problem presented in Grigoriev et al. (2008). In this reduction, all average valuations of the bundles are at least M and at most $M + 1$, where M is a chosen large number. The NP-hardness result for $\alpha \leq 1 + \varepsilon$, $\varepsilon > 0$, follows straightforwardly. Moreover, the reduction works even under stronger restrictions on customers' valuations, namely *monotonicity*: $v_j \leq v_k$ for any $j, k \in J$ such that $I_j \subset I_k$. Thus, we proved the following theorem.

Theorem 1 *The bundle pricing problem is strongly NP-hard even if inhomogeneity $\alpha \leq 1 + \varepsilon$ for any $\varepsilon > 0$, and if the valuations are monotone.*

In the next section we present a parametric approximation algorithm for the bundle pricing problem that complements the NP-hardness result. The proposed $O(nm)$ -time algorithm has performance guarantee $1 + \ln \alpha + \varepsilon$, for any $\varepsilon > 0$. Notice that this is a constant-factor approximation algorithm as soon as the inhomogeneity α of valuations is bounded by some constant, and the semi-logarithmic inapproximability result of Demaine et al. (2008) is not longer valid. We believe that a constant bound on α is not unreasonable in practical applications.

2 An $O(\ln \alpha)$ -approximation algorithm

The idea of the approximation algorithm is as follows. We partition the set of customers J into $O(\ln \alpha)$ subsets S_1, \dots, S_K , such that in each subset any two customers have average valuations different from each other by at most a constant factor $\delta > 1$. Denote by Π_k the maximum revenue for the bundle pricing problem restricted to the set of customers S_k (referred to as S_k -restricted problem). Then $\sum_{k=1}^K \Pi_k$ is clearly an upper bound for the optimum Π of the original problem. Therefore, the highest maximum revenue $\max_{k=1, \dots, K} \Pi_k$ over all restricted problems is at least Π/K . Next, from the fact that the inhomogeneity of the average valuations in S_k is bounded by at most factor of δ , we derive that for the S_k -restricted problem there exists a price vector generating revenue at least Π_k/δ . Thus, taking the price vector yielding the highest revenue over all restricted problems, we generate a revenue at least $\Pi/\delta K$. Finally, we optimize the performance guarantee over parameters K and δ .

To partition the set of customers J into subsets S_1, \dots, S_K , we use straightforward geometric scaling of average valuations. Namely, we assign customer $j \in J$ to subset S_1 if $\bar{b}_j \leq \delta \bar{b}_{\min}$, and we assign her to subset S_k , $k \geq 2$, if $\delta^{k-1} \bar{b}_{\min} < \bar{b}_j \leq \delta^k \bar{b}_{\min}$, where $\delta > 1$ is to be defined later.

By definition of the inhomogeneity α , we have $\bar{b}_{\max} = \alpha \bar{b}_{\min}$. Let K be the largest integer such that $\bar{b}_{\max} \geq \delta^{K-1} \bar{b}_{\min}$ or equivalently $\alpha \geq \delta^{K-1}$. Thus, $K \leq 1 + \log_\delta \alpha = 1 + \ln \alpha / \ln \delta$, and we derived the first ingredient of the approximation algorithm, formulated in the following lemma.

Lemma 1 *For any $\delta > 1$, the number of subsets K is at most $1 + \ln \alpha / \ln \delta$.*

Second, we show that there is a solution to the S_k -restricted problem such that (i) the set of winners $W = S_k$; and (ii) the revenue generated in this solution is at least Π_k/δ . Consider the price vector $p^k = (p_1^k, \dots, p_m^k)$ where for all $i \in I$ we have the uniform price $p_i^k = \min\{\bar{b}_j \mid j \in S_k\}$. Now, consider a customer $j \in S_k$. By definition

of price vector p^k , the price of bundle I_j is $\sum_{i \in I_j} p_i^k \leq \sum_{i \in I_j} \bar{b}_j = v_j$, and therefore $j \in W$. By definition of set S_k , $\max_{j \in S_k} \bar{b}_j / \min_{j \in S_k} \bar{b}_j \leq \delta$, that yields a revenue at least Π_k / δ . Thus, we proved the following lemma.

Lemma 2 *In the S_k -restricted problem, price vector p^k yields a revenue at least Π_k / δ .*

Now, we are ready to present our first approximation result.

Theorem 2 *The bundle pricing problem admits an $e(2 + \ln \alpha)$ -approximation algorithm with computation time $O(nm)$.*

Proof The combination of Lemmas 1 and 2 immediately implies that the revenue generated by the best price vector from $\{p^k \mid k = 1, \dots, K\}$ is at least $\Pi / \delta (1 + \frac{\ln \alpha}{\ln \delta})$,

which is maximized for $\delta = e \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\ln \alpha}} \right)^{-1}$. Then, the approximation guarantee follows from the following two facts. Firstly, $\delta = e \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\ln \alpha}} \right)^{-1}$ is monotone increasing in α and approaches e when α tends to infinity. Secondly, for any $\alpha > 1$, if

$\delta = e \left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\ln \alpha}} \right)^{-1}$, it is not hard to see that $\frac{\ln \alpha}{\ln \delta} \leq 1 + \ln \alpha$.

We arrive at the computation time as follows. Firstly, we conventionally omit the dependance on the input length of the numbers v_j , which are treated as constants. Doing so, because $K \leq 2 + \ln \alpha \leq 2 + \ln \max_{j \in J} v_j$, also parameter K , the size of the partition of customers, does not appear in the analysis. Hence, computing the average valuations for all customers takes $O(nm)$ time. Moreover, the assignment of customers to subsets and finding the uniform price vectors can be done in $O(n)$ time. Computing the revenues also takes $O(n)$ time. Therefore, the total computation time is $O(nm)$. □

3 Improved analysis and asymptotic tightness

There are several directions for improvement of the obtained approximate solution to the bundle pricing problem. First, instead of the constructed uniform price vectors $p^k, k = 1, \dots, K$, we can use price vectors maximizing the revenue in the S_k -restricted problems, with given set of winners $W = S_k$. Notice that, for any set of winners $W \subseteq J$, the price vector maximizing the revenue obtained from W can be found in polynomial time by solving a simple linear program; see (Grigoriev et al. 2009; Guruswami et al. 2005). Unfortunately, this approach does not necessarily lead to any provable improvement of the performance guarantee.

The following approach allows us to slightly improve the performance guarantee; it is simply based on a more careful analysis. By construction of the partition of J , for any two subsets S_k and $S_{k'}$, $k \leq k'$, the average valuation of any customer from S_k is at most the average valuation of a customer from $S_{k'}$. Therefore, for any $k = 1, \dots, K$, and for all $k' \geq k$, if $S_k \subseteq W$, then $S_{k'} \subseteq W$ as well. By definition of the subsets, the maximum average valuation in set S_{k+1} is at most δ times the maximum average valuation in set S_k . Thus, we have that the revenue generated by price vector p^k applied to the set of customers J is at least

$$R_k = \frac{1}{\delta} \Pi_k + \frac{1}{\delta^2} \Pi_{k+1} + \dots + \frac{1}{\delta^{K-k+1}} \Pi_K, \quad \forall k = 1, \dots, K.$$

These equalities can be equivalently represented by the following recurrent formulas

$$R_k = \frac{1}{\delta} \Pi_k + \frac{1}{\delta} R_{k+1}, \quad \forall k = 1, \dots, K - 1; \tag{1}$$

$$R_K = \frac{1}{\delta} \Pi_K. \tag{2}$$

Summing up all Eqs. (1) and (2) and dividing both sides by K , we derive

$$\bar{R} = \frac{1}{K} \sum_{k=1}^K R_k = \frac{1}{K\delta} \sum_{k=1}^K \Pi_k + \frac{1}{K\delta} \sum_{k=1}^K R_k - \frac{1}{K\delta} R_1.$$

Let $R_1 = \phi \bar{R}$. Since $\sum_{k=1}^K \Pi_k \geq \Pi$, we derive

$$\bar{R} \geq \frac{\Pi}{K(\delta - 1) + \phi}.$$

Taking the maximum revenue over all price vectors $p^k, k = 1, \dots, K$, we obtain

$$\max_{k=1, \dots, K} R_k \geq \max\{R_1, \bar{R}\} \geq \max\left\{ \frac{\phi \Pi}{K(\delta - 1) + \phi}, \frac{\Pi}{K(\delta - 1) + \phi} \right\},$$

that is minimized with $\phi = 1$, yielding

$$\max_{k=1, \dots, K} R_k \geq \frac{\Pi}{\delta \left(1 + \frac{\ln \alpha}{\ln \delta} \right) - \frac{\ln \alpha}{\ln \delta}}.$$

Note that $\delta \left(1 + \frac{\ln \alpha}{\ln \delta} \right) - \frac{\ln \alpha}{\ln \delta} < \delta \ln \alpha + \delta$ for any $\delta \geq 1$. Given $\varepsilon > 0$, let $\delta = 1 + \varepsilon / (\ln \alpha + 1)$. Then,

$$\delta \ln \alpha + \delta = \left(1 + \frac{\varepsilon}{\ln \alpha + 1} \right) \ln \alpha + \left(1 + \frac{\varepsilon}{\ln \alpha + 1} \right) = 1 + \ln \alpha + \varepsilon,$$

and we arrive at the following theorem.

Theorem 3 *For any $\varepsilon > 0$, the bundle pricing problem admits an $(1 + \ln \alpha + \varepsilon)$ -approximation algorithm with computation time $O(nm)$, and this performance guarantee is asymptotically tight.*

We are only left to show that the performance guarantee established in Theorem 3 is asymptotically tight. To see this, consider the following instance, which is an adaptation of Example 1 from [Bouhtou et al. \(2007\)](#). Assume, there are M types of items.

Let $q > 1$ be a real number, and let type $i \in \{1, \dots, M\}$ consists of $q^i - q^{i-1}$ distinct items. For simplicity of presentation let us not bother with the fact that the values $q^i - q^{i-1}$ might be fractional; any rounding would do the job, but only overcomplicates the exposition. Observe that the number of items is $m = \sum_{i=1}^M (q^i - q^{i-1}) = q^M - 1$. There are $n = m$ customers, and every customer requests a single dedicated and exclusive item. All customers requesting an item of type $i \in \{1, \dots, M\}$ have valuation $v_i = q^{2M-i}$ for that item. Thus, $\bar{b}_{\min} = q^M$, $\bar{b}_{\max} = q^{2M-1}$, and $\alpha = q^{M-1}$.

In the optimal solution all items are priced at the valuations of the corresponding customers, yielding a total revenue $\sum_{i=1}^M q^{2M-i} (q^i - q^{i-1}) = Mq^{2M-1}(q - 1)$. Now notice that the above described approximation algorithm always produces a uniform price vector, i.e., the prices for all items are identical. In the given example, the uniform pricing capturing the first j types of items generates the total revenue $\sum_{i=1}^j q^{2M-j} (q^i - q^{i-1}) = q^{2M} - q^{2M-j}$, which is maximized with $j = M$. Hence, the optimal uniform pricing generates revenue $q^{2M} - q^M$, which is an upper bound for what the algorithm can achieve. By straightforward calculations, the performance guarantee for this instance can be bounded from below by $LB = M \frac{q-1}{q}$ and, by Theorem 3, it is bounded from above by $UB = 1 + \varepsilon + (M - 1) \ln q$. Now, the tightness of the bound in Theorem 3 follows from the fact that, for any $x > 0$, there exists a choice for M and q such that the ratio UB/LB is at most $1 + x$. Essentially, we let M tend to infinity and q approaches 1 from above.

Acknowledgments We thank two anonymous referees for their very useful suggestions. Joyce van Loon acknowledges support by METEOR, the Maastricht Research School of Economics of Technology and Organizations. A preliminary version with parts of this work appeared in the conference proceedings Grigoriev et al. (2010).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bouhtou M, Grigoriev A, van Hoesel S, van der Kraaij AF, Spieksma FCR, Uetz M (2007) Pricing bridges to cross a river. *Naval Res Logist* 54:411–420
- Demaine ED, Feige U, Hajiaghayi MT, Salavatipour MR (2008) Combination can be hard: approximability of the unique coverage problem. *SIAM J Comput* 38:1464–1483
- Grigoriev A, van Loon J, Sitters R, Uetz M (2009) Optimal pricing of capacitated networks. *Networks* 53:79–87
- Grigoriev A, van Loon J, Sviridenko M, Uetz M, Vredeveld T (2008) Optimal bundle pricing with monotonicity constraint. *Oper Res Lett* 36:609–614
- Grigoriev A, van Loon J, Uetz M (2010) On the complexity of the highway pricing problem. In: van Leeuwen J, Muscholl A, Peleg D, Pokorný J, Rumpe B (eds) *SOFSEM 2010: theory and practice of computer science. Lecture notes in computer science*, vol 5901. Springer, Berlin, pp 465–476
- Guruswami V, Hartline JD, Karlin AR, Kempe D, Kenyon C, McSherry F (2005) On profit-maximizing envy-free pricing. In: *Proceedings of 16th annual ACM-SIAM symposium on discrete algorithms, SODA 2005*, ACM Press, USA, pp 1164–1173
- Hartline JD, Koltun V (2005) Near-optimal pricing in near-linear time. In: Dehne FKHA, López-Ortiz A, Sack J-R (eds) *Algorithms and data structures—WADS 2005. Lecture notes in computer sciences*, vol 3608. Springer, Berlin, pp 422–431