Alexander Grigoriev, Joyce van Loon, Marc Uetz

On the Complexity of the Highway Pricing
Problem

RM/08/030

METE☾R

On the Complexity of the
Highway Pricing Problem

Alexander Grigoriev[1], Joyce van Loon[1⋆], and Marc Uetz[2]

[1] Maastricht University, Quantitative Economics,
P.O.Box 616, NL–6200 MD Maastricht, The Netherlands,
{a.grigoriev,j.vanloon}@ke.unimaas.nl
[2] University of Twente, Applied Mathematics,
P.O. Box 217, NL–7500 AE Enschede, The Netherlands,
m.uetz@utwente.nl

Abstract. The highway pricing problem asks for prices to be determined for segments of a single highway such as to maximize the revenue obtainable from a given set of customers with known valuations. The problem is (weakly) NP-hard and a recent quasi-PTAS suggests that a PTAS might be in reach. Yet, so far it has resisted any attempt for constant-factor approximation algorithms. We relate the tractability of the problem to structural properties of customers' valuations. We show that the problem becomes NP-hard as soon as the average valuations of customers are not homogeneous, even under further restrictions such as monotonicity. Moreover, we derive an efficient approximation algorithm, parameterized along the inhomogeneity of customers' valuations. Finally, we discuss extensions of our results that go beyond the highway pricing problem.

Keywords: Pricing problems, highway pricing problem, computational complexity, approximation algorithm.

1 Introduction

We consider the *highway pricing problem*, introduced by Guruswami et al. [8]. The problem is motivated by determining revenue-maximizing tolls to be charged for segments of a highway. The highway is thought of as a simple path, and capacity is considered unlimited. There are potential customers, each of them requesting to travel a sub-path of the highway, and the maximal valuation for utilizing the requested sub-path is considered public knowledge. The objective is to find prices to be charged for the segments of the highway so as to maximize the total revenue obtained by the customers.

More formally, let $I = \{1, \ldots, m\}$ represent the highway segments, and regard them as consecutive edges on a simple path. Let $J = \{1, \ldots, n\}$ denote the set of potential customers. Every customer $j \in J$ requests a sub-path of the highway, denoted $I_j \subseteq I$, and we assume that each I_j is of the form $I_j = \{k, k+1, \ldots, \ell\}$, $k \leq \ell$. The valuation v_j for traveling sub-path I_j is publicly known. This is quite reasonable when assuming that the valuation is a monetary expression for the time saving that can be realized by using the highway instead of the next-fastest alternative route. We assume $v_j > 0$, for otherwise that customer can be deleted from the instance. Given a vector of prices $p = (p_1, \ldots, p_m)$, containing one price for each highway segment, denote by $W = \{j \in J \mid \sum_{i \in I_j} p_i \leq v_j\}$ the set of *winners*.

Definition 1. *The highway pricing problem asks for a vector of prices $(p_1, \ldots, p_m)$, one for each segment of the highway, such that the total revenue $\sum_{j \in W} \sum_{i \in I_j} p_i$ extracted from the set W of winners is maximal.*

1.1 Related Work

The complexity of the highway problem was left open in [8], but it was shown (weakly) NP-hard by Briest and Krysta [2]. Guruswami et al. [8] propose a polynomial time dynamic programming algorithm when the valuations are bounded by a constant, and a pseudo-polynomial time dynamic programming algorithm when the lengths of the sub-paths are bounded by a constant. Note that the problem can be interpreted as a bilevel linear program, and if either the price vector or the set of winners is known, the problem is polynomially solvable [6, 8], even under the requirement of integral prices. Balcan and Blum [1] derive an $O(\log m)$-approximation algorithm for the highway problem, improving upon the previous $O(\log m + \log n)$-approximation of Guruswami et al. [8], where m is the number of highway segments and n is the number of customers. Under the monotonicity condition that the total price of any given path is no more than the total price of a longer path, Grigoriev et al. [7] show that a $O(\log B)$-approximation exists, where B is an upper bound on the valuations. Furthermore, Grigoriev et al. [6] derive an FPTAS, assuming that the maximum capacity of any segment of the highway is bounded by a constant. Finally, Elbassioni et al. [4] present a quasi-polynomial time approximation scheme for both the capacitated and uncapacitated version of the problem, thereby suggesting that a PTAS is likely to exist.

1.2 Motivation & Results

Intrigued by the gap between (weak) NP-hardness on the one hand, and only logarithmic polynomial-time approximation algorithms on the other hand, in this paper we interpret customers' valuations in such a way that we come a step closer towards understanding this complexity gap. To start with, let us make the following definition, illustrated also by Example 1 below.

Definition 2 (Inhomogeneity of valuations). *For any instance of the highway pricing problem, define $\bar{v}_j = v_j/|I_j|$ as the average (per segment) valuation of customer j, and define the inhomogeneity of valuations as*

$$\alpha = \max_{j,k \in J} \left\{ \frac{\bar{v}_j}{\bar{v}_k} \right\} .$$

Example 1. Figure 1 shows an example with three segments, $I = \{1, 2, 3\}$, and six customer requests $J = \{1, \ldots, 6\}$. The left part of this figure shows the underlying highway with its alternative roads and distances, and the right part shows the corresponding instance of the highway problem. The valuation for traveling from the start of segment k until the end of segment ℓ is denoted $v_{k,\ell}$. This instance has inhomogeneity $\alpha = 2$; comparing the valuations for $\{1, 2, 3\}$ and $\{2\}$.

$v_{1,1} = 7$
$v_{2,2} = 10$
$v_{3,3} = 8$
$v_{1,2} = 12$
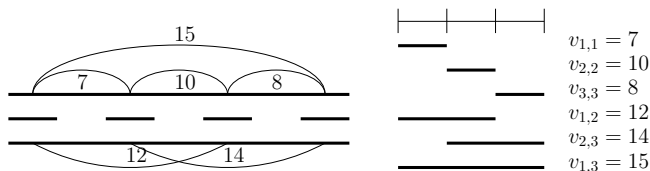$v_{2,3} = 14$
$v_{1,3} = 15$

Fig. 1. An instance of the highway pricing problem.

Notice that $\alpha \geq 1$, and that the problem becomes trivial as soon as the valuations are *homogeneous* (that is, $\alpha = 1$), since this corresponds to the case where all customers' valuations per segment are identical; see Section 2.

Our first result is to show that, in contrast to the trivially solvable homogeneous case, the problem with inhomogeneity of valuations is (weakly) NP-hard. While this does not sound very surprising, the main point is that this NP-hardness result 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 highway pricing problem.

Furthermore, the NP-hardness result remains true even if we impose further restrictions on customers' valuations, such as monotonicity, that is,

$$v_j \ \leq \ v_k \qquad \text{for all } I_j \subseteq I_k ,$$

and monotonicity of average valuations, that is,

$$\frac{v_j}{|I_j|} \ \geq (\leq, \text{resp.}) \ \frac{v_k}{|I_k|} \qquad \text{for all } I_j \subseteq I_k .$$

Our second result is a parametric approximation algorithm for the highway pricing problem that complements the NP-hardness result. The proposed algorithm has performance guarantee $\mathrm{O}(\log \alpha)$ and computation time

3

$O(n(\log n + m))$, where the constant hidden in the O-notation of the performance bound is not more than e. More specifically, it is easy to see that an α-approximation exists, and for large values of α we show how to improve this bound to $1 + \ln \alpha + \varepsilon$ for any $\varepsilon > 0$. Notice that this is a constant-factor approximation algorithm as soon as the inhomogeneity α of customers' valuations is bounded by some constant. We believe that such a constant bound is not unreasonable in practical applications, and note that $\alpha \leq m$ for the case of monotone and decreasing average valuations.

Finally, we briefly comment on the fact that the $O(\log \alpha)$ approximation result even holds for the more general bundle pricing problem where customers are interested in arbitrary bundles instead of sub-paths only. In this context, notice that if there exists any constant upper bound on the inhomogeneity α then the semi-logarithmic inapproximability result of Demaine et al. [3] for that problem is not longer valid. For that problem we also derive a (strong) NP-hardness result, again for any constant upper bound on the inhomogeneity of the valuations.

2   Complexity of the highway problem with inhomogeneous valuations

We start with the short argument that the highway problem with homogeneous average valuations is trivially solvable: consider the average valuation $\bar{v}$, which is, by homogeneity, the same for each customer, and define the price $p_i = \bar{v}$ for every segment $i \in I$. Clearly, each customer contributes her entire valuation to the revenue, and the obtained solution is optimal.

Surprisingly enough, even if we allow only arbitrarily small deviations of homogeneous valuations, the highway problem becomes intractable. The following theorem shows that the problem with inhomogeneous valuations remains NP-hard even in further restricted settings.

Theorem 1. *The highway problem is NP-hard even when restricted to the instances satisfying the following conditions:*

1. *the inhomogeneity $\alpha \leq 1 + \varepsilon$ where ε is an arbitrary positive constant;*
2. *customers valuations are monotone, i.e., $v_j \leq v_k$ for any $j, k \in J$ such that $I_j \subseteq I_k$;*
3. *customers average valuations are monotone decreasing, i.e., $\bar{v}_k \leq \bar{v}_j$ for any $j, k \in J$ such that $I_j \subseteq I_k$.*

Proof. The reduction is from the PARTITION problem, and extends an idea by Briest and Krysta [2]. PARTITION: Given integers $a_1, \ldots, a_{2L}$ and A, does there exist a set $S \subseteq \{1, \ldots, 2L\}$ such that $\sum_{\ell \in S} a_\ell = \sum_{\ell \notin S} a_\ell = A$? This problem is known to be NP-hard, even under the additional restriction that $|S| = L$; see [5]. We may assume that $L > 3/\varepsilon$, for otherwise PARTITION is solvable in polynomial time. Without loss of generality, we also assume that $0 \leq a_1 \leq \ldots \leq a_{2L}$ and

4

$a_\ell \leq A$ for all $\ell = 1, \ldots, 2L$. Let $a'_\ell = a_\ell + (4L + 2)A$ for all $\ell = 1, \ldots, 2L$, and $A' = (4L^2 + 2L + 1)A$. Note that $\sum_{\ell=1}^{2L} a'_\ell = 2A'$.

We now create an instance $\mathcal{H}$ of the highway problem with $7L + 3$ segments combined in *gadgets*. Gadget $\ell = 1, \ldots, 2L$ contains two segments, $i = 2\ell - 1$ and $i = 2\ell$. Each of these two segments are requested by $2L - 1$ customers with valuation a'_ℓ. The combination of two segments, $2\ell - 1$ and 2ℓ, is requested by one customer with valuation $(2 - \frac{1}{L})a'_\ell$. Finally, gadget $2L + 1$ contains $3L + 3$ segments, where the first three segments, $4L+1, 4L+2, 4L+3$, are requested by one customer with valuation $\frac{12}{4L+3}A'$ and the last $3L$ segments, $4L+4, \ldots, 7L+3$, are requested by 3 customers with valuation $\frac{12L}{4L+3}A'$. All segments in gadget $2L + 1$ are also requested by 3 customers with valuation $\frac{12L}{4L+3}A'$. There is one *big* customer, who requests the first $4L + 3$ items with valuation $(4 - \frac{1}{L})A'$. Instance $\mathcal{H}$ is displayed in Figure 2, where the number of customers interested in a sub-path is presented by d.
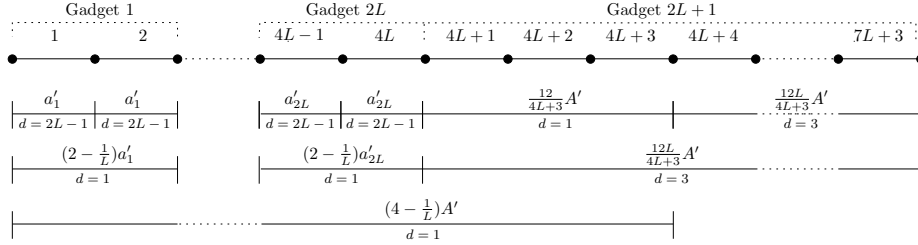

Fig. 2. Instance $\mathcal{H}$.

Though it requires a quite extensive case study, one can straightforwardly verify that conditions (1)-(3) of Theorem 1 are satisfied. For the first condition, we have that $\alpha = 1 + O(1/L)$ implying that it can be upper bounded by $1 + \varepsilon$ where ε is any positive constant.

Now we claim that there exists a feasible solution to PARTITION if and only if there is a feasible solution to instance $\mathcal{H}$ of the highway problem with a total revenue of at least $\left(8L + \frac{72L}{4L+3} - \frac{1}{L}\right)A'$.

($\Rightarrow$) Given a set $S \subseteq \{1, \ldots, 2L\}$ such that $\sum_{\ell \in S} a_\ell = \sum_{\ell \notin S} a_\ell = A$ and $|S| = L$. For all $\ell \in \{1, \ldots, 2L\}$, let $p_{2\ell-1} = p_{2\ell} = a'_\ell$ if $\ell \in S$ and $p_{2\ell-1} = p_{2\ell} = (1 - \frac{1}{2L})a'_\ell$ if $\ell \notin S$. Furthermore, we set $p_{4L+1} = \ldots = p_{7L+2} = 0$ and $p_{7L+3} = \frac{12L}{4L+3}A'$. Applying this price vector, the revenue without contribution of the big customer is equal to $(4L - 2)a'_\ell$ in each gadget $\ell = 1, \ldots, 2L$. The big customer contributes her entire valuation $(4 - \frac{1}{L})A'$. In gadget $2L+1$, the customer requesting segments $4L + 1, 4L + 2, 4L + 3$ gets this path for free. The other customers in this gadget contribute their respective valuations. The total revenue generated with this

pricing vector equals

$$(4L-2)\sum_{\ell=1}^{2L} a'_\ell + \left(4 - \frac{1}{L}\right)A' + 6\cdot\frac{12L}{4L+3}A' = \left(8L + \frac{72L}{4L+3} - \frac{1}{L}\right)A'.$$

($\Leftarrow$) Given is an optimal solution to instance $\mathcal{H}$ with a total revenue at least $(8L + \frac{72L}{4L+3} - \frac{1}{L})A'$. First, we observe that in such optimal solution, segments $4L+1, 4L+2, 4L+3$ are necessarily priced to 0 and the total price of the remaining segments in gadget $2L+1$ is $\frac{12L}{4L+3}A'$, yielding revenue $\frac{72L}{4L+3}A'$. To see this, we notice that the total demand on the first three segments in this gadget is 5 and on the latter $3L$ segments the demand is 6. Therefore, if the total price on the first three segments of gadget $2L+1$ is $0 < x \leq \frac{12}{4L+3}A'$, the total revenue obtained in the gadget is at most $\frac{72L}{4L+3} - x$, that is, we receive x from the big customer and at most $x + 3(\frac{12L}{4L+3}A' - x) + 3(\frac{12L}{4L+3}A')$ from the customers in gadget $2L+1$. The above suggested pricing does not decrease revenue generated in gadgets $1, \ldots, 2L$, and generates the total revenue in gadget $2L+1$ equal to $\frac{72L}{4L+3}$.

Second, in the optimal solution to the highway problem, there could be only two alternative pricing strategies in gadgets $\ell = 1, \ldots, 2L$: either $p_{2\ell-1} = p_{2\ell} = a'_\ell$ or $p_{2\ell-1} + p_{2\ell} = (2 - \frac{1}{L})a'_\ell$, where both prices do not exceed a'_ℓ. In both realizations, the contribution of the gadget (without big customer) to the total revenue is $(4L-2)a'_\ell$. Therefore, in the optimal solution to instance $\mathcal{H}$ with revenue at least $(8L + \frac{72L}{4L+3} - \frac{1}{L})A'$, the big customer must contribute her entire valuation. This amount is to be spent in the first $4L$ segments as the price of segments $4L+1, 4L+2$ and $4L+3$ is set to 0.

Define set $S = \{\ell \in \{1, \ldots, 2L\} : p_{2\ell-1} = p_{2\ell} = a'_\ell\}$. The payment of the big customer is $\sum_{\ell \in S} 2a'_\ell + \sum_{\ell \notin S}(2 - 1/L)a'_\ell$. As this must be equal to the valuation of the big customer, we have $\sum_{\ell \in S} a'_\ell = \sum_{\ell \notin S} a'_\ell = A'$ and consequently, $\sum_{\ell \in S} a_\ell = \sum_{\ell \notin S} a_\ell = A$. $\qquad\square$

3 $O(\log\alpha)$-approximation algorithm

The idea for the approximation algorithm is as follows. We partition the set of customers J into $O(\ln\alpha)$ subsets $S_1, \ldots, S_K$, 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 highway 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 $\max_{k=1,\ldots,K} \Pi_k$ 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 pricing vector yielding the highest revenue over all restricted problems, we generate a total 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 $S_1, \ldots, S_K$, we use the following recursive procedure running in K steps. At step $k = 1, \ldots, K$, we construct subset S_k. Consider the set of customers J_k not yet assigned to any of the subsets $S_1, \ldots, S_{k-1}$, assuming $J_1 = J$. Add all customers $j \in J_k$ to S_k for which $\bar{v}_j \leq \delta^k \bar{v}_{\min}$, where $\bar{v}_{\min} = \min_{j \in J}\{\bar{v}_j\}$ and $\delta > 1$ to be defined later. Set $J_{k+1} = J_k \setminus S_k$ and recurse on this set.

By definition of the inhomogeneity α, we have $\bar{v}_k \leq \alpha \bar{v}_j$ for every pair of customers $k, j \in J$. Then, by straightforward induction on k, one can prove that the ratio between the highest and the lowest average valuations in J_k is at most α/δ^{k-1}, yielding $K \leq 1 + \log_\delta \alpha = 1 + \ln \alpha / \ln \delta$. Thus, 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 pricing vector $p^k = (p_1^k, \ldots, p_m^k)$ where price p_i^k of segment $i \in I$ is determined as follows. Let $S_{ik} \subseteq S_k$ be the set of customers requesting segment i. If $S_{ik} = \emptyset$, then price p_i^k can be chosen arbitrarily. If $S_{ik} \neq \emptyset$, define $p_i^k = \min\{\bar{v}_j | \ j \in S_{ik}\}$. Now, consider a customer $j \in S_k$. By definition of price vector p^k, the price of sub-path I_j is $\sum_{i \in I_j} p_i^k \leq \sum_{i \in I_j} \bar{v}_j = v_j$, and therefore $j \in W$. By definition of set S_k, $\max_{j \in S_k} \bar{v}_j / \min_{j \in S_k} \bar{v}_j \leq \delta$, that yields the revenue of the solution is 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.*

Clearly, the combination of Lemma 1 and Lemma 2 immediately implies that the total revenue generated by the best price vector p^* from $\{p^k | \ k = 1, \ldots, K\}$ is at least $\Pi/\delta(1 + \frac{\ln \alpha}{\ln \delta})$, which is maximized for $\delta = e^{\left(\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{\ln \alpha}}\right)^{-1}}$. Notice that for big α the value of δ is close to e. Therefore, we have the following result.

Theorem 2. *Price vector p^* yields a total revenue at least $\Pi/(e \ln \alpha + e)$ for the highway problem, where Π is the maximal revenue, and it can be computed in $O(n(\log n + m))$ time.*

We arrive at the computation time as follows. First, we order the customers according to their average valuation (increasingly), which takes $O(n \log n)$ time. Then, for all $k = 1, \ldots, K$, we use binary search to create set S_k in $O(\log n)$ time, and for all items $i = 1, \ldots, m$ we determine the set of customers that request the item in $O(n)$ time, and the item price and the revenue in constant time. So, the total runtime is $O(n \log n + K(\log n + nm))$, which is in $O(n(\log n + m))$, as K is a constant.

There are several directions for improvement of the obtained approximate solution to the highway problem. First, instead of the constructed price vectors

$p^k, k = 1, \ldots, K$, 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 [6, 8]. Unfortunately, this approach does not necessarily lead to any provable improvement of the performance guarantee.

The second approach allows us to improve the performance guarantee, and is based on more careful analysis of the revenue generated by price vector p^* when applied to the entire set J instead of S_k only. 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 $k = 1, \ldots, K$, 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

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

These equalities can be equivalently represented by the following recurrent formulas

$$R_k = \frac{1}{\delta}\Pi_k + \frac{1}{\delta}R_{k+1}, \qquad \forall k = 1, \ldots, K-1, \tag{1}$$

with an additional equality

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

Summing up all Equations (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 $p^k, k = 1, \ldots, K$, we obtain

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

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

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

Clearly, price vector p^* yields a total revenue at least $\Pi/(\delta(1+\frac{\ln\alpha}{\ln\delta}) - \frac{\ln\alpha}{\ln\delta})$. Note that $\delta(1 + \frac{\ln\alpha}{\ln\delta}) - \frac{\ln\alpha}{\ln\delta} < \delta\ln\alpha + \delta$. 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 (Improved Bound). *Price vector p^* yields a total revenue at least $\Pi/(1 + \ln \alpha + \varepsilon)$ for the highway problem for any $\varepsilon > 0$, and it can be computed in $O(n(\log n + m))$ time.*

4 General Bundle Pricing

As a matter of fact, in all arguments developed in the previous sections, we did not make use of the fact that the subsets I_j requested by customers are sub-paths of a path. Hence, the results hold for the more general bundle pricing problem where customers request arbitrary subsets of a given set of items, each of which available in unlimited supply (digital goods, for example). This problem is in general known to be inapproximable by a semi-logarithmic factor in the number of customers n [3]. This inapproximablity result is no longer valid as soon as the inhomogeneity is bounded by a constant, since we have:

Corollary 1. *Given $\varepsilon > 0$, the bundle pricing problem admits an approximation algorithm that yields a revenue at least $(1 + \ln \alpha + \varepsilon)^{-1}$ times the optimal revenue, with computation time $O(n(\log n + m))$.*

For this problem, we can even derive a stronger negative result than for the more restrictive highway pricing problem.

Theorem 4. *The bundle pricing problem is strongly NP-hard, even when restricted to the instances satisfying the following conditions:*

1. *the inhomogeneity $\alpha \leq 1 + \varepsilon$ where ε is an arbitrary positive constant;*
2. *customers valuations are monotone, i.e., $v_j \leq v_k$ for any $j, k \in J$ such that $I_j \subseteq I_k$;*
3. *customers average valuations are monotone decreasing, i.e., $\bar{v}_k \leq \bar{v}_j$ for any $j, k \in J$ such that $I_j \subseteq I_k$.*

Proof. We show that the bundle pricing problem is strongly NP-hard by using a reduction from the strongly NP-hard problem INDEPENDENTSET [5]. Given a graph $G = (V, E)$ and integer $s \leq |V|$. Does there exist a set of vertices that are pairwise non-adjacent with cardinality at least s. We define an instance $\mathcal{I}$ of the pricing problem as follows. Given an $\varepsilon > 0$, let $M > \max\{1/\varepsilon, s + 1/2\}$. For every vertex $v \in V$ we create two *vertex-items*, v^1 and v^2, and for every edge $e \in E$ we introduce two *edge-items*, e^1 and e^2. Every vertex- and edge-item is requested by $2M^2 + 2M - 1$ customers with valuation $M + 1$. For every vertex $v \in V$, there is one customer interested in bundle $\{v^1, v^2\}$ and similarly, for every edge $e \in E$, there is one customer interested in bundle $\{e^1, e^2\}$. These customers have valuation $2M + 2 - 1/M$. There is one customer interested in item x with valuation $M + 1$, and there are 2 customers interested in bundle y of size M with valuation M^2. Also, there are two customers requesting bundle $\{x, y\}$ (of size $M + 1$) with valuation $M^2 + M$. Then, for every edge $e = \{u, v\} \in E$, there

is one customer interested in bundle $\{u^1, u^2, v^1, v^2, e^1, e^2\} \cup \{x\}$ with valuation $7M+6-2/M$. One customer requests all vertex items and item x, that is, bundle $\{v^1, v^2 : v \in V\} \cup \{x\}$, with valuation $(2M + 2 - 1/M)|V| + M + (1/M)s$. The instance is displayed in Figure 3.
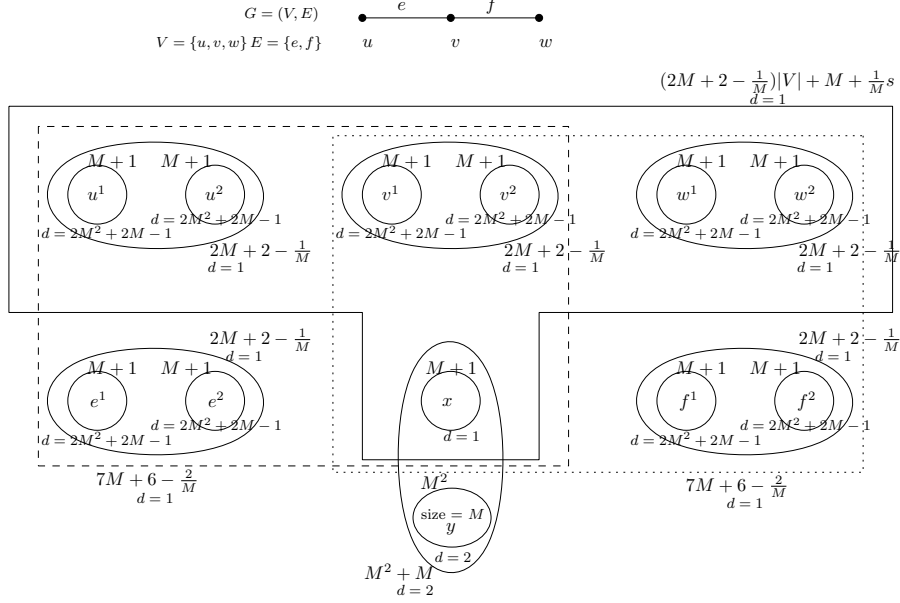

Fig. 3. An instance of the bundle pricing problem created from original graph G above.

Let us give a short intuition as to why we need these particular bundles. The bundles on the vertex- and edge-items determine which vertices are in the independent set of G and bundles $\{u^1, u^2, v^1, v^2, e^1, e^2\} \cup \{x\}$ assure later that a feasible solution to the general bundle pricing problem corresponds to an independent set in G. Bundle $\{v^1, v^2 : v \in V\} \cup \{x\}$ assures that a feasible solution to the pricing problem corresponds to an independent set of cardinality s. Finally, bundles $\{x\}, \{y\}$ and $\{x, y\}$ are present to fulfill the conditions required in this theorem.

The single-item bundles have the largest average valuation of $M + 1$, and bundles $\{y\}$ and $\{x, y\}$ have the smallest average valuation of M, thus $\alpha = 1 + 1/M < 1 + \varepsilon$. Though it requires an extensive case study, one can straightforwardly verify that conditions (2) and (3) are also satisfied.

We define π_i as the revenue obtained from the customers requesting a bundle from set $\{i^1, i^2, \{i^1, i^2\}\}$ for all $i \in I = V \cup E$. We define π_e as the revenue

from the customers requesting $\{u^1, u^2, v^1, v^2, e^1, e^2\} \cup \{x\}$ for some $e = \{u, v\} \in E$. We define π_{xy} as the revenue received from customers requesting a bundle from set $\{x, y, \{x, y\}\}$, and finally, π_V represents the revenue obtained from the customer requesting $\{v^1, v^2 : v \in V\} \cup \{x\}$. Obviously, the total revenue is $\pi = \sum_{i \in I} \pi_i + \sum_{e \in E} \pi_e + \pi_{xy} + \pi_V$. Also, let C be a constant equal to $(|V| + |E|)(4M^3 + 8M^2 + 2M - 2) + |E|(7M + 6 - 2/M) + 4M^2 + 4M$. We claim that there exists an independent set in G of size s if and only if there exists a solution to the general bundle pricing problem with revenue at least $C + s/M$.

Given an independent set $V' \subseteq V$ of size $|V'| = s$. Define $E_0 = \{e = \{u, v\} \in E : u, v \notin V'\}$. Let $p_i = (p_{i^1}, p_{i^2})$ be defined by $p_i = (M + 1, M + 1)$ if $i \in V \cap V'$ or $i \in E \cap E_0$ and $p_i = (M + 1 - \frac{1}{2M}, M + 1 - \frac{1}{2M})$ if $i \in V \setminus V'$ or $i \in E \setminus E_0$. Also, let $p_x = M$ and $p_y = M^2$, where p_y denotes the sum of all M item prices in bundle $\{y\}$. Under this pricing strategy, we see that $\pi_i = 4M^3 + 8M^2 + 2M - 2$ for all $i \in I = V \cup E$, irrespective of which pricing is used for item i. Then, every edge $e = \{u, v\} \in E$ contains one item priced at $(M + 1, M + 1)$ and two at $(M + 1 - \frac{1}{2M}, M + 1 - \frac{1}{2M})$ by definition of the pricing and set E_0. As $p_x = M$, we have $\pi_e = 2(M + 1) + 4(M + 1 - \frac{1}{2M}) + M = 7M + 6 - 2/M$. The customer requesting all vertex-items and item x spends $(2M + 2 - 1/M)|V \setminus V'| + (2M + 2)|V'| + p_x = (2M + 2 - 1/M)|V| + M + (1/M)s$. Then, the total revenue is $\pi = (|V| + |E|)\pi_i + |E|\pi_e + \pi_{xy} + \pi_V = C + (1/M)s$.

For the converse, we are given a solution to instance $\mathcal{I}$ with revenue at least $C + (1/M)s$. First, we consider π_{xy}. If the customer requesting bundle $\{x, y\}$ is not a winner, the maximum revenue is $M + 1 + 2M^2$. Otherwise, let p_x be the price for item x. Then, the maximum revenue is $p_x + 2(M^2 + M - p_x) + (M^2 + M)$, where $p_x \in [M, M + 1]$ such that all customers are winners. Then, $\pi_{xy} \leq 4M^2 + 3M$ (attained when $p_x = M$). For every item $i \in I$, we have $\pi_i = \max\{2(2M^2 + 2M - 1)(M + 1), (2M^2 + 2M - 1 + 1)(2M + 2 - 1/M)\}$. Both values are equal and therefore, $\pi_i = 4M^3 + 8M^2 + 2M - 2$. Clearly, for every $e \in E$, the revenue π_e is at most the valuation $7M + 6 - 2/M$. Now, we know that the revenue from the customer requesting bundle $\{v^1, v^2 : v \in V\} \cup \{x\}$ is

$$\pi_V = \pi - (|V| + |E|)\pi_i - |E|\pi_e - \pi_{xy} \geq (2M + 2 - 1/M)|V| + M + (1/M)s.$$

Thus, the minimum revenue is at least equal to the valuation. As this customer cannot contribute more than the valuation, it should be equality throughout. This also means that all other revenues described above attain their maximum, thus $p_x = M$ and $p_y = M^2$. Now, let $V' = \{v \in V : p_v = (M + 1, M + 1)\}$ and $E_0 = \{e \in E : p_e = (M + 1, M + 1)\}$. As $\pi_e = 7M + 6 - 2/M$ and $p_x = M$ for all $e = \{u, v\} \in E$, we know that either $u \in V'$ and $v \notin V'$, $e \notin E_0$, or $v \in V'$ and $u \notin V'$, $e \notin E_0$, or $e \in E_0$ and $u, v \notin V'$. Thus, for each edge, either one vertex is in V' or both are not in. Hence, V' is an independent set. Furthermore, the customer requesting bundle $\{v^1, v^2 : v \in V\} \cup \{x\}$ pays

$$(2M + 2 - 1/M)|V \setminus V'| + (2M + 2)|V'| + p_x = (2M + 2 - 1/M)|V| + M + (1/M)|V'|.$$

As this payment is equal to the revenue, which in turn has to be equal to the valuation, we know that $|V'| = s$. $\qquad\square$

5 Conclusions

Clearly, the existence of a quasi-PTAS for the highway pricing problem suggests that a PTAS might be in reach [4]. Yet, we leave it as an open problem to derive a PTAS, even for bounded inhomogeneity of valuations.

References

1. M. F. Balcan and A. Blum. Approximation algorithms and online mechanisms for item pricing. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 29–35. ACM, 2006.
2. P. Briest and P. Krysta. Single-minded unlimited supply pricing on sparse instances. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1093–1102. ACM-SIAM, 2006.
3. E. D. Demaine, U. Feige, M.T. Hajiaghayi, and M. R. Salavatipour. Combination can be hard: Approximability of the unique coverage problem. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 162–171. ACM-SIAM, 2006.
4. K. Elbassioni, R. Sitters, and Y. Zhang. A quasi-ptas for profit-maximizing pricing on line graphs. In L. Arge and E. Welzl, editors, *Proceedings of the 15th Annual European Symposium on Algorithms*, volume 4698 of *Lecture Notes in Computer Science*, pages 451–462. Springer, 2007.
5. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 1979.
6. A. Grigoriev, J. van Loon, R. Sitters, and M. Uetz. Optimal pricing of capacitated networks. *Networks*, 2008, to appear.
7. A. Grigoriev, J. van Loon, M. Sviridenko, M. Uetz, and T. Vredeveld. Bundle pricing with comparable items. In L. Arge, M. Hoffmann, and E. Welzl, editors, *Algorithms - ESA 2007*, volume 4698 of *Lecture Notes in Computer Science*, pages 475–486. Springer, 2007.
8. V. Guruswami, J. D. Hartline, A. R. Karlin, D. Kempe, C. Kenyon, and F. McSherry. On profit-maximizing envy-free pricing. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1164–1173. ACM-SIAM, 2005.