UDC 519.1

B. GOLDENGORIN
Department of Operations, University of Groningen, The Netherlands
Khmelnitskiy Private University of Economics and Business, Khmelnitskiy, Ukraine
V. KUZ'MENKO
Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, Kiev, Ukraine

DATA COMPRESSION FOR p-MEDIAN PROBLEMS

In this paper we coin data truncation properties of p-Median Problem (PMP) instances defined on a bipartite graph with m sites and n clients. Both truncations are related to a column (client) of an PMP instance matrix and based on the pseudo-Boolean polynomial (pBp) of the PMP. The first truncation is induced by combining of like terms in the pBp and the second one is induced by the reduction of pBp degree from $m-1$ to $m-p$. We use our PMP representation to explain why some p-median problem instances are more difficult to solve to optimality than other instances of the same cardinality of feasible solutions sets. Based on the uniqueness of pBp we define instances of this problem that are equivalent, in the sense that each feasible solution has the same objective function value in all such instances, and provide an efficient description of all equivalent instances by a polytope. We also show that the problem of optimal aggregation of pBp terms into the minimum number of aggregated clients is equivalent to the problem of finding the maximum cardinality of an antichain defined on the set of partially ordered pBp terms (Dilworth' decomposition theorem). Our computational experiments demonstrate an essential compactification in benchmark instances.

1. Introduction

The *p*-Median Problem (PMP) is one of well known problems within minisum location-allocation problems. A detailed introduction to this problem and solution methods appear in Reese [1] and [2]. For a directed weighted graph $G=(V,A,C)$, with number of vertices $|V|=n$, set of arcs $(i,j)\in A\subseteq V\times V$, and weights (distances, similarities, etc) $C=\{c(i,j):(i,j)\in A\}$, the PMP consists of determining p nodes (the median nodes) such that $1\le p\le n$ minimizing the total sum of weights to all other nodes of the graph.

The PMP is a generalization of classical Fermat's (respectively, Weber's [3]) problem defined on three different points (respectively, weighted points to model client demand) in a plane, with the purpose (objective function) to find a median point in the plane such that the sum of the distances from each of the points to the median point induced by the triangle spanned on these points is minimized. Hakimi [4, 5] has generalized the Weber problem to the problem of finding such a vertex on a graph *(absolute median)* that minimizes the sum of the weighted distances between that absolute median and the vertices of the graph. Hakimi has shown that an optimal absolute median is always located at a vertex of the graph, providing a discrete representation of a continuous problem allowing that the absolute median may be located at any point along the graph's arcs. Similarly Hakimi [5] has generalized the absolute median to the PMP, again providing a discrete representation of a continuous problem by restricting the set of feasible solutions to the vertices.

Another common model within minisum location-allocation problems is the Uncapacitated Facility Location Problem (UFLP), often referred to as the Simple Plant Location Problem (SPLP) (see [6]) or the warehouse location problem (see Balinski [7] and Khumawala [8]). The UFLP is similar to the PMP, and the methods used to solve one are often adapted to solve the other. The objective function of the SPLP is one of determining the cheapest method of meeting the demands of a set of clients $J=\{1,...,n\}$ from plants that can be located at some candidate sites $I=\{1,...,m\}$. The costs involved in meeting the client demands include the fixed cost of setting up a plant at a given site, and the per unit transportation cost of supplying a given client from a plant located at a given site. The PMP and SPLP differ in the following details. First, SPLP involves a fixed cost for locating a facility at a given vertex, and the PMP does not. Second, unlike the PMP, SPLP does not have a constraint on the maximum number of facilities. Typical SPLP formulations separate the set of potential facilities (sites location, cluster centers) from the set of demand points (clients). In the PMP these sets are identical, i.e. $I = J$. Such problems are well known problems in clustering analysis (see Brusco and Kohn [9] and references within). Both problems form underlying models in several combinatorial problems, like set covering, set partitioning, information retrieval, simplification of logical Boolean expressions, airline crew scheduling, vehicle dispatching (see Christofides [10]), assortment (see [11 – 15]) and is a subproblem for various location analysis problems (see Revelle and Laporte [16]).

An instance of the SPLP has an optimal solution in which each client is satisfied by exactly one plant. A similar observation is valid for the PMP. In Hammer [17] this fact is used to derive a pseudo-Boolean representation of the SPLP. The pseudo-Boolean polynomial (PBP) developed in that work has terms that contain both a literal and its complement. Subsequently, in Beresnev [18] a different pseudo-Boolean form has been developed in which each term contains only literals or only their complements. We find this form easier to manipulate, and hence use Beresnev's formulation in this paper which we term as *Hammer-Beresnev polynomial.*

Based on the Hammer-Beresnev formulation for the SPLP [19] have derived a pegging rule within a branch-and-peg algorithm. The test series reveals the advantage of the suggested branch and peg approach, whose computation times are significantly lower than that of comparable branch and bound techniques (see [20]). Goldengorin [21] have incorporated the Hammer-Beresnev polynomial in Goldengorin's data correction approach (see [22]). These authors present reduction rules that are significantly more powerful than those suggested by Khumawala [8]. Recently Alekseeva [23] have applied the Hammer-Beresnev polynomial for a formulation of the PMP with the purpose to analyze the complexity of different local search heuristics.

The PMP is NP-hard (Kariv and Hakimi [24]), and has many applications in location (see [20] and references within) and clustering analysis (see Mirkin [25] and references within). A recent computational study by Avella [26] shows that PMP instances with $|I \times J| > 360{,}000$ are difficult for commercial MIP codes, mainly due to memory restrictions.

Problem reduction is a very common technique in integer programming and combinatorial optimization; see, for example, Andersen and Andersen [27], Crowder [28], Goldengorin [21], Hoffman and Padberg [29], Martin [30], Martin [31], Suhl and Szymanski [32]. Classical reductions of PMP instances are based either on reduction tests (see Avella and Sforza [33]) or on good lower bounds (see Briant and Naddef [34]). In this paper, we present a reduction of PMP instances using a pseudo-Boolean formulation of PMP due to Hammer [17] and Beresnev [18].

Since the PMP is NP-hard and many polynomially solvable special cases are well known in the literature (see the 1-median problem on a cactus in Burkard and Krarup [35]). In this context Burkard [36, p. 155] presents an open problem. For the PMP, this problem can be stated as follows:

"Suppose we are given a PMP instance defined on a cost matrix *C* which does not belong to a polynomially solvable class of PMP instances. Is it possible to modify *C* into a cost matrix *D* belonging to a polynomially solvable class of PMP such that an optimal solution to the original problem instance is as close as possible to the optimal solution of the modified instance?"

We show that our pseudo-Boolean formulation allows us to find such modifications if the polynomially solvable class of PMP instances is defined algebraically in terms of the elements in its cost matrix. For this, we describe the concept of equivalent instances. Moreover, we reduce the problem of finding an equivalent cost matrix *D* with the minimum number of columns to the given matrix *C* relates to the well known Dilworth's decomposition theorem (see Theorem 14.2 in Schrijver [37]).

While this paper does not suggest any new algorithm for solving the PMP it presents some fundamental properties of PMP derived from the pseudo-Boolean representation of the problem. Our paper is organized as follows. In Section 2 of this paper, we adjust the Hammer-Beresnev's pseudo-Boolean formulation of the Simple Plant Location Problem (SPLP) (see Hammer [17]) to the PMP, and show that combining of like terms in the Hammer-Beresnev's pseudo-Boolean polynomial leads to the aggregation of entries in the given PMP instance. In Section 3, based on the truncation of degree of Hammer-Beresnev polynomial from $(m-1)$ to $(m-p)$ we are further able to aggregate the entries of PMP instance by introducing the so called *p-truncated* columns of the PMP matrix *C* in Section 3 and use it to develop rules for preprocessing PMP instances. We also show that the pseudo-Boolean representation allows us to comment on relative difficulties in obtaining provably optimal solutions to different PMP instances. Both sections 2 and 3 include computational analysis of benchmark instances similar to instances used in [26]. Section 4 defines the concept of equivalent instances and describes an algebraic method of modifying the cost matrix of a PMP instance without disturbing the optimality of any optimal solution to the original instance. This answers Burkard *et al's* open problem affirmatively in the context of PMP. It also indicates the relationships with the minimum number of aggregated columns and Dilworth's decomposition theorem. Section 5 summarizes the main results of the paper, and points to directions for future research.

2. A Pseudo-Boolean Formulation of the PMP

Recall that given sets $I = \{1,2,...,m\}$ of sites in which plants can be located, $J = \{1,2,...,n\}$ of clients, a matrix $C = [c_{ij}]$ of costs of supplying each $j \in J$ from each $i \in I$, the number p of plants to be opened, and unit demand at each client site, the p-Median Problem (PMP) is one of finding a set $S \subseteq I$ with $|S| = p$, such that the total cost

$$f_C(S) = \sum_{j \in J} \min\{c_{i,j} \,|\, i \in S\}$$

is minimized. An instance of the problem is described by a m.n matrix $C = [c_{ij}]$ and the number $1 \le p \le |I|$. We assume that the entries of *C* are nonnegative and finite, i.e. $C \in \Re_{+}^{m\,n}$. The PMP is thus the problem of finding

$$S^* \in \arg\min\{f_C(S) : \varnothing \subset S \subseteq I, |S| = p\}. \quad (1)$$

An $m \times n$ *ordering matrix* $\prod = [\pi_{ij}]$ is a matrix each of whose columns $\prod_j = (\pi_{1j},...,\pi_{mj})^T$ define a permutation of 1, . . . ,*m*. There may exist several ordering matrices for a given instance of the PMP. Given a matrix *C*, the set of all ordering matrices $\prod$ such that $c_{\pi_{1j}j} \le c_{\pi_{2j}j} \le ... \le c_{\pi_{mj}j}$, for $j = 1,...,n$, is denoted by perm(*C*).

Corresponding to an ordering matrix $\prod = [\pi_{ij}]$, an $m \times n$ *difference matrix* $\Delta = [\delta_{ij}]$ can be constructed, in which

$$\delta_{1k} = c_{\pi_{1k}k}$$
$$\delta_{rk} = c_{\pi_{rk}k} - c_{\pi_{(r-1)k}k} \text{ for } 2 \le r \le m \tag{2}$$

Defining

$$y_i = \begin{cases} 0 & if\ i \in S \\ 1 & otherwise, \end{cases} \text{ for each } i = 1, \ldots, m \tag{3}$$

we can indicate any solution S by a vector $y = (y_1, y_2, ..., y_m)$. Its total cost is given by

$$\mathrm{B}_{C,\Pi}(y) = \sum_{j=1}^{n} \{\delta_{1j} + \sum_{k=2}^{m} \delta_{kj} \cdot \prod_{r=1}^{k-1} y_{\pi_{rj}}\} \tag{4}$$

Note that a solution y is called feasible if $\sum_{i=1}^{m} y_i = m - p$.

In [21] it is shown that the total cost function is identical for all permutations in *perm(C)*. Hence we can remove the **Π** in $\mathrm{B}_{C,\Pi}(y)$ without introducing any confusion. We call this pseudo-Boolean representation of the total cost $\mathrm{B}_C(y)$ the Hammer-Beresnev polynomial representation since this representation of the total cost was first presented in the context of uncapacitated facility location problems independently in Hammer [17] and Beresnev [18].

We can formulate (1) in terms of Hammer-Beresnev polynomials as

$$y^* \in \arg\min \{B_C(y) : y \in \{0,1\}^m, \sum_{i=1}^{m} y_i = m - p\}. \tag{5}$$

Hammer-Beresnev polynomials allow a compact description of p-median problems, since it allows combining of like terms. Consider for example, the following PMP instance where *m = 4, n = 5, p = 2* and

$$C = \begin{bmatrix} 7 & 15 & 10 & 7 & 10 \\ 10 & 17 & 4 & 11 & 22 \\ 16 & 7 & 6 & 18 & 14 \\ 11 & 7 & 6 & 12 & 8 \end{bmatrix}. \tag{6}$$

A possible ordering matrix for this problem is given by

$$\Pi = \begin{bmatrix} 1 & 3 & 2 & 1 & 4 \\ 2 & 4 & 3 & 2 & 1 \\ 4 & 1 & 4 & 4 & 3 \\ 3 & 2 & 1 & 3 & 2 \end{bmatrix}$$

and the corresponding difference matrix is given by

$$\Delta = \begin{bmatrix} 7 & 7 & 4 & 7 & 8 \\ 3 & 0 & 2 & 4 & 2 \\ 1 & 8 & 0 & 1 & 4 \\ 5 & 2 & 4 & 6 & 8 \end{bmatrix}$$

The Hammer-Beresnev polynomial corresponding to C is

$$B_C(y) = [7 + 3y_1 + 1y_1y_2 + 5y_1y_2y_4] + [7 + 0y_3 + 8y_3y_4 + 2y_1y_3y_4] + [4 + 2y_2 + 0y_2y_3 + 4y_2y_3y_4] +$$
$$+ [7 + 4y_1 + 1y_1y_2 + 6y_1y_2y_4] + [8 + 2y_4 + 4y_1y_4 + 8y_1y_3y_4],$$

whose terms can be aggregated into the polynomial

$$33 + 7y_1 + 2y_2 + 2y_4 + 2y_1y_2 + 8y_3y_4 + 4y_1y_4 + 11y_1y_2y_4 + 10y_1y_3y_4 + 4y_2y_3y_4.$$

Note that the original $B_C = (y)$ has 20 terms including the terms with zero coefficients, and after combining of like terms there are just 10 terms.

Table 1 shows the reductions obtained through combining of terms in Hammer-Beresnev polynomials for benchmark PMP instances considered in [26]. The OR, ODM, and TSP instances are available from [38], [34], and [39], respectively. Here, *m, C,* and $B_C = (y)$ correspond to notations developed earlier and Reduction refers to the reduction in terms obtained due to combining of terms in the Hammer-Beresnev polynomial representation. For example, in the pmed15 instance, the cost matrix C had 90,000 entries, while the Hammer-Beresnev polynomial had only 17,102 terms, leading to a percentage reduction of $((90000 - 17102) \times 100)/90000 = 81.00\%$.

Hammer-Beresnev polynomials are easily manipulated in computer programs, refer to [19] for details on data structures to store and manipulate these polynomials.

Table 1

Reductions of terms in benchmark instances

library	instance	m	Entries in C matrix	Terms in $\mathcal{B}_C(\mathbf{y})$	Reduction(%)
OR	pmed15	300	90,000	17,102	81.00
OR	pmed26	600	360,000	25,083	93.03
OR	pmed40	900	810,000	30,756	96.20
ODM	BN48	42	411	329	19.95
ODM	BN1284	1284	88,542	85,416	3.53
ODM	BN3773	3773	349,524	341,775	2.22
ODM	BN5535	5535	666,639	654,709	1.79
TSP	D657	657	430,992	367,355	14.77
TSP	fl1400	1400	1,958,600	836,557	57.29
TSP	pcb3038	3038	9,226,406	5,759,404	37.58

The number of terms in the Hammer-Beresnev representation of a PMP instance can be further reduced by exploiting the fact that for any feasible solution y to the PMP instance, $\sum_{i=1}^{m} y_i = m - p$ Consider for example, the Hammer-Beresnev polynomial derived for the PMP instance corresponding to the cost matrix (6). Since *p = 2, exactly (4 − 2) = 2* of the y_i will equal zero in any feasible solution. Therefore, each cubic term in the polynomial will evaluate to zero, and the truncated polynomial

$$B_{C,p=2}(y) = 33 + 7y_1 + 2y_2 + 2y_4 + 2y_1y_2 + 8y_3y_4 + 4y_1y_4$$

would adequately describe the PMP instance. Notice that this polynomial has only seven terms. Such truncations are possible for all PMP instances with $p \le m$ and the truncated *Hammer-Beresnev polynomial* for a PMP instance with cost matrix *C* is given by

$$\mathrm{B}_{C,p}(y) = \sum_{j=1}^{n} \{\delta_{1j} + \sum_{k=2}^{m-p} \delta_{kj} \cdot \prod_{r=1}^{k-1} y_{\pi_{rj}}\} \tag{7}$$

We formalize this result through the following theorem.

Theorem 1. For any PMP instance C with $p \le m$ the following assertions hold: 1. The degree of truncated Hammer-Beresnev polynomial $\mathrm{B}_{C,p}(y)$ is at most m − p; 2. The truncated Hammer-Beresnev polynomial is equal to the Hammer-Beresnev polynomial for any feasible solution y to the PMP instance, i.e. $\mathrm{B}_C(y) = \mathrm{B}_{C,p}(y)$.

Proof. The assertions follows from the fact that in a PMP instance, exactly p components of any feasible solution y equal 0, and therefore, any term in a Hammer-Beresnev polynomial expressed as a product of *m − p* or more components of y will evaluate to zero in any feasible solution.

Corollary 1. A corollary to Theorem 1 is a reformulation of the definition of PMPs in terms of truncated Hammer-Beresnev polynomials.

Corollary 2. A PMP instance with m possible facilities, n clients, and a cost matrix C can be represented as

$$y^* \in \arg\min \{B_{C,p}(y) : y \in \{0,1\}^m, \sum_{i=1}^{m} y_i = m - p\} \tag{8}$$

Proof. The proof is trivial.

Theorem 1 also shows that the largest p entries in any column of the cost matrix *C* will be not involved in obtaining an optimal solution to the PMP. This leads to the following *p*-truncation operation.

Definition 1. For any $p \le m$ the column j of matrix C is called *p*-truncated if the values of the largest p elements in the column are replaced with $c_{\pi(m-p+1)_j j}$.

So the 1-truncation of a column (corresponding to p = 1) leaves it unchanged, and a 2-truncation of a column replaces the value of its largest element with the value of the second largest element in the column.

Corollary 3. In a PMP problem instance, every column of PMP matrix C may be *p*-truncated without affecting the optimality of the optimal solution to the instance.

Proof. The proof is straightforward from Theorem 1 and the discussion above.

3. Preprocessing PMP Instances

The *p*-truncation operation described in the previous section allows us to reduce the search space for an optimal solution to a PMP instance.

Theorem 2. Assume that in a given PMP instance, all the the entries corresponding to a particular row i in the cost matrix C are changed when *p*-truncation operations are performed on all columns of C. Then there exists an optimal solution y^* to the instance with $y_i^* = 1$.

Proof. Consider any column *j* of *C*. Since the entry for row i has changed after the *p*-truncation operation for this column, $\pi_{ij} > m - p + 1$. Hence there will be no term in the truncated Hammer-Beresnev polynomial containing y_i that is derived from column *j*. If this is true for all columns in *C*, then the truncated Hammer-Beresnev polynomial for the instance will not contain any term containing y_i . The result follows from this observation.

Consider for example, the p-median problem instance defined by the cost matrix in (6). If $p = 3$, the truncated cost matrix is given by

$$C_3 = \begin{bmatrix} 7 & 7 & 6 & 7 & 10 \\ 10 & 7 & 4 & 11 & 10 \\ 10 & 7 & 6 & 11 & 10 \\ 10 & 7 & 6 & 11 & 8 \end{bmatrix} \tag{9}$$

Clearly, all the entries in the third row of C have changed as a result of the p-truncation procedure and hence, from Theorem 2, there exists an optimal solution with $y_3 = 1$ Setting $y_3 = 1$ immediately solves the problem, so that the optimal solution is found to be y = (0, 0, 1, 0).

Given a cost matrix corresponding to a PMP instance, it is interesting to study the variation in the degree of difficulty preprocessing the instance with changing values of p.

As the value of p increases, the number of entries in any column whose values are revises through p-truncation increases. So the higher the value of p is, the more the chance that a row of C is preprocessed out through Theorem 2. Let $p^{'}$ be the smallest value of p for which the p-truncation of columns of C allow Theorem 2 to be used to preprocess at least one row in C. This explains why for $p_0 < m/2$, PMP instances with $p = p_0$ are more difficult to solve than instances on the same cost matrix with $p = m - p_0$ even though the number of feasible solutions to the two are identical.

Also, let p^* denote the minimum number of rows in C which contain all the column minima of C. In other words, a PMP instance defined on C with $p > p^*$ would have open facilities which do not serve any client. Hence, for $p_1 > p^*$, the number of optimal solutions is bounded below by $\binom{m - p^*}{p_1}$. So for PMP instances with $p > p^*$, it become progressively difficult to prove optimality of an optimal solution when the value of p increases from p^* to $(m - p^*)/2$. Then it becomes progressively easier to prove when p increases further. Table 2 presents $p^{'}$ and p^* values for benchmark instances introduced in table 1.

Table 2

p' and p^* values for benchmark instances

library	*instance*	m	p'	$p^\star$
OR	pmed15	300	180	285
OR	pmed26	600	452	581
OR	pmed40	900	644	882
ODM	BN48	42	27	35
ODM	BN1284	1284	653	1211
ODM	BN3773	3773	3385	3742
ODM	BN5535	5535	2179	5503
TSP	D657	657	477	653
TSP	fl1400	1400	1177	1395
TSP	pcb3038	3038	3026	3033

4. Equivalent PMP Instances

More than one PMP instances can have identical truncated Hammer-Beresnev polynomials since the truncated Hammer-Beresnev polynomial allows terms to be aggregated. If two PMP instances do have the same truncated Hammer-Beresnev polynomial, then the same solution would have the same cost in the two instances, and hence, the same solution would be optimal for both instances. Such instances motivate the following definition.

Definition 2. Two PMP instances defined on cost matrices C and D are called equivalent if C and D are of the same size and if $B_{C,p}(y) = B_{D,p}(y)$.

Truncated Hammer-Beresnev polynomials of PMP instances can be generated in polynomial time, and have a number of terms that is polynomial in the size of the instance. Therefore it is possible to check the equivalence of two instances in polynomial time, even though the PMP itself is a NP-hard problem.

Note however that the condition of equivalence is only a sufficient condition for two PMP instances to have the same optimal solution. For example, two PMP instances with cost matrices

$$C = \begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix} \text{ and } D = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

and with p = 1 have different truncated Hammer-Beresnev polynomials ($B_{C,p}(y) = 6 + 4y_1$ and $B_{D,p}(y) = 2 + 2y_1$) but the same unique optimal solution, (0, 1).

We now consider the set of all PMP instances D that are equivalent to a given PMP instance C. This set can be defined as

$$P_C = \{D \in \Re_+^{m \times n} : B_C = B_D\} \tag{10}$$

We show that the set P_C can be described by a system of linear inequalities.

Let us assume that $\Psi, \Pi \in perm(C)$. The choice of the particular Ψ and Π is unimportant since the truncated Hammer-Beresnev polynomials for all permutations within perm(·) are in a PMP instance are identical (see [40]). Let the difference matrices corresponding to *C* and *D* be Δ^C and Δ^D respectively. The truncated Hammer-Beresnev polynomial for *C* is

$$B_{C,p}(y) = \sum_{j=1}^{n} \delta_{1j}^{C} + \sum_{j=1}^{n}\sum_{k=2}^{m-p} \delta_{kj}^{C} \cdot \prod_{r=1}^{k-1} y_{\Psi_{rj}} \tag{11}$$

while that for D is

$$B_{D,p}(y) = \sum_{j=1}^{n} \delta_{1j}^{D} + \sum_{j=1}^{n}\sum_{k=2}^{m-p} \delta_{kj}^{D} \cdot \prod_{r=1}^{k-1} y_{\pi_{rj}} \tag{12}$$

For the PMP defined on *D* to be equivalent to the PMP defined on *C*, $B_{D,p}(y)$ has to equal $B_{C,p}(y)$ Equating like terms in $B_{C,p}(y)$ and $B_{D,p}(y)$, we see that for equivalence, entries in Δ^D have to satisfy the following equations:

From equating constant terms:

$$\sum_{j=1}^{n} \delta_{1j}^{D} - \sum_{j=1}^{n} \delta_{1j}^{C} = 0 \tag{13}$$

From equating linear and nonlinear terms:

$$\sum_{\{\Psi_{1j},\ldots,\Psi_{(k-1)j}\}=\{\pi_{1j},\ldots,\pi_{(k-1)j}\}} \delta^D kj - \delta^C k, j = 0 \qquad k = 2,\ldots,m-p; \quad j = 1,\ldots,n. \tag{14}$$

Further, since $\mathbf{\Pi} \in perm(D)$ is a permutation matrix for a PMP instance,

$$\delta_{ij}^{D} \geq 0 \;\text{ for }\; i = 1,\ldots,m; \quad j = 1,\ldots,n. \tag{15}$$

Hence, given a cost matrix *C*, any solution Δ^D to the set of inequalities (13) – (15) will be a difference matrix for an equivalent instance. Given a permutation matrix $\Psi \in perm(C)$ and a difference matrix Δ^D, it is trivial to construct the cost matrix *D* of a PMP instance equivalent to a PMP instance with cost matrix *C*.

Remark 1. Note that for a PMP instance defined on a cost matrix D to be equivalent to a PMP instance defined on a cost matrix C, perm(D) has to be identical to perm(C).

Remark 2. Note that we reduce the problem of finding an equivalent matrix D with the minimum number of columns to the given matrix C to the following well known Dilworth's decomposition theorem (see Theorem 14.2 in Schrijver [37]):

"The set of terms T_a with positive coefficients in a pseudo-Boolean polynomial are subsets of partially ordered set T, and hence, the minimum number of chains covering T_a (nothing else as the minimum number of aggregated columns of C) is equal to the maximum size of an antichain (the maximum number of non-embedded terms)."

The maximum size of antichain found for example (6) is equal to three, and the corresponding Hammer-Beresnev polynomial $B_{C,p=2}(y)$, aggregated matrix *D* and one of its permutation matrix $\mathbf{\Pi}_D$ are as follows:

$$B_{C,p=2}(y) = [33 + 7y_1 + 4y_1y_4] + [0 + 2y_2 + 2y_1y_2] + [0 + 2y_4 + 8y_3y_4],$$

with matrices

$$D = \begin{bmatrix} 33 & 2 & 10 \\ 44 & 0 & 10 \\ 44 & 4 & 2 \\ 40 & 4 & 0 \end{bmatrix} \text{ and } \prod\nolimits_D = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 1 & 3 \\ 2 & 3 & 1 \\ 3 & 4 & 2 \end{bmatrix}. \tag{16}$$

5. Concluding Remarks and Directions for Future Research

The size required to represent a p-median problem (PMP) instance is indicated by the size of the cost matrix for the instance. It is often possible to represent the instance in a more compact manner. In Section 2 we have presented a representation of the PMP through a pseudo-Boolean polynomial called the truncated Hammer-Beresnev polynomial which achieves this compactification. This compactification is mainly achieved through the combining of like terms in the polynomial, and the truncation of the polynomial to degree $m - p$ for a p median problem with m candidate facilities. Computations presented in the section show that the compactification is significant for OR and TSP, but negligible for ODM benchmark problem instances.

In Section 3 we present a preprocessing procedure for p-median problems based on its truncated Hammer-Beresnev polynomial representation. This representation allows us to perform a p-truncation operation on the cost matrix representing the problem instance, which in turn allows us to remove certain facilities from the consideration set for optimal solutions. Additionally, it allows us to explain why certain p-median problem instances are more difficult to solve than others.

In Section 4 we show how we can construct PMP instances that have the same optimal solutions as a given PMP instance. These instances are called equivalent instances. Construction of equivalent instances is practically useful; given a PMP instance, we can search for an equivalent PMP instance belonging to a known polynomially solvable class of PMP instances, solve the equivalent instance easily, and hence come up with an optimal solution to the original instance. It also allows us to use data correcting algorithms (see [21] to generate good quality solutions in reasonable time).

It is interesting that the above-mentioned results and properties of PMP derived from its PBP are much more difficult to discover from the mathematical programming formulation of PMP (see [26]).

The paper leads to two interesting courses of future research on p-median problems. The first direction is to use the concept of equivalence described here to extend the set of polynomially solvable PMP instances. The second direction is to design new exact and heuristic algorithms for solving large-scale instances of PMP based on the truncated Hammer-Beresnev polynomial and the preprocessing scheme demonstrated in Section 3.

List of the used references

1. J. Reese. Solution Methods for the p-Median Problem: An Annotated Bibliography // Networks. – N. 48 (3). – 2006. – P. 125 – 142.
2. N. Mladenovic, J. Brimberg, P. Hansen, J. A. Moreno-Perez. The p-median problem: A survey of metaheuristic approaches // European Journal of Operational Research. – N. 179. – 2007. – P. 927 – 939.
3. Z. Drezner, J. Guyse. Application of decision analysis to theWeber facility location problem // European Journal of Operational Research. – N. 116. – 1999. – P. 69 – 79.
4. S. L. Hakimi. Optimum locations of switching centers and the absolute centers and medians of a grap // Operations Research. – N. 12. – 1964. –P. 450 – 459.
5. S. L. Hakimi. Optimum distribution of switching centers in a communication network and some related graph theoretic problems // Operations Research. – N. 13. – 1965. – P. 462 – 475.
6. G. Cornuejols, G. L. Nemhauser, and L. A. Wolsey. The Uncapacitated Facility Location Problem. Ch.3, Discrete Location Theory, R. L. Francis and P. B. Mirchandani (eds). – New York: Wiley-Interscience, 1990.
7. M. Balinski. Integer Programming, Methods, Uses and Computation // Management Science. – N. 12. – 1965. – P. 253 – 313.
8. B. M. Khumawala. An efficient branch-and-bound algorithm for the warehouse location problem // Management Science. – N. 18. – 1972. – P. 718 – 731.
9. M. J. Brusco, H.-F. Kohn. Optimal partitioning of a data set based on the p-median problem // Psychometrika. – N. 73 (1). – 2008. – P. 89 – 105.
10. N. Christofides. Graph Theory: An Algorithmic Approach. – London: Academic Press Inc. Ltd., 1975.
11. V. L. Beresnev, E. Kh. Gimadi, V. T. Dementyev. Extremal Standardization Problems. – Novosibirsk: Nauka, 1978 (in Russian).
12. B. Goldengorin. Requirements of Standards: Optimization Models and Algorithms. – ROR, Hoogezand, The Netherlands, 1995.
13. P. C. Jones, T. J. Lowe, G. Muller, N. Xu, Y. Ye, J. L. Zydiak. Specially structured uncapacitated facility location problems // Operations Research. – N. 43. – 1995. – P. 661 – 669.
14. D. W. Pentico. The assortment problem: A survey // European Journal of Operational Research. – N. 190. – 2008. – P. 295 – 309.
15. A. Tripathy, Sural, Y. Gerchak. Multidimensional assortment problem with an application // Networks. – N. 33. – 1999. – P. 239 – 245.
16. S. Revelle, G. Laporte. The plant location problem: new models and research prospects // Operations Research. – N. 44. – 1996. – P. 864 – 874.
17. P. L. Hammer. Plant Location — A Pseudo-Boolean Approach // Israel Journal of Technology. – N. 6. – 1968. – P. 330 – 332.
18. V. L. Beresnev. On a problem of mathematical standardization theory // Upravliajemyje Sistemy. – N. 11. – 1973. – P. 43 – 54 (in Russian).
19. B. Goldengorin, D. Ghosh, G. Sierksma. Branch and peg algorithms for the simple plant location problem // Computers & Operations Research. – N. 30. – 2003. – P. 967 – 981.
20. C. S. Revelle, H. A. Eiselt, M. S. Daskin. A bibliography for some fundamental problem categories in discrete location science // European Journal of Operational Research. – N. 184. – 2008. – P. 817 – 848.
21. B. Goldengorin, G.A. Tijssen, D. Ghosh, G. Sierksma. Solving the simple plant location problems using a data correcting approach // Journal of Global Optimization. – N. 25. – 2003. – P. 377 – 406.
22. B. Goldengorin. A Correcting Algorithm for Solving Discrete Optimization Problems // Soviet Math. Doklady 27. – 1983. – P. 620 – 623.
23. E. Alekseeva, Yu. Kochetov, A. Plyasunov. Complexity of Local Search for the p-Median Problem // European Journal of Operational Research. – N. 191. – 2008. – P. 736 – 752.

24. O. Kariv, L. Hakimi. An Algorithmic Approach to Network Location Problems, Part II: The p-Medians // SIAM Journal of Applied Mathematics. – N. 37 (3). – 1979. – P. 539 – 560.
25. B. Mirkin. Clustering For Data Mining: A Data Recovery Approach. – Chapman & Hall/Crc Computer Science. – Chapman & Hall/CRC, 2005.
26. P. Avella, A. Sassano, I. Vasil'ev. Computational Study of Large-Scale p-Median Problems // Mathematical Programming Ser. A. – N. 109. – 2007. – P. 89 – 114.
27. E. D. Andersen, K. D. Andersen. Presolving in linear programming // Math. Programming. – N. 71. – 1995. – P. 221 – 245.
28. H. Crowder, E. Johnson, M. W. Padberg. Solving large-scale zero one linear programming problems // Operations Research. – N. 31. – 1983. – P. 803 – 834.
29. K. L. Hoffman, M. W. Padberg. Improved LP-representations of zero-one linear programs for branch-and-cut // ORSA J. Comput. – N. 3. – 1991. – P. 121 – 134.
30. Martin. Integer Programs with Block Structure. – Technische Universitt Berlin, Habilitations-Schrift, Berlin, Germany, 1998.
31. Martin. General Mixed Integer Programming: Computational Issues for Branch-and-Cut Algorithms // Lecture Notes in Computer Science 2241. – 2001. – P. 1 – 25.
32. U. H. Suhl, R. Szymanski. Super node processing of mixed-integer models // Comput. Optim. Appl. 3. – 1994. – P. 317 – 331.
33. P. Avella, A. Sforza. Logical reduction tests for the p-median problem // Annals of Operations Research. – N. 86. – 1999. – P. 105 – 115.
34. O. Briant, D. Naddef. The optimal diversity management problem // Operations Research. – N. 52 (4). – 2004. – P. 515 – 526.
35. R. E. Burkard, J. Krarup. A linear algorithm for the pos/neg-weighted 1-median problem on a cactus // Computing. – N. 60. – 1998. – P. 193 – 215.
36. R. E. Burkard, B. Klinz, R. Rudolf. Perspective of Monge properties // Discrete Applied Mathematics. – N. 70. – 1996. – P. 95 – 161.
37. Schrijver. Combinatorial Optimization. Polyhedra and Efficiency. – Springer, 2003.
38. OR-Library // http://mscmga.ms.ic.ac.uk/info.html.
39. TSP-library // http://www.iwr.uni-heidelberg.de/groups/comopt/ software/TSPLIB95/.
40. F. AlBdaiwi, B. Goldengorin, G. Sierksma. Equivalent instances of the simple plant location problem // Computers and Mathematics with Applications. – 2008 // doi:10.1016/j.camwa.2008.10.081