1 Introduction and motivation

The combinatorial optimization problem presented here arises from a real application. A company sells small appliances and must provide spare parts from a central distribution center to several regional distribution centers and sale points. The types of spare parts to be stocked and distributed have different prices. For the sake of easiness in the administrative procedures, the company wants to group them into clusters, such that all spare parts in the same cluster are sold at the same price. The price assigned to each cluster is a decision variable and it must be chosen so that the expected annual income remains equal to a predefined value, estimated on the basis of the expected demand for each type of spare part. This clustering operation may involve a huge number of spare part types and clusters and therefore it must be computer-assisted with a decision support system. The goal is to keep the price of each cluster close to the original price of the spare parts in it. The number of clusters is a user-defined parameter and the decision support system should allow for the analysis of the trade-off between the number of clusters and the maximum difference between the original price of any spare part and the price associated with its cluster.

The problem resembles the well-known K-center problem an a line. However, a major difference is the constraint on the required profit, i.e. the lower bound on the weighted sum of the positions of the representative points in each cluster. Its effect is that locating the cluster representatives at the centers, may not be enough, i.e. it may provide an insufficient total weighted sum of their positions. Therefore the optimal representative of a cluster (the *centroid* in the remainder) may be located to the right of the cluster center, on the real line.

2 Models and properties

The problem can be formulated as follows.

Data. The following data are given:

- a set N of n points, representing the spare part types;
- a weight $w_i \in \Re_+$ for each point $i \in N$, representing the expected demand of spare parts of type i;
- a value $p_i \in \Re_+$ for each point $i \in N$, representing the price of spare parts of type i;
- a coefficient $\alpha \geq 1$, representing the required profit margin.

Variables. A solution is a set of K clusters $\{C_1, \ldots, C_K\}$, with $C_k \subset N \ \forall k = 1, \ldots, K$, such that $\bigcup_{k=1}^{K} C_k = N$ and $C_{k'} \cap C_{k''} = \emptyset \ \forall k' \neq k''$. For each cluster C_k, its centroid is a point on the line in position $q_k \in \Re_+$.

Therefore a solution can be represented with the following variables. A binary assignment variable x_{ik} represents the assignment of point $i \in N$ to cluster C_k. A continuous variable $q_k \in \Re_+$ indicates the position of the centroid of cluster C_k on the real line, (i.e. the price associated with the cluster).

Constraints. A solution is feasible when the following assignment constraints are satisfied:

$$\sum_{k=1}^{K} x_{ik} = 1 \quad \forall i \in N.$$

The constraint on the required profit is translated into a lower bound on the weighted sum of the positions of the centroids:

$$\sum_{k=1}^{K} \sum_{i=1}^{n} w_i q_k x_{ik} \geq \alpha \sum_{i=1}^{n} w_i p_i. \tag{1}$$

This non-linear constraint states that the total expected profit when spare parts are sold at the price q_k of their clusters must be at least α times the income that would be obtained by selling each spare part $i \in N$ at its price p_i. For brevity, in the remainder we keep the term *profit* to refer to the left hand side of (1) (possibly restricted to a subset of clusters).

Objectives. The problem has two conflicting objectives.

1. Minimize the maximum difference between p_i and q_k for any point $i \in N$ in cluster C_k. In the remainder we refer to this quantity as the *offset*, for brevity.

2. Minimize the number of clusters.

Objective 2 can be replaced by a constraint

$$x_{ik} = 0 \quad \forall k > K$$

so that the optimal solution of the problem is computed with respect to Objective 1 for different values of K in a used-defined range. This allows to enumerate all Pareto-optimal solutions of the two-objectives problem.

The objective function corresponding to Objective 1 is

$$\text{minimize } z$$

subject to

$$\begin{aligned} z &\geq q_k - p_i - \overline{p}(1 - x_{ik}) && \forall i \in N, \forall k = 1, \ldots, K \\ z &\geq p_i x_{ik} - q_k && \forall i \in N, \forall k = 1, \ldots, K \end{aligned}$$

where $\overline{p} = \max_{i \in N}\{p_i\}$.

Owing to the presence of non-linear constraints and binary variables and the possibly large size of the instances, the use of general-purpose MINLP solvers cannot be considered a viable option to find provably optimal solutions.

2.1 Additional constraints

We consider two additional constraints that may be imposed to define a feasible solution.

Constraint 1: Non-overlapping clusters. This constraint requires that clusters do not overlap, i.e. there is no pair of points $i \in N$ and $j \in N$ with $p_i < p_j$ that are assigned to clusters $C_{k'}$ and $C_{k''}$, respectively, with $q_{k'} > q_{k''}$.

Constraint 2: Bounded centroids. This constraint requires that the position of the centroid of each cluster be within the interval defined by the minimum and the maximum position of the points in the cluster. In particular, when a cluster C_k includes only one point $i \in N$, then the constraint imposes $q_k = p_i$.

Example. The following toy instance shows the effect of Constraints 1 and 2.

Consider an instance with $N = 4$ points, in positions $p = [100, 140, 150, 205]$ and weight $w = [4, 2, 1, 1]$ to be clustered into $K = 2$ clusters. The total profit is 1185. Assume a value of $\alpha = \frac{1220}{1185}$ such that the target profit is set to 1220.

Consider a solution A with clusters $\{1, 2\}$ and $\{3, 4\}$ (complying with Constraint 1). The target profit can be achieved by setting $q_1 = 140$ and $q_2 = 190$ with an offset $z = 40$.

Consider a solution B with clusters $\{1, 3\}$ and $\{2, 4\}$ (violating Constraint 1). The target profit can be achieved by setting $q_1 = 137.5$ and $q_2 = 177.5$ with an offset $z = 37.5$.

From the comparison, solution B is better than solution A.

Consider now a solution C with clusters $\{1\}$ and $\{2, 3, 4\}$ (complying with Constraint 1). It can be obtained by repairing solution B, i.e. upgrading to cluster 2 all points of cluster 1 whose position is larger than the minimum position of points in cluster 2. In our example $p_3 > p_2$ which is the minimum position in cluster 2: hence point 3 is upgraded from cluster 1 to cluster 2. The resulting solution complies with Constraint 1, by construction. The target profit can be achieved by setting $q_1 = 132.5$ and $q_2 = 172.5$ with offset $z = 32.5$.

Note that in solution C, the centroid of cluster 1 is out of the cluster bounds, which violates Constraint 2.

Repairing solution B to enforce Constraint 2 produces a solution D with q_1 forced to be equal to $p_1 = 100$. Therefore the target profit can be achieved only by setting $q_2 = 205$, with offset $z = 65$.

Symmetrically, consider a solution E with clusters $\{1, 2, 3\}$ and $\{4\}$, that can be obtained by repairing B in a different way (by point downgrading). Since q_2 would be fixed to $p_4 = 205$, then the target profit is achieved by setting $q_1 = 145$, implying offset $z = 45$.