# MATHEMATICAL PROGRAMMING BASED EXACT AND HEURISTIC SOLUTION APPROACHES FOR A CLUSTERING PROBLEM WITH LOCALIZED FEATURE SELECTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖZDENUR BÜYÜK HABACI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

AUGUST 2024

Approval of the thesis:

**MATHEMATICAL PROGRAMMING BASED EXACT AND HEURISTIC SOLUTION APPROACHES FOR A CLUSTERING PROBLEM WITH LOCALIZED FEATURE SELECTION**

submitted by **GÖZDENUR BÜYÜK HABACI** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Zeynep Pelin Bayındır
Head of Department, **Industrial Engineering** _____

Prof. Dr. Sinan Gürel
Supervisor, **Industrial Engineering, METU** _____

Prof. Dr. Cem İyigün
Co-supervisor, **Industrial Engineering, METU** _____

**Examining Committee Members:**

Assoc. Prof. Dr. Mustafa Kemal Tural
Industrial Engineering, METU _____

Prof. Dr. Sinan Gürel
Industrial Engineering, METU _____

Prof. Dr. Cem İyigün
Industrial Engineering, METU _____

Assoc. Prof. Dr. Hakan Çerçioğlu
Industrial Engineering, Gazi University _____

Assist. Prof. Dr. Nader Ghaffarinasab
Industrial Engineering, METU _____

Date: 22.08.2024

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Surname: Gözdenur Büyük Habacı

Signature :

# ABSTRACT

## MATHEMATICAL PROGRAMMING BASED EXACT AND HEURISTIC SOLUTION APPROACHES FOR A CLUSTERING PROBLEM WITH LOCALIZED FEATURE SELECTION

Habacı, Gözdenur Büyük
M.S., Department of Industrial Engineering
Supervisor: Prof. Dr. Sinan Gürel
Co-Supervisor: Prof. Dr. Cem İyigün

August 2024, 117 pages

Clustering is an unsupervised machine learning problem that is widely studied in different contexts of business and science. The complexity of real-world data, often characterized by high dimensionality, poses significant challenges to traditional clustering methods. Feature selection is the most used technique to cope with the high dimensionality in clustering problems. Most feature selection methods select a common set of features to define all clusters, which is called global feature selection. Localized feature selection methods consider that the relevant set of features may differ across the clusters and select a set of features for each cluster separately. In this thesis, we address a clustering problem that aims to group data points and select a cluster center and a set of relevant features for each cluster. The objective is to minimize the sum of Euclidean distances between data points and their cluster center over each cluster's relevant set of features. We propose two Mixed-Integer Second-Order Cone Programming formulations, a matheuristic method, and an iterative heuristic method for the problem. We present the computational performance of the proposed methods on generated data sets.

Keywords: clustering, localized feature selection, mixed integer second-order cone programming, matheuristics, heuristics

# ÖZ

## YERELLEŞTİRİLMİŞ ÖZELLİK SEÇİMİ İLE KÜMELEME PROBLEMİ İÇİN MATEMATİKSEL MODELLEMEYE DAYALI KESİN ÇÖZÜM VE SEZGİSEL ÇÖZÜM YAKLAŞIMLARI

Habacı, Gözdenur Büyük
Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Sinan Gürel
Ortak Tez Yöneticisi: Prof. Dr. Cem İyigün

Ağustos 2024 , 117 sayfa

Kümeleme, iş ve bilimin farklı alanlarında yaygın olarak incelenen bir makine öğrenme problemidir. Yüksek boyutlu gerçek dünya verilerinin karmaşıklığı, geleneksel kümeleme yöntemleri için önemli zorluklar oluşturur. Özellik seçimi, kümeleme problemlerindeki yüksek boyutlulukla başa çıkmak için en çok kullanılan yaklaşımdır. Çoğunlukla, tüm kümeleri tanımlamak için ortak bir özellik kümesi seçilir. Yerelleştirilmiş özellik seçimi yaklaşımı, her küme için farklı özelliklerin önemli olabileceğini dikkate alır ve her küme için ayrı bir özellik kümesi seçer. Bu tezde, veri noktalarını gruplamayı, her küme için bir küme merkezi ve ilgili özellik kümesi seçmeyi amaçlayan bir kümeleme problemini ele alıyoruz. Amaç, veri noktaları ile küme merkezleri arasındaki Öklid mesafelerinin toplamını, her kümenin ilgili özellik kümesi üzerinde en aza indirmektir. Problem için iki Karma Tamsayı İkinci Dereceden Konik Programlama formülasyonu, bir matsezgisel yöntem ve yinelemeli bir sezgisel yöntem öneriyoruz. Önerilen yöntemlerin üretilen veri kümeleri üzerindeki hesaplama performansını sunuyoruz.

Anahtar Kelimeler: kümeleme, yerelleştirilmiş özellik seçimi, karma tamsayı ikinci dereceden konik programlama, matsezgiseller, sezgiseller

To my family...

## ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my supervisor, Prof. Dr. Sinan Gürel, and my co-supervisor, Prof. Dr. Cem İyigün, for their guidance, encouragement, and continuous support. Their expertise and dedication played an instrumental role in shaping this thesis and my academic career.

I would also like to thank the examining committee members, Assoc. Prof. Dr. Mustafa Kemal Tural, Assoc. Prof. Dr. Hakan Çerçioğlu and Assist. Prof. Dr. Nader Ghaffarinasab, for their time and valuable comments, which have significantly contributed to my work.

I am grateful to my family for their continuous support and being always by my side. To my mother, father, and brother, thank you for your encouragement, understanding, and patience throughout this journey. The deepest thanks to my beloved husband for being with me in every step of my life. I couldn't have done this without you.

Lastly, I would like to acknowledge all of my friends, colleagues, and professors who have always supported me along the way.

## TABLE OF CONTENTS

## LIST OF TABLES

TABLES

## LIST OF FIGURES

FIGURES

## LIST OF ALGORITHMS

ALGORITHMS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| CLFS | Clustering with Localized Feature Selection |
| $CLFS^A$ | Data Point-Cluster Assignment Problem |
| $CLFS^{AF}$ | Data Point-Cluster Assignment & Feature Selection Problem |
| $CLFS^C$ | Cluster Center Selection Problem |
| $CLFS^F$ | Cluster Feature Selection Problem |
| MILP | Mixed-Integer Linear Programming |
| MINLP | Mixed-Integer Nonlinear Programming |
| MISOCP | Mixed-Integer Second-Order Cone Programming |
| $\%Gap_R$ | MIP Relative Gap |
| $\%Gap_{Avg}$ | Average Gap |
| $\%Gap_{Best}$ | Best Solution Gap |
| $\%Gap_{Worst}$ | Worst Solution Gap |
| $\%Eql$ | Percentage of Initializations that the Matheuristic and Heuristic Found the Same Solution |
| $\%Hrs$ | Percentage of Initializations that the Heuristic Found Better |
| $\%Mth$ | Percentage of Initializations that the Matheuristic Found Better |
| $\#Base$ | Number of Data Sets where the Base Objective Was Found |
| $\#Best$ | Number of Best Solutions |
| $\#IBase$ | Number of Initializations that the Base Objective Was Found |
| $\#Opt$ | Number of Data Sets where the Optimal Solution Was Found |
| $\#TL$ | Number of Data Sets that the Time Limit Was Exceeded |
| $\#Worst$ | Number of Worst Solutions |

# CHAPTER 1

# INTRODUCTION

Technological advances have led to enormous growth and diversity of data in all sectors. Organizations need to use efficient methods to analyze their data automatically and quickly to gain valuable insights for informed decision-making. Data mining has emerged as a field that eases the extraction of useful information from very large collections of data. Clustering, an unsupervised learning method, is widely used in data mining to find groups of similar data points. It has been applied in numerous areas, such as customer segmentation, event detection, and image processing.

Data collections are stored and analyzed as data sets, which can be structured in a tabular format. In data sets, rows correspond to data points, representing entities or objects in the problem, such as customers, students, or devices. Columns correspond to features, which are specific attributes or characteristics of data points, such as age, grade, or color. A cell in the data set, located at the intersection of the $i^{th}$ row and $j^{th}$ column, stores the value of feature $j$ for data point $i$. Clustering aims to group a given set of data points according to the similarities among their feature values.

One of the most important challenges in clustering problems is to extract useful information from high-dimensional data sets, i.e., data sets with many features. While some features are relevant to the clustering structure, others may be irrelevant or redundant. The redundant features may mask the hidden clustering structure and reduce the learning performance of clustering methods and their computational efficiency. As the number of features increases, the masking effect is observed increasingly. It becomes more difficult to understand the patterns among data points, called the "Curse of Dimensionality." To deal with this, feature selection is widely used in clustering problems.

Most feature selection methods proposed for clustering perform global feature selection, which selects a common subset of features for all clusters. This assumes the same set of features can be used to identify all clusters. However, the global selection of features oversimplifies the inherent complexity, as the set of relevant features for clustering may vary across clusters. Therefore, clustering methods that perform global feature selection may not identify the clusters in different feature subspaces. On the other hand, localized feature selection allows clustering methods to find clusters in different feature subspaces, where a set of features is selected for each cluster. Clustering with localized feature selection requires determining clusters of data points and choosing the relevant features of each cluster simultaneously.

This thesis considers a partitioning clustering problem where $n$ data points exist and $m$ features are measured. It is assumed that all feature values are continuous. The aim is to group (i.e., partition) $n$ data points into a predetermined number of clusters, $p$, to optimize a partitioning criterion. Each data point is assigned to one cluster, and a cluster center is selected from the data points within each cluster. The objective is to minimize the total Euclidean distance from data points to their corresponding cluster center. A set of $q$ most relevant features is selected for each cluster, and the Euclidean distance between the data points and the cluster center is calculated based on the selected set of features. The selected features may differ for each cluster, and a feature can be chosen for multiple or no clusters. We call this problem clustering with localized feature selection and abbreviate CLFS in the rest of the thesis.

The CLFS problem differs from the existing studies in the literature in that a localized feature selection is performed simultaneously with the clustering of data points. Each cluster is defined based on a different set of features. The objective of the problem also differs in that the total Euclidean distance from data points to their cluster center is aimed to be minimized. Euclidean distance is a more robust distance measure, especially in cases where extreme differences may occur between feature values of data points.

We provide two different Mixed Integer Nonlinear Programming (MINLP) formulations for the CLFS problem: a center-based and a cluster-based. We show that the proposed formulations can be reformulated as Mixed-Integer Second-Order Cone

Programming (MISOCP) models. Furthermore, we propose a matheuristic method combining mathematical programming with a heuristic approach and an iterative heuristic method to solve the CLFS problem. The computational performance of the proposed methods has been tested on generated data sets.

**Contribution of the Thesis:**

This thesis contributes to the literature in several aspects. First, it is the first study to address the clustering problem with localized feature selection, where the total Euclidean distance is the clustering criterion. Second, it proposes two MINLP formulations for the CLFS problem and shows the models can be reformulated as MISOCP models. Third, we developed a matheuristic method that combines mathematical programming with a heuristic approach for the CLFS problem. This is the first study to propose a matheuristic method for the clustering problem with feature selection. Last, we propose a heuristic method demonstrating a strong computational performance to solve the CLFS problem.

The thesis is organized as follows: Chapter 2 summarizes the clustering problem, feature selection, and localized feature selection in clustering with the related studies. Chapter 3 defines the CLFS problem and proposes mathematical formulations. Chapter 4 proposes a matheuristic method for the CLFS problem. Chapter 5 provides an iterative heuristic algorithm for the CLFS problem. Chapter 6 presents the results of computational experiments on generated data sets. Lastly, 7 provides the main findings from this thesis study and discusses the further improvement directions.

# CHAPTER 2

# LITERATURE REVIEW

Researchers from diverse disciplines, from statistics to genetics, have studied clustering problems for many years, which has led to the emergence of various techniques for clustering. One of the main challenges with clustering is extracting useful knowledge from data sets where some of the features are irrelevant or redundant to the cluster structure. Feature selection is the most common method to cope with redundant features in clustering problems.

This chapter reviews the literature on clustering problems, feature selection, and localized feature selection in clustering. Then, we summarize the literature on mathematical programming-based solution approaches for clustering problems with localized feature selection.

Section 2.1 overviews the clustering problem and common clustering approaches. Section 2.2 reviews the feature selection in clustering, while the global and localized feature selections are discussed in Section 2.3. This study proposes two mathematical models, a matheuristic and a heuristic method for clustering problems with localized feature selection. Therefore, the studies related to the mathematical programming, matheuristic, and heuristic approaches for clustering with localized feature selection are discussed in Section 2.4, 2.5, and 2.6 respectively.

## 2.1 Clustering Problems

Clustering aims to uncover the hidden clustering structure of a data set according to the similarities between data points. Following this, clustering methods group similar

data points in a cluster according to the relevant features while assigning different data points to separate clusters.

In a clustering problem, each feature stores information about a characteristic of the data points. To demonstrate, in a movie recommendation system, system users are data points, and attributes of users such as age, favorite genre, or average screen time are the features of the data points. Features can take both qualitative and quantitative values. In this example, age and the average screen time of the user are quantitative features that take discrete and continuous values, respectively. On the other hand, the user's favorite genre is a qualitative feature.

The proximity between data points can be defined using a distance function such as Manhattan, Euclidean, and Jaccard distances based on the feature type. Euclidean and Manhattan distances are mostly preferred to define similarity in continuous data sets, while Jaccard distance is generally selected for categorical and binary data sets. In our study, we work with continuous features and define the distance between data points based on Euclidean distance.

Another important issue for cluster analysis is determining the objective function. A clustering method should produce low similarity between clusters (separation) and high similarity within a cluster (compactness). Although separation and compactness are closely related performance measures, the objective function differs according to whether compactness or separation is aimed to be maximized. Also, the objective function differs according to how the separation and compactness are defined. For example, compactness can be measured using the average or maximum pairwise distance between the data points or the total distance between the center and the data points within the clusters. The objective function of the cluster analysis should be determined according to the nature of its application. In our study, we work on a clustering problem that aims to maximize the compactness of clusters measured by the proximity between the centers and data points.

Clustering has been applied in numerous areas, such as customer segmentation, event detection, image processing, and pattern recognition. Therefore, plenty of clustering techniques have been developed for different applications. The best-known clustering methods are hierarchical clustering and partitioning.

Hierarchical clustering provides a hierarchical structure of clusters represented with special graphs called dendrograms. Hierarchical clustering methods are either agglomerative or divisive. Agglomerative clustering methods initially consider each data point as a cluster. Then, they repeatedly combine the two nearest clusters into one cluster. Divisive clustering methods start with one cluster of all objects and divide it into sub-clusters recursively. The combination or division at each level continues until a stopping criterion is satisfied. BIRCH (Zhang et al., 1996), CURE (Guha et al., 1998), and ROCK (Guha et al., 2000) are examples of agglomerative clustering methods, while DIANA and MONA (Kaufman & Rousseeuw, 2009) are examples of divisive clustering methods. A drawback of hierarchical clustering is that it is computationally expensive since the pairwise distances between all cluster pairs are computed at each step. On the other hand, they result in a nested grouping of objects at different similarity levels, so they do not require the number of clusters in advance.

Partitioning (point assignment) clustering methods maintain a set of clusters by partitioning data points to a specified number of clusters to optimize a partitioning criterion. Each cluster is represented by a center. Partitioning methods start with the initialization of clusters with a random center. Then, they assign each data point to its nearest cluster and determine the centers of clusters again. Assignments and center selections continue until the clusters stabilize. The most popular partitioning methods are K-means (MacQueen et al., 1967) and k-medoids(PAM) (Kaufman & Rousseeuw, 2009).

Partitioning methods take less time and memory than hierarchical clustering methods. However, they require a predetermined number of clusters. In these methods, the quality of the solution depends on the initial cluster settings. In this thesis, we also consider a partitioning clustering problem where clusters are represented by their medoids (a data point closest to other data points within the cluster), and data points are partitioned into a predetermined number of clusters.

There are many other approaches for clustering, such as density-based, grid-based, and model-based clustering methods. Finding a common categorization for clustering techniques is difficult due to the diversity and overlapping nature of approaches. Reviews of clustering problems and clustering techniques are provided in Jain et al.,

1999, Rokach and Maimon, 2005, Xu and Wunsch, 2005, Jain, 2010 and Saxena et al., 2017.

Traditional clustering methods tend to consider all features in a data set to learn as much as possible from the data. However, not all features are always relevant to the clustering structure. The following section reviews dimension reduction and feature selection approaches in clustering.

## 2.2 Clustering with Feature Selection

High-dimensional data sets, data sets with a high number of features, present additional complexity to clustering problems. Among the numerous features in a data set, only some are relevant to the clustering structure, while others may be irrelevant or redundant. These irrelevant features can obscure the true clustering structure, reduce the effectiveness of clustering algorithms, and lower computational efficiency. As the number of features increases, the number of irrelevant features increases, and this masking effect is observed even more. This is referred to as the "Curse of Dimensionality."

Dimension reduction methods have emerged to decide which features should be considered while clustering data points. These methods are mainly categorized into feature selection and feature extraction. Feature selection is selecting a distinctive set of features, while feature extraction utilizes transformation and combining features into a lower dimensional feature space (Xu & Wunsch, 2005). Both dimension reduction methods improve learning performance and reduce computational complexity. However, feature selection provides more readable and interpretable results than feature extraction as it preserves the original features in the cluster analysis (Alelyani et al., 2018). Examples of feature extraction methods are Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA).

Feature selection methods are commonly used to improve efficiency and quality of clustering decisions, although it is difficult to define feature relevance using unlabeled data. Feature selection methods are generally categorized into filter, wrapper, and hybrid methods in the clustering literature. Filter methods select a relevant set of

features based on filtering criteria prior to clustering. In contrast, wrapper methods utilize a clustering algorithm to evaluate the sets of features using a heuristic search strategy. Filter methods are superior to wrapper methods in terms of computational efficiency, while wrapper methods produce features that produce better clustering performance. As can be understood from the name, hybrid methods combine filter and wrapper methods to take advantage of both. A review of feature selection methods in clustering is provided in Alelyani et al., 2018 and Hancer et al., 2020. This thesis follows wrapper approaches, i.e., it selects relevant features during clustering.

Feature selection can be performed considering the whole cluster structure or considering each cluster separately. The next section reviews the global feature selection and localized feature selection in clustering.

## 2.3 Clustering with Global versus Localized Feature Selection

Most feature selection methods proposed for clustering perform global feature selection, which selects a common set of features for all clusters. Global feature selection oversimplifies the inherent complexity, as the set of relevant features for clustering may vary across clusters. Therefore, clustering methods that perform global feature selection may not identify the clusters in different feature subspaces. Localized feature selection methods where relevant features are selected for each cluster separately have emerged to overcome this problem.

Clustering with localized feature selection aims to find clusters oriented in different feature subspaces. In these problems, data point assignments and relevant cluster features should be determined simultaneously since these decisions depend on each other. When assigning a data point to a cluster, the relevant set of features for a cluster is used to determine the similarities between data points. On the other hand, relevant features are selected for each cluster created.

Figure 2.1 illustrates the clustering without feature selection, clustering with global feature selection, and with localized feature selection. Figure 2.1 (a) shows a resulting clustering structure in a data set where each row represents a data point and each column corresponds to a feature. It is seen that the data points are grouped into two

clusters represented by different colors, considering their similarities according to all features in the data set. Figure 2.1 (b) illustrates the clustering problems with global feature selection where all data points are clustered according to a common set of features. Figure 2.1 (c) shows the clustering problems with localized feature selection where each cluster of data points can be identified using a different set of features.

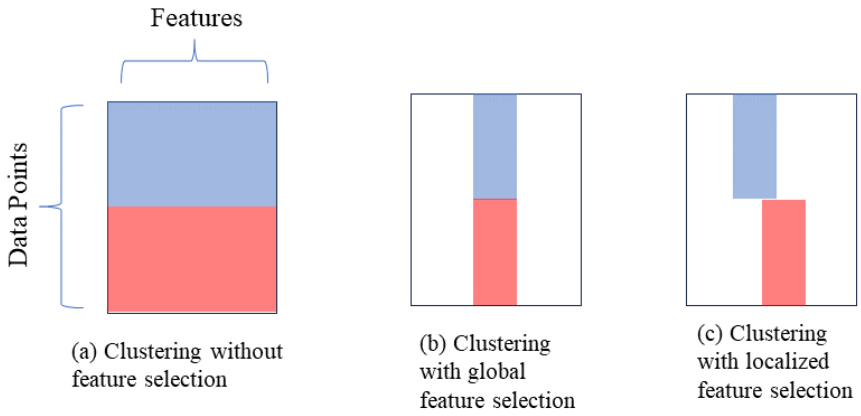


Figure 2.1: Clustering Problems According to Feature Selection

The first study to define localized feature selection in clustering is Li et al., 2008a. This study considers a clustering problem that aims to cluster data points and select relevant features for each cluster separately to optimize a clustering criterion, adjusted, and normalized scatter separability. Scatter separability covers both within-cluster similarity and between-cluster separability. Our study considers a clustering problem that aims to only maximize within-cluster similarity measured with the Euclidean distances between data points and their cluster centers. Cluster centers are also decided in our problem, in addition to clusters and relevant features. Our problem mainly differs from Li et al., 2008a in terms of the objective function and the selection of cluster centers.

In their later study, Li et al., 2008b propounded a new approach for clustering problems with localized feature selection. They define the problem from a statistical perspective and state that each cluster can be represented by a Gaussian distribution. The objective of the problem is to simultaneously determine distribution parameters and

relevant features for each cluster to maximize a likelihood function. They consider a model-based clustering problem in this study while we work on a partitioning clustering problem in ours. Our problem differs from Li et al., 2008b in terms of the clustering approach, the decision variables, and the objective function.

A closely related study to this thesis is by Öz, 2019. Öz, 2019 studied the clustering problem with localized feature selection that aims to determine clusters, cluster centers, and features to minimize the total Manhattan distance between the data points and their cluster centers. This problem is also studied by Şener, 2022. Unlike Öz, 2019 and Şener, 2022, we use Euclidean distance, the most commonly used similarity measure for continuous data, to describe the similarities between data points in our problem. It is worth noting that Euclidean distance introduces additional complexity to the problem since it is a nonlinear function.

Bi-clustering and subspace clustering approaches can also be related to clustering with localized feature selection problems, although they have different perspectives and goals. Kriegel et al., 2009 state that the primary challenge for clustering high-dimensional data is that different sets of features identify different clusters and call this phenomenon "local feature relevance". They propose to use bi-clustering and subspace clustering to cluster high-dimensional data with local feature relevance. Subspace clustering aims to uncover all clusters in different feature subspaces. Similarly, bi-clustering aims to identify submatrices of the data set based on the patterns between data points and features.

Our problem is similar to subspace clustering and bi-clustering in that a localized feature selection is performed. However, it differs from bi-clustering and subspace clustering in that we assign each data point to exactly one cluster. On the contrary, a data point can be assigned to multiple clusters or no cluster in subspace clustering and bi-clustering approaches.

Next, we present a literature review of mathematical programming-based exact solution approaches related to clustering problems with localized feature selection.

## 2.4 Mathematical Programming-Based Solution Approaches for Clustering with Localized Feature Selection

Partitioning clustering can be seen as an optimization problem in that a partitioning criterion is usually optimized by making assignment and center selection decisions. The clustering decisions are made under certain restrictions. First, each data point has to be assigned to one cluster. Second, a center may have to be selected for each cluster. Last, the selected center has to be one of the data points within the cluster.

The operations research community has studied the clustering problem, especially due to its similarity to location problems. If the customers are considered data points and the facilities are considered cluster centers, then the traditional p-median problem can be seen as a partitioning clustering problem. If the facilities will be located in one of the customer locations, then the problem is a K-medoids problem. However, there are only two features, the x-coordinate and y-coordinate, in the p-median problem, while there can be many features in a partitioning clustering problem. Also, both dimensions are considered in a p-median problem, while a feature selection can be performed for clustering.

Over the years, mathematical programming models have been proposed to obtain an exact solution for clustering and its applications. The key issues in these studies are definitions of problem setting, objective function, and decision variables.

Vinod, 1969 is one of the first studies that show that Integer Programming (IP) can be used to formulate clustering problems. First, they studied a clustering problem similar to warehouse location problems. The objective of the problem is to minimize the total cost of assignment, where the assignment costs are predetermined for each data point. They developed an IP formulation for the problem. Second, they studied a clustering problem that aims to minimize the within-group sum of square distances between centers and data points. They developed an IP formulation and alternative linearizations for the problem. In the following years, Rao, 1971 extend the study of Vinod, 1969 by providing different objective function formulations for the clustering problem. They modify the mathematical models for minimizing the sum of average, total, and maximum within-group distances.

The contribution of mathematical programming to the clustering field is not limited to the early work. Mathematical programming-based exact solution approaches have been proposed for diverse clustering problems. Olafsson et al., 2008 reviews the intersection of data mining and operations research by discussing operations research's contributions to the clustering field.

In these studies, the cluster structure is identified using all features in the data set. However, as discussed before, redundant and irrelevant features should usually be eliminated, especially from high-dimensional data sets, to increase the clustering performance.

Benati and Garcıa, 2014 studied a clustering problem with feature selection. Their problem aims to minimize the total Manhattan distance between data points and their corresponding cluster centers by determining clusters, centers, and relevant features. They developed a mixed integer nonlinear programming (MINLP) model and its two linearizations simultaneously performing feature selection and clustering. Their models perform a global feature selection where a common set of relevant features is selected to define clusters.

In their later study, Benati et al., 2018 considered a clustering problem with feature selection where the cluster centers are predetermined. The objective is the same as the problem in Benati and Garcıa, 2014, and they perform a global feature selection. They proposed two MILP formulations for this centers-given clustering problem. Compared to our problem, both Benati and Garcıa, 2014 and Benati et al., 2018 use global feature selection instead of localized feature selection, and the proposed models do not support using Euclidean distance as a similarity measure. Also, the models in Benati and Garcıa, 2014 are proposed for binary data sets while we focus on continuous data sets.

As discussed before, Öz, 2019 studied a clustering problem with localized feature selection. Their problem aims to determine clusters, cluster centers and relevant features to minimize the total Manhattan distance from data points to their cluster center. They proposed a MINLP model and its three linearizations for the clustering problem with localized feature selection where relevant features are selected for each cluster separately. A limitation of this study is the proposed linearizations are not applicable

when Euclidean distance is selected for cluster analysis. In this thesis, we use Euclidean distance to define the distance between data points. This led to the MINLP formulation of the problem, which we showed to be representable by second-order conic inequality. To take the comparison further, Öz, 2019 proposes only a center-based formulation of the problem, while we propose both center-based and cluster-based formulations.

In this thesis, we studied the clustering problem with localized feature selection. The problem aims to minimize the total Euclidean distance between data points and their corresponding center by determining clusters, cluster centers, and relevant features selected for each cluster separately. We propose two different MINLP formulations for the problem and show that they can be reformulated as MISOCP models. To our knowledge, this is the first study that uses a SOCP approach to formulate the clustering problem with feature selection.

Mathematical programming-based exact solution approaches provide high-quality clustering solutions for smaller data sets. However, solving mathematical models with existing solvers may not be practical for large-sized problem instances. Matheuristic methods have been developed to solve large-size optimization problems. The next section provides the literature review for the matheuristic approaches related to clustering with localized feature selection.

## 2.5 Matheuristics for Clustering with Localized Feature Selection

Matheuristic methods that combine mathematical programming with heuristic approaches are used in operations research to solve complex optimization problems. These methods take advantage of both accuracy and computational efficiency. Although numerous mathematical programming models have been proposed for solving complex clustering problems, the number of studies related to matheuristics is limited.

Gnägi and Baumann, 2021 studied a capacitated clustering problem. The problem aims to minimize the total distance from centers to their assigned data points where each data point has a weight, and the total weight in each cluster is restricted. They

developed a matheuristic method for this capacitated clustering problem. The idea of the matheuristic is to solve a linear program for the assignment problem with fixed centers subject to capacity constraints on each sub-problem and update the centers between the solution of two consecutive sub-problems.

Our study provides a matheuristic method that combines MISOCP models with a heuristic algorithm to solve large-scale CLFS problems. To the best of our knowledge, this is the first study that proposes a matheuristic method for the clustering problem with feature selection.

Matheuristic methods combine mathematical models with heuristic approaches to decrease the computation time of the models. However, solving mathematical models can be costly for even less complex problems with large data sets. Heuristic solution approaches are commonly used for clustering problems to obtain fast clustering solutions for large-sized problems, especially in real-life applications. The next section provides a literature review of the heuristic approaches for clustering with localized feature selection.

## 2.6 Heuristics for Clustering with Localized Feature Selection

Many heuristic methods are developed based on various strategies to solve complex clustering problems by trading off optimality for speed and practicability. The most known heuristic methods for partitional clustering are K-means (MacQueen et al., 1967) and K-medoids (PAM) (Kaufman & Rousseeuw, 2009).

K-means has the idea of deciding assignments of data points to clusters by fixing the centers and then selecting centers for fixed assignments iteratively until the decisions of assignments and centers do not change. Centers are the mean of the data points in the clusters called centroids. On the other hand, PAM considers centers as medians of data points within the clusters. Therefore, PAM is less precise to outliers than K-means. Many improvements have been proposed for both K-means and K-medoids over the years. For example, Park and Jun, 2009 proposed a heuristic algorithm by improving the initial selection of centers in K-medoids. These algorithms decide the clustering structure using all features in the data set without feature selection. How-

ever, feature selection methods have also been proposed for partitioning clustering.

Brusco and Cradit, 2001 presents a feature selection heuristic for K-means based on the adjusted Rand index. The proposed heuristic method iteratively evaluates each feature's contribution to the clustering structure by comparing the clustering results with and without this feature. At each step, unselected features are evaluated based on the adjusted Rand index, and features that improve the index are added to the set of selected features.

Ólafsson and Yang, 2005 proposes a partitioning method for feature selection in clustering. They use the nested partitions method, which partitions the feature space into sets and iteratively evaluates their contributions to the clustering quality. In their later study, Yang and Olafsson, 2006 improves the prior study with an adaptive sampling of features in each step.

Steinley and Brusco, 2008 presents a feature weighting and selection heuristic for K-means clustering based on variance-to-range ratio. The proposed algorithm iteratively selects the most informative set of features by assigning weights and evaluates the clustering solution.

Li et al., 2008a studied a clustering problem with localized feature selection where an index that covers both within-cluster similarity and between-cluster separability is aimed to be optimized. They developed a heuristic algorithm to solve this problem. The algorithm computes adjusted and normalized scatter separability and conducts a sequential backward search to find the set of relevant features for each cluster separately.

Benati et al., 2018 studied the clustering problem with localized feature selection where the cluster centers are given. They developed two heuristic algorithms for the problem. They define the feature selection problem as finding a set of features such that the total Manhattan distance from data points to their closest center is minimized. One of these heuristics, called q-vars, outperforms the other proposed heuristic. The main idea of q-vars is to iteratively assign data points to centers and select relevant features until no further improvement in the objective function is observed. Also, they show the application of the q-vars algorithm to K-means clustering. In the heuristics

of feature selection in clustering described so far, a global set of features is selected to define the distance between data points.

Şener, 2022 studied clustering problems with localized feature selection, especially for large data sets. They developed a metaheuristic framework for the problem. The proposed metaheuristic combines a genetic approach with a neighborhood search. They also modified the metaheuristic for the high dimensional data sets. Their problem differs from ours in that they aim to minimize a partitioning criterion based on the Manhattan distance between data points. Furthermore, they propose a metaheuristic approach, while we propose mathematical programming-based approaches to the problem.

Öz, 2019 studied a clustering problem with localized feature selection using Manhattan distance. They developed two different heuristics for the problem. The first heuristic algorithm uses a mathematical model to decide relevant features and centers for each cluster. Then, the data points are assigned to their closest centers. The second heuristic algorithm iteratively decides assignments, centers, and local feature relevance by fixing the other two decisions using different algorithms.

This thesis proposes a heuristic method that iteratively decides assignments, centers, and local feature relevance by fixing the other two decisions. However, our problem differs from Öz, 2019 in that Euclidean distance is used to determine the distance between data points. Also, the heuristic algorithm and the feature selection subproblem differ from Öz, 2019. The feature selection sub-problem is an NP-hard problem in our context, while it can be solved with a basic heuristic algorithm in Öz, 2019.

To summarize, clustering is a widely studied problem in many disciplines, and many approaches and techniques have been developed. This study considers a partitioning clustering approach where each data point is assigned to one cluster, and clusters are represented by their centers. Feature selection methods have been developed to select relevant features to define cluster structure among many features in a data set. Localized feature selection approaches select relevant features for each cluster separately to obtain cluster structures in different feature subspaces. The number of studies on clustering with localized feature selection is very limited in the literature. The most closely related study to our problem is Öz, 2019.

Our study differs from Öz, 2019 in four ways. First, their problem is based on the Manhattan distance, and their proposed mathematical models cannot be applied to Euclidean distance. Second, we propose a cluster-based approach in formulations in addition to center-based approaches, which is also considered by Öz, 2019. Third, we propose a matheuristic method for the clustering problem with localized feature selection. Fourth, our proposed heuristic algorithm differs from that proposed by Öz, 2019.

This thesis contributes to the clustering literature in several ways. First, it formulates and solves a clustering problem with localized feature selection. It is the first study where the total Euclidean distance is the clustering criterion in a clustering problem with feature selection. Second, we propose two MINLP formulations for the problem and show the models can be reformulated as MISOCP. Third, we propose a matheuristic method for the problem as the first matheuristic approach proposed for a clustering problem with feature selection. Last, we propose a heuristic method for the large-size applications of the problem.

The next chapter explains the problem and two MINLP formulations with their conversions to MISOCP.

# CHAPTER 3

# PROBLEM DEFINITION AND MATHEMATICAL FORMULATION

In Chapter 2, we introduced brief information about clustering and feature selection in clustering with the related literature to the clustering with localized feature selection. This chapter defines the studied problem and proposes two mathematical formulations.

## 3.1 Problem Definition

Suppose that $n$ data points exist, and for each data point $i$, $m$ features are measured. $v_{ik}$ is the value of feature $k$ for data point $i$. Assume that all $v_{ik}$ values are continuous. From the given data points, $p$ clusters will be formed. Each point will be assigned to exactly one cluster. A center will be selected for each cluster from the data points within the cluster. For each cluster, $q$ relevant features will be selected. The features selected may differ for each cluster, and a feature can be chosen for multiple clusters or no clusters at all. The distance between data points within a cluster will be determined according to the selected features using Euclidean distance (L2-norm). The distance between the data points $i$ and $j$ based on feature $k$ is $|v_{ik} - v_{jk}|$. The overall distance between data points $i$ and $j$ within a cluster is $\sqrt{\sum_{k \in F}(v_{ik} - v_{jk})^2}$ where $F$ is the set of relevant features for the cluster. The objective is to minimize the total distance between data points and their corresponding cluster center. We name this problem clustering with localized feature selection (CLFS).

The decisions to solve the CLFS problem can be categorized into three groups. The first decision group assigns data points to clusters to form clusters of data points. The second group selects a data point as the center for each cluster. These decisions affect

the objective function in the sense that the distance from the data point is measured to the center of the cluster to which the data point is assigned. The third decision group selects relevant features for each cluster. These decisions also affect the objective function because the distance between data points and their corresponding center is determined according to the selected features of the cluster.

One of the main challenges with this problem is that the number of feasible solutions rapidly grows with respect to the data size. For a CLFS problem specified with a number of clusters $p$, a number of data points $n$, a number of features $m$, and a number of selected features $q$, the number of feasible solutions can be expressed as

$$\binom{n}{p} \cdot \binom{m}{q}^{p} \cdot p^{n-p} \tag{3.1}$$

The first term represents the number of feasible solutions while selecting $p$ cluster centers out of $n$ data points. The second term is the number of possible feature selections for $p$ clusters. The last term counts the possible number of assignments to cluster centers. It is crucial to note that the last term significantly contributes to the growth of feasible solutions since it grows exponentially to $n$. Table 3.1 shows the number of feasible solutions for different settings of the CLFS problem. It can be seen that the number of solutions is high for even small-size problems, and it rapidly grows with the number of clusters, features, and data points. Therefore, it is computationally costly to explore all feasible solutions to find the optimal solution for the CLFS problem.

We propose two mathematical formulations for the CLFS problem. First, clusters can be represented by their centers, and assignments and feature selections can be made for the centers. Second, clusters can be represented by a cluster index, and all three groups of decisions can be made directly for the relevant cluster indices. We call these formulations the center-based approach and cluster-based approach, respectively. The following sections will explain the mathematical models according to center-based and cluster-based approaches.

Table 3.1: Number of Feasible Solutions According to $p$, $m$, $q$ and $n$

| p | m | q | n | Number of Feasible Solutions |
|---|---|---|---|---|
| 2 | 4 | 2 | 10 | $4.15x10^{5}$ |
| 2 | 4 | 2 | 20 | $1.79x10^{9}$ |
| 2 | 4 | 2 | 40 | $7.72x10^{15}$ |
| 2 | 6 | 2 | 10 | $2.59x10^{6}$ |
| 2 | 6 | 2 | 20 | $1.12x10^{10}$ |
| 2 | 6 | 2 | 40 | $4.82x10^{16}$ |
| 2 | 6 | 3 | 10 | $4.61x10^{6}$ |
| 2 | 6 | 3 | 20 | $1.99x10^{10}$ |
| 2 | 6 | 3 | 40 | $8.58x10^{16}$ |
| 3 | 4 | 2 | 10 | $5.67x10^{7}$ |
| 3 | 4 | 2 | 20 | $3.18x10^{13}$ |
| 3 | 4 | 2 | 40 | $9.61x10^{23}$ |
| 3 | 6 | 2 | 10 | $8.86x10^{8}$ |
| 3 | 6 | 2 | 20 | $4.97x10^{14}$ |
| 3 | 6 | 2 | 40 | $1.5x10^{25}$ |
| 3 | 6 | 3 | 10 | $2.1x10^{9}$ |
| 3 | 6 | 3 | 20 | $1.18x10^{15}$ |
| 3 | 6 | 3 | 40 | $3.56x10^{25}$ |

## 3.2 A Center-Based Formulation

In the center-based formulation, clusters are represented by centers without an index of the cluster. The decision variable $y_j$ determines whether a data point $j$ is selected as the center or not. The decision variable $x_{ij}$ is introduced to determine if data point $i$ is assigned to the cluster with center $j$ or not. Another decision variable, $z_{jk}$, is used to determine whether feature $k$ is selected for the cluster of center $j$ or not. The center-based mathematical model is constructed as follows:

$$(F1) \quad \min \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot z_{jk}} \cdot x_{ij} \tag{3.2}$$

s.t.

$$x_{ij} \leq y_j \qquad \forall i, j \tag{3.3}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i \tag{3.4}$$

$$\sum_{j=1}^{n} y_j = p \tag{3.5}$$

$$\sum_{k=1}^{m} z_{jk} = q \cdot y_j \qquad \forall j \tag{3.6}$$

$$x_{ij} \geq 0 \qquad \forall i, j \tag{3.7}$$

$$y_j \in \{0, 1\} \qquad \forall j \tag{3.8}$$

$$z_{jk} \in \{0, 1\} \qquad \forall j, k \tag{3.9}$$

The objective is to minimize the total Euclidean distance between data points and their corresponding cluster center over the selected features for each cluster. Constraint (3.3) imposes that the data point $i$ can only be assigned to center $j$ if $j$ is selected as the center of a cluster. Constraint (3.4) establishes that every data point is assigned to one center. Constraint (3.5) sets the number of centers to the number of clusters, $p$. Constraint (3.6) provides that a feature $k$ can be selected for a cluster with center $j$ only if $j$ is chosen as a center while limiting the total number of features selected for a cluster to $q$. Constraints (3.8) and (3.9) imply that the decision variables $y_j$ and $z_{jk}$ are binary variables. On the other hand, the binary decision variable $x_{ij}$ is relaxed to the non-negative continuous variable at constraint (3.7) thanks to Property 1.

**Observation 1.** *$F1$ has an optimal solution where all $x_{ij}$ variables are binary.*

*Proof.* Consider solving $F1$ for given feasible $y_j$ and $z_{jk}$ values. The resulting problem is a linear program and is equivalent to assigning each data point $i$ to its closest cluster center $j$, which implies an optimal solution with 0-1 $x_{ij}$ values for the linear program. This observation is also valid for solving $F1$ with the optimal values of $y_j$ and $z_{jk}$. □

$F1$ is a mixed integer nonlinear programming (MINLP) model with a nonlinear objective function (3.2). In Property 1, we show that $F1$ can be reformulated as a MISOCP formulation that can be solved by off-the-shelf Branch and Bound solvers such as CPLEX and Gurobi.

**Property 1.** *$F1$ can be reformulated as a mixed integer second-order cone programming (MISOCP) model.*

*Proof.* A nonnegative auxiliary variable $w_{ijk}$ can be introduced to replace the nonlinear term, $z_{jk} \cdot x_{ij}$, in objective function (3.2) by adding the following constraints to $F1$:

$$x_{ij} + z_{jk} - 1 \leq w_{ijk} \quad \forall i, j, k \tag{3.10}$$

$$w_{ijk} \geq 0 \quad \forall i, j, k \tag{3.11}$$

$w_{ijk}$ takes the value of 1 if $x_{ij}$ and $z_{jk}$ take the value of 1, 0 otherwise. Then, the objective function (3.2) can be reformulated as

$$\min \quad \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\sum_{k=1}^{m} \left(v_{ik} - v_{jk}\right)^2 \cdot w_{ijk}} \tag{3.12}$$

Multiplication of variables after adding auxiliary variable $w_{ijk}$ is averted; however, the objective function (3.12) is still non-linear and non-convex due to the root of the sum of squares.

A nonnegative decision variable $d_{ij}$ can be introduced to represent the Euclidean distance realized between data point $i$ and the center $j$ with the following constraint:

$$\sqrt{\sum_{k=1}^{m} \left(v_{ik} - v_{jk}\right)^2 \cdot w_{ijk}^2} \leq d_{ij} \quad \forall i, j \tag{3.13}$$

$d_{ij}$ takes the value of Euclidean distance between $i$ and $j$ if data point $i$ is assigned to center $j$, 0 otherwise. Since the decision variable $w_{ijk}$ only takes the value 1 and 0, it can be squared in the Euclidean distance function to convert the non-convex function into a second-order conic inequality. By doing these, the objective function (3.12) could be reconstructed linearly as follows:

$$\min \quad \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} d_{ij} \tag{3.14}$$

□

The center-based MISOCP model is called $F1_{MISOCP}$ and is given below.

$$(F1_{MISOCP}) \quad \min \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} d_{ij} \tag{3.14}$$

s.t.

$$x_{ij} + z_{jk} - 1 \leq w_{ijk} \qquad \forall i, j, k \tag{3.10}$$

$$w_{ijk} \geq 0 \qquad \forall i, j, k \tag{3.11}$$

$$\sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot w_{ijk}^2} \leq d_{ij} \quad \forall i, j \tag{3.13}$$

(3.3), ..., (3.9)

Second-order Order Cone Programming (SOCP) models are convex optimization models with a linear objective function that allows for affine combinations of variables constrained over the intersection of second-order (Lorentz) cones. The next section proposes the cluster-based formulation for the CLFS problem.

### 3.3 A Cluster-Based Formulation

In the cluster-based formulation, clusters are represented by the set $l = 1, 2, ..., p$. The decision variable $x_{il}$ represents whether the data point $i$ is assigned to cluster $l$ or not. The decision variable $y_{jl}$ determines whether a data point $j$ is selected as the center for cluster $l$ or not. Selection of the relevant features of clusters is performed with the decision variable $z_{lk}$, which takes the value of 1 if feature $k$ is selected for cluster $l$, 0 otherwise.

The cluster-based mathematical model is constructed as follows:

$$(F2) \quad \min \quad \sum_{l=1}^{p} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot z_{lk} \cdot x_{il} \cdot y_{jl}} \tag{3.15}$$

s.t.

$$\sum_{l=1}^{p} x_{il} = 1 \qquad \forall i \tag{3.16}$$

$$y_{jl} \leq x_{jl} \qquad \forall j, l \tag{3.17}$$

$$\sum_{j=1}^{n} y_{jl} = 1 \qquad \forall l \tag{3.18}$$

$$\sum_{k=1}^{m} z_{lk} = q \qquad \forall l \tag{3.19}$$

$$y_{jl} \in \{0, 1\} \qquad \forall j, l \tag{3.20}$$

$$z_{lk} \in \{0, 1\} \qquad \forall l, k \tag{3.21}$$

(3.15) gives the objective function of the problem. Constraint (3.16) states that a data point $i$ must be assigned to one cluster. Constraint (3.17) forces that a data point $j$ can be selected as the center for cluster $l$ if it is assigned to that cluster. Constraint (3.18) states that one cluster center must be selected for each cluster. (3.19) forces that $q$ features to be selected for each cluster. Constraints (3.20) and (3.21) are the set constraints.

**Observation 2.** *$F2$ has an optimal solution where all $x_{il}$ variables are binary.*

*Proof.* Consider solving $F2$ for given feasible $y_{jl}$ and $z_{lk}$ values. The resulting problem is a linear program and is equivalent to assigning each data point $i$ to the cluster $l$ with the closest cluster center, which implies an optimal solution with 0-1 $x_{il}$ values for the linear program. This observation is also valid for solving $F2$ with the optimal values of $y_{jl}$ and $z_{lk}$. □

$F2$ is a mixed integer nonlinear programming (MINLP) model with a nonlinear objective function (3.15).

**Property 2.** *$F2$ can be reformulated as a mixed integer second-order cone programming (MISOCP) model.*

*Proof.* A nonnegative auxiliary variable $w_{ijlk}$ can be introduced to replace the nonlinear term, $z_{lk} \cdot x_{il} \cdot y_{jl}$, in objective function (3.15) by adding the following constraints to $F2$:

$$x_{il} + y_{jl} + z_{lk} - 2 \leq w_{ijlk} \quad \forall i, j, l, k \tag{3.22}$$

$$w_{ijlk} \geq 0 \quad \forall i, j, l, k \tag{3.23}$$

$w_{ijlk}$ takes the value of 1 if $x_{il}$, $y_{jl}$ and $z_{lk}$ take the value of 1, 0 otherwise. Then, the objective function (3.15) can be reformulated as

$$\min \quad \sum_{l=1}^{p} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot w_{ijlk}} \tag{3.24}$$

Multiplication of variables after adding auxiliary variable $w_{ijlk}$ is averted; however, the objective function (3.24) is still non-linear and non-convex due to the root of the sum of squares.

A nonnegative decision variable $d_{ij}$ can be introduced to represent the Euclidean distance realized between data point $i$ and the center $j$ with the following constraint:

$$\sqrt{\sum_{l=1}^{p} \sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot w_{ijlk}^2} \leq d_{ij} \quad \forall i, j \tag{3.25}$$

$d_{ij}$ takes the value of Euclidean distance between $i$ and $j$ if data point $i$ is assigned to the cluster with center $j$, 0 otherwise. Since the decision variable $w_{ijlk}$ only takes the value 1 and 0, it can be squared in the Euclidean distance function to convert the non-convex function into a second-order conic inequality. By doing these, the objective function (3.24) could be reconstructed linearly as follows:

$$\min \quad \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} d_{ij} \tag{3.26}$$

□

A deficiency of the cluster-based formulation is that it involves symmetry in that the cluster indices can be permuted in any feasible solution. In a clustering problem with $p$ clusters, assigning cluster indices to the clusters yields $p!$ alternative feasible

solutions for the same cluster structure with selected centers and relevant features. In other words, for a given solution, $p!$ alternative feasible solutions with the same objective function value are obtained by re-indexing the clusters. Symmetry in the formulation causes the solutions in the Branch and Bound tree to be duplicated when searching for feasible solutions, increasing the computational time of the model.

**Property 3.** *The symmetry in clustering solutions in $F2$ can be broken via the following constraint:*

$$\left(\sum_{j=1}^{n} j \cdot y_{jl}\right) + 1 \leq \sum_{j=1}^{n} j \cdot y_{jl+1} \quad l = 1, 2, ..., p-1 \tag{3.27}$$

*Proof.* Constraint (3.27) forces to assign the lower cluster index to the cluster with the center that has a lower point index. This constraint ensures that each cluster's center index is smaller than the index of the next cluster's center. Since each cluster center has a different point index, only one possible assignment of cluster indices exists for a particular cluster structure. □

Cluster-based MISOCP model is called $F2_{MISOCP}$ and is given below.

$$(F2_{MISOCP}) \quad \min \quad \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} d_{ij} \tag{3.26}$$

s.t.

$$x_{il} + y_{jl} + z_{lk} - 2 \leq w_{ijlk} \qquad \forall i, j, l, k \tag{3.22}$$

$$w_{ijlk} \geq 0 \qquad \forall i, j, l, k \tag{3.23}$$

$$\sqrt{\sum_{l=1}^{p} \sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot w_{ijlk}^2} \leq d_{ij} \quad \forall i, j \tag{3.25}$$

$$\left(\sum_{j=1}^{n} j \cdot y_{jl}\right) + 1 \leq \sum_{j=1}^{n} j \cdot y_{jl+1} \qquad l = 1, 2, ..., p-1 \tag{3.27}$$

(3.16), ..., (3.21)

The performance of $F1_{MISOCP}$ and $F2_{MISOCP}$ are tested and compared on some generated data sets. The details of the computational study are discussed in Chapter

6. It is examined that both models can find the optimal solution within the time limit (1 hour) for smaller data sets regarding the number of data points and features. However, the models' computation times accelerate, and the solution quality found within the time limit deteriorates when data sets grow. Therefore, we developed a matheuristic method to solve the CLFS problem with less computation time on larger data sets. The proposed matheuristic method is explained in the following chapter.

# CHAPTER 4

# A MATHEURISTIC METHOD FOR THE CLFS PROBLEM

In Chapter 3, two different MISOCP formulations are proposed for the CLFS problem. Although both models effectively solve CLFS problems for smaller data sets, the computation time of the models accelerates with the growth of the number of data points, features, and clusters. Matheuristic methods that combine mathematical programming with heuristic approaches are used to solve complex optimization problems. This chapter proposes a matheuristic method to solve CLFS problems.

As mentioned before, the CLFS problem has three groups of decisions. The first group determines each data point's cluster. The second group selects a center for each cluster from the data points within the cluster. The third group selects a set of relevant features for each cluster. The $F1_{MISOCP}$ and $F2_{MISOCP}$ models simultaneously give all three groups of decisions, resulting in many decision variables and, therefore, high computation time. We know that if centers are given for a CLFS problem, the remaining problem is to select a set of features for each cluster and assign each data point to the cluster with the closest center according to the selected features. Also, if clusters of data points and their relevant set of features are given, the problem is only selecting the center for each cluster, which is the data point closest to others within the cluster.

In this study, our approach is decomposing the three groups of decisions into two subproblems so that the subproblems can be solved with less computation time. The first subproblem is the Assignment & Feature Selection ($CLFS^{AF}$) problem. $CLFS^{AF}$ problem aims to give the assignment and feature selection decisions in the CLFS problem when the centers are known. The second subproblem is the Center Selection ($CLFS^{C}$) problem. $CLFS^{C}$ problem aims to determine the centers in the CLFS

problem when the assignments and relevant features of clusters are known.

The remainder of the chapter is organized as follows: Section 4.1 proposes the $CLFS^{AF}$ problem. Section 4.2 proposes $CLFS^{C}$ problem. Lastly, Section 4.3 explains the algorithm of the matheuristic method.

## 4.1 Assignment & Feature Selection Problem ($CLFS^{AF}$)

In Assignment & Feature Selection Problem ($CLFS^{AF}$), cluster centers, $C$, are given. The problem is assigning data points to clusters and selecting a set of $q$ relevant features for each cluster to minimize the total distance between the points and their cluster centers.

A center-based approach can be used to formulate the $CLFS^{AF}$ problem:

$$(F1^{AF}) \quad \min \quad \sum_{i=1}^{n} \sum_{j \in C} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot z_{jk} \cdot x_{ij}} \tag{4.1}$$

s.t.

$$\sum_{j \in C} x_{ij} = 1 \qquad \forall i \tag{4.2}$$

$$\sum_{k=1}^{m} z_{jk} = q \qquad j \in C \tag{4.3}$$

$$x_{ij} \geq 0 \qquad j \in C \quad \forall i \tag{4.4}$$

$$z_{jk} \in \{0, 1\} \qquad j \in C \quad \forall k \tag{4.5}$$

The decision variable $x_{ij}$ indicates if the data point $i$ is assigned to cluster with center $j$ or not. The decision variable $z_{jk}$ shows if the feature $k$ is selected for the cluster with center $j$ or not. Constraint (4.1) is the objective function, which is the sum of Euclidean distances between data points and their corresponding centers. Constraint (4.2) forces each data point to be assigned to one center in $C$. Constraint (4.3) implies that $q$ features must be selected for each cluster center in $C$. Constraint (4.4) is the nonnegativity constraint for $x_{ij}$, and Constraint (4.5) is the set constraint for $z_{jk}$. Although $x_{ij}$ is defined as a nonnegative variable, it takes the value of one or zero in the optimal solution according to a similar approach in Property 1.

It can be seen that $F1^{AF}$ is a MINLP model with a nonlinear objective function. Property 1 states that $F1$ can be reformulated as a MISOCP model. Similarly, $F1^{AF}$ can be reformulated as a MISOCP model by taking a similar approach to Property 1.

The MISOCP reformulation, $F1^{AF}_{MISOCP}$, is given below.

$$(F1^{AF}_{MISOCP}) \quad \min \sum_{i=1}^{n} \sum_{j \in C} d_{ij} \tag{4.6}$$

s.t.

$$x_{ij} + z_{jk} - 1 \leq w_{ijk} \qquad j \in C \quad \forall i,k \tag{4.7}$$

$$w_{ijk} \geq 0 \qquad j \in C \quad \forall i,k \tag{4.8}$$

$$\sqrt{\sum_{k=1}^{m} (v_{ik} - v_{jk})^2 \cdot w_{ijk}^2} \leq d_{ij} \qquad j \in C \quad \forall i \tag{4.9}$$

(4.2), ..., (4.5)

First, a nonnegative auxiliary variable $w_{ijk}$ is introduced to replace the nonlinear term $z_{jk} \cdot x_{ij}$ in (4.1). Constraint (4.7) and Constraint (4.8) ensure that $w_{ijk}$ takes the value of 1 if both $x_{ij}$ and $z_{jk}$ have the value of 1, 0 otherwise. Second, a nonnegative decision variable $d_{ij}$ is introduced to replace the Euclidean distance realized between data point $i$ and its center $j$. SOC constraint (4.9) ensures that $d_{ij}$ take the value of Euclidean distance between data point $i$ and center $j$ when data point $i$ is assigned to center $j$, 0 otherwise. Then, $d_{ij}$ is placed in the objective function (4.6) instead of the Euclidean distance between data point $i$ and center $j$.

A cluster-based approach also can be used to formulate the $CLFS^{AF}$ problem:

$$(F2^{AF}) \quad \min \quad \sum_{l=1}^{p} \sum_{i=1}^{n} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{C_l k})^2 \cdot z_{lk}} \cdot x_{il} \tag{4.10}$$

s.t.

$$\sum_{l=1}^{p} x_{il} = 1 \qquad \forall i \tag{4.11}$$

$$\sum_{k=1}^{m} z_{lk} = q \qquad \forall l \tag{4.12}$$

$$x_{il} \geq 0 \qquad \forall i,l \tag{4.13}$$

$$z_{lk} \in \{0,1\} \qquad \forall l,k \tag{4.14}$$

$C_l$ represents the center index of cluster $l$. The decision variable $x_{il}$ indicates if the data point $i$ is assigned to cluster $l$ or not. The decision variable $z_{lk}$ shows if the feature $k$ is selected for cluster $l$ or not. Constraint (4.10) is the objective function, which is the sum of Euclidean distances between data points and the center of clusters they assigned. Constraint (4.11) forces each data point to be assigned to one cluster. Constraint (4.12) implies that $q$ features must be selected for a cluster $l$. Constraint (4.13) is the nonnegativity constraint for $x_{il}$, and Constraint (4.14) is the set constraint for $z_{lk}$. The nonnegative variable $x_{il}$ takes the value of one or zero in the optimal solution according to a similar approach to Property 2.

Since the cluster center representing each cluster is known in advance in the $CLFS^{AF}$ problem, it can be seen that both center-based and cluster-based approaches result in similar mathematical models. The only difference is that data points and features are assigned to the cluster center representing each cluster in $F1^{AF}$ while they are assigned to clusters represented by indices in $F2^{AF}$.

It can be seen that $F2^{AF}$ is also a MINLP model with a nonlinear objective function. Property 2 states that the $F2$ model can be reformulated as a MISOCP model. Similarly, $F2^{AF}$ can be reformulated as a MISOCP model by taking a similar approach as in Property 2.

The cluster-based MISOCP formulation for the $CLFS^{AF}$ problem, $F2^{AF}_{MISOCP}$, is given below.

$$(F2^{AF}_{MISOCP}) \quad \min \sum_{l=1}^{p} \sum_{i=1}^{n} d_{il} \tag{4.15}$$

s.t.

$$x_{il} + z_{lk} - 1 \leq w_{ilk} \qquad \forall i, l, k \tag{4.16}$$

$$w_{ilk} \geq 0 \qquad \forall i, l, k \tag{4.17}$$

$$\sqrt{\sum_{k=1}^{m} \left(v_{ik} - v_{C_l k}\right)^2 \cdot w_{ilk}^2} \leq d_{il} \quad \forall i, l \tag{4.18}$$

(4.11), ..., (4.14)

First, a nonnegative auxiliary variable $w_{ilk}$ is defined to avoid the multiplication of

$z_{lk}$ and $x_{il}$ in (4.10). Constraint (4.16) and Constraint (4.17) ensure that $w_{ilk}$ takes the value of 1 if both $x_{il}$ and $z_{lk}$ have the value of 1, 0 otherwise. Second, a nonnegative decision variable $d_{il}$ is introduced to replace the Euclidean distance realized between data point $i$ and the center of its cluster $l$. SOC constraint (4.18) ensures that $d_{il}$ take the value of Euclidean distance between data point $i$ and $C_l$ if data point $i$ is assigned to cluster $l$, 0 otherwise.

$F1^{AF}_{MISOCP}$ and $F2^{AF}_{MISOCP}$ can be solved using optimization solvers such as Gurobi and CPLEX. The next section proposes the second subproblem, Center Selection ($CLFS^C$).

## 4.2 Center Selection Problem ($CLFS^C$)

In the $CLFS^C$ problem, clusters of data points, $A$, and relevant features of clusters, $F$, are given in advance. The problem for each cluster is choosing the data point with the least total distance to other data points within the cluster as the cluster center. Since the problem is decomposed into clusters, center-based and cluster-based formulations are the same for each cluster.

The linear programming model for each cluster $l$ is given below.

$$(F^C) \quad \min \sum_{i \in A_l} \sum_{\substack{j \in A_l \\ j \neq i}} \sqrt{\sum_{k \in F_l} (v_{ik} - v_{jk})^2} \cdot y_j \tag{4.19}$$

s.t.

$$\sum_{j \in A_l} y_j = 1 \tag{4.20}$$

$$y_j \geq 0 \qquad j \in A_l \tag{4.21}$$

$A_l$ is the given set of data points assigned to cluster $l$; where $F_l$ is the given set of relevant features for cluster $l$. The decision variable for center selection, $y_j$, takes the value of 1 if data point $j$ is the cluster center and 0 otherwise. The objective function (4.19) is the total distance between the cluster center and the data points within the cluster. $\sqrt{\sum_{k \in F_l} (v_{ik} - v_{jk})^2}$ is a parameter for each pair of data points $i$ and $j$. Therefore, the objective function (4.19) is a linear combination of decision variables $y_j$. Constraint (4.20) implies that one center has to be selected for the cluster. Con-

straint (4.21) states that $y_j$ takes a nonnegative value. Since the coefficient matrix in constraint (4.20) is a vector of 1's, the determinant of each square submatrix obtained from it is equal to 1. Therefore, the coefficient matrix is totally unimodular, yielding an integer optimal solution for $y_j$.

The $F^C$ model can be solved with optimization solvers such as Gurobi and CPLEX. The objective value of the $CLFS^C$ problem is the summation of the objective values of $F^C_{MISOCP}$ models of $p$ clusters.

Until now, we have discussed the mathematical formulations for the described subproblems of the matheuristic method. The next section will explain how these subproblems are used in an algorithm to solve the CLFS problem.

## 4.3 A Matheuristic Algorithm for CLFS

The matheuristic method works based on solving $CLFS^{AF}$ and $CLFS^C$ iteratively to solve the CLFS problem. The output of one subproblem serves as the input to the other. In this way, assignment and feature selection decisions are made by fixing cluster centers, and center selection decisions are made by fixing assignment and relevant features in the CLFS problem.

Figure 4.1 demonstrates subproblems and their relationships in the matheuristic method.

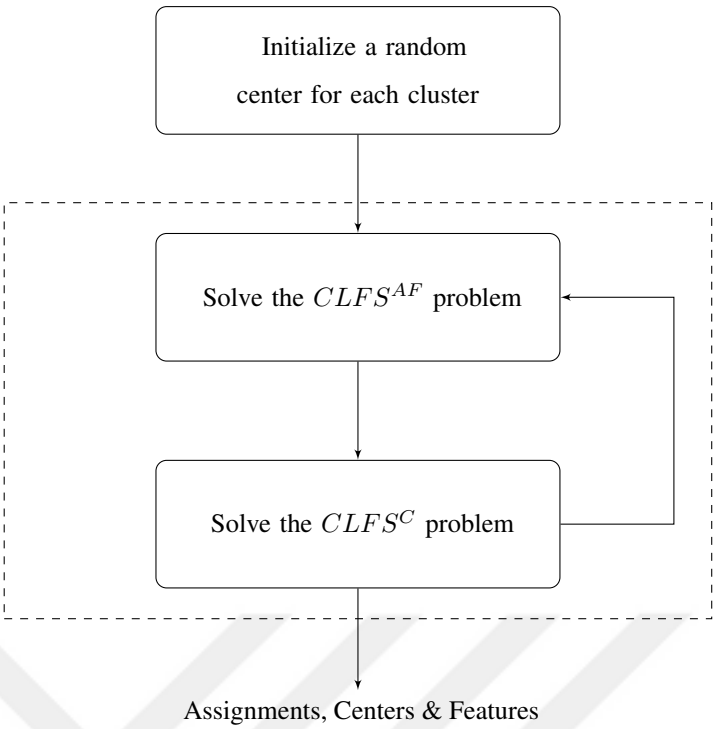


Figure 4.1: Representation of the Matheuristic Method for CLFS

The algorithm begins with the random initialization of cluster centers. First, $CLFS^{AF}$ is solved for given centers. Assignments and relevant features for each cluster are obtained. Next, $CLFS^{C}$ is solved according to these assignments and features, and cluster centers are determined. Then, $CLFS^{AF}$ is solved for the cluster centers obtained from the $CLFS^{C}$ problem. The iterations continue until the objective values of $CLFS^{AF}$ and $CLFS^{C}$ converge to the same value.

$F1^{AF}_{MISOCP}$ or $F2^{AF}_{MISOCP}$ is used to solve the $CLFS^{AF}$ problem and $F^{C}$ is used to solve the $CLFS^{C}$ problem. An optimization solver can be utilized to find the optimal solution for both problems.

The algorithm guarantees convergence of objective values across iterations because the next solution gives an objective value at least as good as the previous solution, thanks to the optimization in the subproblems.

A deficiency of the algorithm is that the clustering solution highly depends on the initial cluster centers. Therefore, the algorithm can be restarted with different cluster center initializations to reduce the impact of random selection.

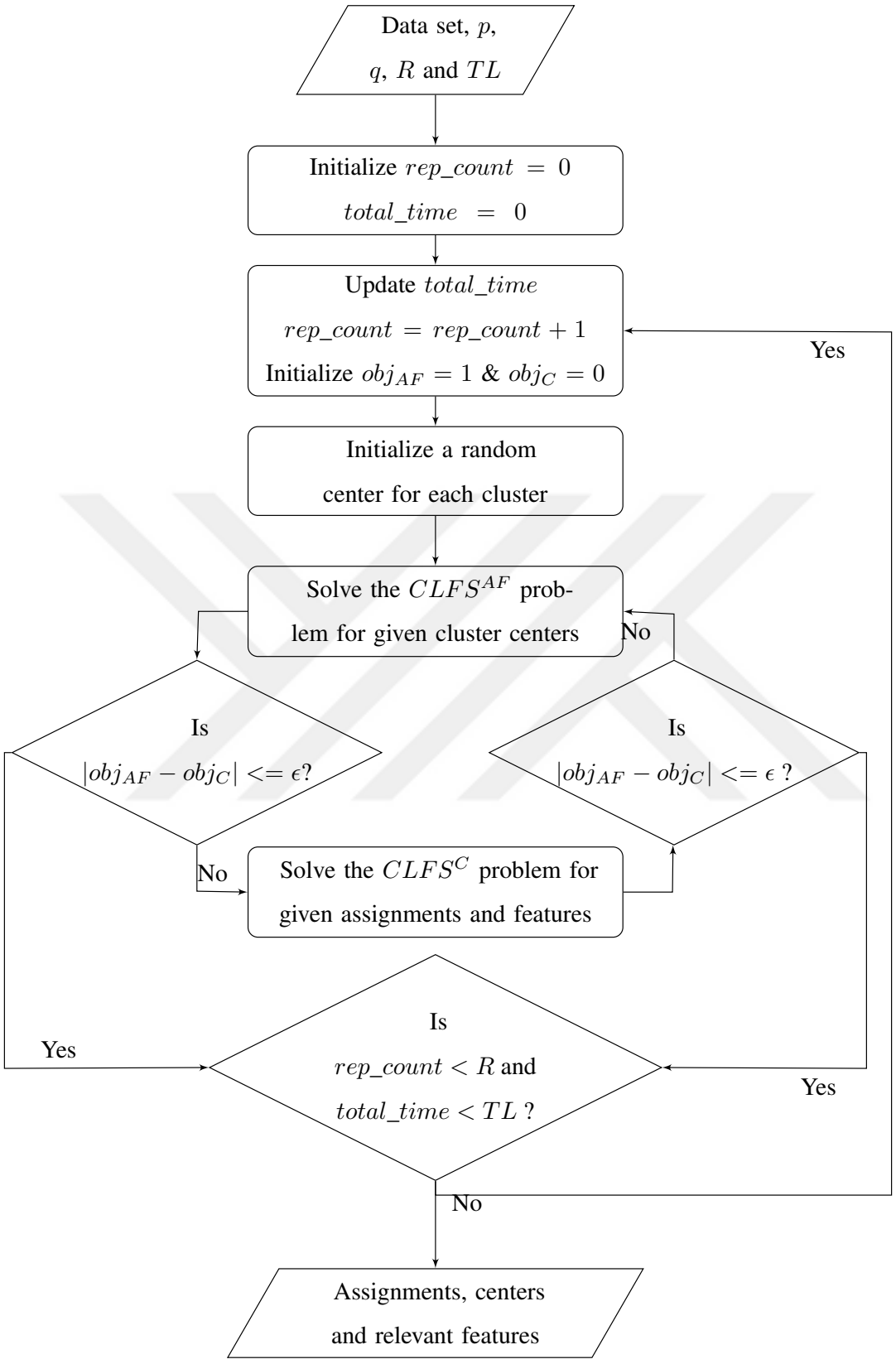


Figure 4.2: Flowchart of the Matheuristic Method for CLFS

Figure 4.2 provides a flowchart of the algorithm. A pseudocode of the algorithm is provided in Appendix A.

The algorithm gets the data set, the number of clusters, $p$, the number of features that can be selected for each cluster, $q$, the number of random initializations for cluster centers, $R$, and the time limit for the algorithm, $TL$, as the inputs. It gives the assignments, centers, and relevant features of clusters as the output.

$rep\_count$ is the count of random initialization, initialized to zero at the beginning of the algorithm. Random initializations continue while $rep\_count$ is less than $R$ and the total time of the algorithm is less than $TL$. For each initialization, $rep\_count$ is increased by one, and the total time is updated.

For each random initialization of cluster centers, an $obj_{AF} = 1$ and $obj_C = 0$ is also defined. $obj_{AF}$ represents the objective value of $CLFS^{AF}$ problem and $obj_C$ represents the objective value of $CLFS^C$ problem. Then, $CLFS^{AF}$ and $CLFS^C$ problems are solved iteratively until the difference between $obj_{AF}$ and $obj_C$ is less than a small threshold value, $\epsilon$.

We tested the performance of the matheuristic method on the generated data sets. The results of the computational study are discussed in Chapter 6. The matheuristic method takes much less computation time than $F1_{MISOCP}$ and $F2_{MISOCP}$ models. The results also show that it performs well in terms of accuracy. However, the computation time of the matheuristic method also accelerates with the growth of the data sets since the method's sub-problems are still solved using mathematical programming. Therefore, we developed a heuristic method to solve the CLFS problem on larger data sets with less computation time. The heuristic method for the CLFS problem is proposed in the following chapter.

# CHAPTER 5

# A HEURISTIC METHOD FOR THE CLFS PROBLEM

The computational performance of $F1_{MISOCP}$ and $F2_{MISOCP}$ models diminish with the growth of the data set. In Chapter 4, a matheuristic that combines mathematical programming with a heuristic algorithm is proposed to solve the CLFS problem. Although the solution time of the problem shows an improvement with the use of the matheuristic method, it is still difficult to solve the problem for larger data sets. This chapter proposes an iterative heuristic method to solve the CLFS problem on larger data sets with less computation time.

As discussed before, the CLFS problem has three groups of decisions. The first group is the assignment decisions of data points to clusters. The second group includes center selection decisions. The third group determines the set of relevant features for each cluster. The $F1_{MISOCP}$ and $F2_{MISOCP}$ models simultaneously give all three groups of decisions. However, giving all these decisions in a model results in many decision variables, and therefore, high computation times are encountered. We aim to decompose these three groups of decisions into subproblems such that each subproblem gives one of the groups of decisions. In this way, the number of decisions in each subproblem and, therefore, the solution time can be decreased.

The first subproblem is the Feature Selection Problem ($CLFS^F$). $CLFS^F$ aims to select the set of relevant features of each cluster when the data points and the center of clusters are given. The second subproblem is the Assignment Problem ($CLFS^A$). $CLFS^A$ aims to determine the assignments of data points to clusters when the centers and relevant features of clusters are given. The third subproblem is the Center Selection Problem ($CLFS^C$). $CLFS^C$ is to select the cluster centers among the points within the clusters when the assignment and relevant features are given.

It can be observed that if centers and features are given for a CLFS problem, the remaining problem, $CLFS^A$, is to assign each data point to the cluster with the closest center. Moreover, when the clusters and their relevant features are given, the remaining problem, $CLFS^C$ selects the data point closest to other data points within the cluster as the center for each cluster. Therefore, the exact solutions for these two subproblems can be found easily without solving a mathematical model. The only difficult problem is $CLFS^F$, selecting relevant features when the clusters and their centers are given.

In the rest of the chapter, Section 5.1 proposes the $CLFS^F$ problem. Section 5.2 proposes the $CLFS^A$ problem and Section 5.3 proposes the $CLFS^C$ problem. Lastly, Section 5.4 explains the iterative heuristic algorithm.

## 5.1 Feature Selection Problem ($CLFS^F$)

$CLFS^F$ aims to select the most relevant set of features for each cluster such that the center is closest to data points within the cluster where the assigned data points and centers are already known.

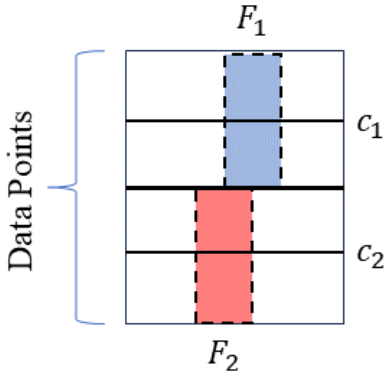


Figure 5.1: Feature Selection Problem ($CLFS^F$)

$CLFS^F$ is illustrated with a two-cluster setting in Figure 5.1. It can be seen that the partition of data points and the centers, $c_1$ and $c_2$, are already determined, and the

problem is selecting the relevant features for each cluster shown with dashed lines.

Since the clusters are already known, the problem can be formulated for each cluster separately. For each cluster $l$, $CLFS^F$ can be mathematically formulated as follows:

$$(F^F) \quad \min \quad \sum_{i \in A_l} \sqrt{\sum_{k=1}^{m} (v_{ik} - v_{C_l k})^2 \cdot z_k} \tag{5.1}$$

s.t.

$$\sum_{k=1}^{m} z_k = q \tag{5.2}$$

$$z_k \in \{0, 1\} \qquad \forall k \tag{5.3}$$

$A_l$ is the given set of data points assigned to cluster $l$. $C_l$ is the center of cluster $l$. $z_k$ is the binary decision variable that determines whether feature $k$ is relevant for the cluster or not. The objective function 5.1 is the sum of Euclidean distances between data points and the center over the selected features. Constraint 5.2 forces the selection of $q$ features for the cluster. Constraint 5.3 is the set constraint for the binary decision variable $z_k$.

$CLFS^F$ problem can be solved by converting the nonlinear binary programming model, $F^F$, to a MISOCP by taking a similar approach as in Property 1 and Property 2. However, in this heuristic method, we aim to solve subproblems without solving a mathematical programming model to reduce the solution time. Therefore, a heuristic algorithm is developed for the $CLFS^F$ problem. The next section proposes the heuristic method for the feature selection.

### 5.1.1 A Heuristic Method for $CLFS^F$

$CLFS^F$ problem assumes that the set of data points in a cluster and its center is given. The problem for a cluster is to select the subset of $q$ features that minimize the total distance of the center from the data points within the cluster. Each cluster's relevant set of features is selected separately, and a solution for the $CLFS^F$ is achieved.

The subset of features for each cluster that minimizes the total Euclidean distance can only be determined by a complete enumeration of all possible solutions. How-

ever, complete enumeration is costly for high-dimensional data sets for two reasons. First, $\binom{m}{q}^p$ possible solutions exist. Second, each possible solution requires $n - p$ distance calculations. Therefore, a heuristic algorithm, Algorithm 1, is proposed for the $CLFS^F$ problem.

Algorithm 1 selects $q$ features one by one based on their contributions to the total Euclidean distance. First, for a given cluster $l$, the total Squared Euclidean distance, $d_k^l$, is calculated for each feature $k$ as follows:

$$d_k^l = \sum_{i \in A_l} (v_{ik} - v_{C_l k})^2$$

$A_l$ is the set of data points in cluster $l$ and $C_l$ is the center of cluster $l$. Second, the features are sorted in ascending order of $d_k^l$. Assume that the indices of features are revised for each cluster so that $d_{1'}^l \leq d_{2'}^l \leq ... \leq d_{m'}^l$, i.e., feature $1'$ has the smallest $d_k^l$, feature $2'$ has the second smallest $d_k^l$ and so on.

Let $F_l$ be the set of features selected so far. The total Euclidean distance realized for cluster $l$ so far, $D_{F_l}^l$ is calculated as follows:

$$D_{F_l}^l = \sum_{i \in A_l} \sqrt{\sum_{k \in F_l} (v_{ik} - v_{C_l k})^2}$$

The algorithm starts by selecting feature $1'$ as the first feature, i.e., $F_l = \{1'\}$. It can be seen that feature $1'$ also minimizes the total Euclidean distance when there is one feature in $F_l$. If more than one feature to be selected, then $D_{F_l \cup \{2'\}}^l$ and $D_{F_l \cup \{3'\}}^l$ are compared. If $D_{F_l \cup \{2'\}}^l < D_{F_l \cup \{3'\}}^l$, update $F_l$ so that $F_l = \{1', 2'\}$. Otherwise, compare $D_{F_l \cup \{3'\}}^l$ and $D_{F_l \cup \{4'\}}^l$ . If $D_{F_l \cup \{3'\}}^l < D_{F_l \cup \{4'\}}^l$, update $F_l$ so that $F_l = \{1', 3'\}$. Otherwise, move on to compare $D_{F_l \cup \{4'\}}^l$ and $D_{F_l \cup \{5'\}}^l$ and so on. If no feature is selected until feature $m'$, then feature $m'$ is selected so that $F_l = \{1', m'\}$.

Only the remaining features are considered in each selection of a feature. For instance, if $F_l = \{1', 3'\}$, then first $D_{F_l \cup \{2'\}}^l$ and $D_{F_l \cup \{4'\}}^l$ will be compared in the next iteration. If $D_{F_l \cup \{2'\}}^l < D_{F_l \cup \{4'\}}^l$, then $F_l$ is updated so that $F_l = \{1', 3', 2'\}$. Else, the algorithm moves on to compare $D_{F_l \cup \{4'\}}^l$ and $D_{F_l \cup \{5'\}}^l$. The selection of features continues until $q$ features are selected for each cluster.

In the pseudocode of Algorithm 1, $A_i$ is the cluster index of each data point $i$. $obj_F$ represents the objective value of the $CLFS^F$ problem. The algorithm takes $A_i$ and

**Algorithm 1** A Heuristic Algorithm for $CLFS^F$

**Data:** $A_i, C_l$

**Result:** $F_l, obj_F$

Initialize $obj_F = 0$, $d_k^l = 0$ and $D_{F_l}^l = 0$.

**for** $l = 1, 2, ..., p$ **do**

  **for** $k = 1, 2, ..., m$ **do**

    **for** $i = 1, 2, ..., n$ **do**

      **if** $A_i == l$ **then**

        $d_k^l \leftarrow d_k^l + (v_{ik} - v_{C_l k})^2$

  Sort features in ascending order of $d_k^l$ →k'

  $F_l = \{1'\}$ and $U_l = U_l - \{1'\}$

  $D_{F_l}^l \leftarrow \sqrt{d_{1'}^l}$

  **if** $q > 1$ **then**

    **for** $r = 1, 2, ..., q-1$ **do**

      Initialize $c = 0$

      **while** $c \leq m - r - 2$ **do**

        $c \leftarrow c + 1$

        $f \leftarrow U_{l_c}$

        $s \leftarrow U_{l_{c+1}}$

        **for** $i = 1, 2, ..., n$ **do**

          **if** $A_i == l$ **then**

            $D_{F_l \cup \{f'\}}^l \leftarrow D_{F_l \cup \{f'\}}^l + \sqrt{\sum_{k \in F_l \cup \{f'\}} (v_{ik} - v_{C_l k})^2}$

            $D_{F_l \cup \{s'\}}^l \leftarrow D_{F_l \cup \{s'\}}^l + \sqrt{\sum_{k \in F_l \cup \{s'\}} (v_{ik} - v_{C_l k})^2}$

        **if** $D_{F_l \cup \{f'\}}^l < D_{F_l \cup \{s'\}}^l$ **then**

          $D_{F_l}^l \leftarrow D_{F_l \cup \{f'\}}^l$

          $F_l = F_l \cup \{f'\}$ and $U_l = U_l - \{f'\}$

          Break

        **else if** $c == m - r - 1$ **then**

          $D_{F_l}^l \leftarrow D_{F_l \cup \{s'\}}^l$

          $F_l = F_l \cup \{s'\}$ and $U_l = U_l - \{s'\}$

          Break

  $obj_F \leftarrow obj_F + D_{F_l}^l$

$C_l$ as input and gives $F_l$ and $obj_F$ as outputs. It starts with initializing $obj_F$, $d_k^l$ and $D_{F_l}^l$ to zero. For each cluster $l$, $d_k^l$ is calculated for each feature $k$, and features are sorted in ascending order of $d_k^l$. $k'$ represents the feature with $k^{th}$ smallest value of $d_k^l$.

First, feature $1'$ is selected and added to $F_l$. $U_l$ is the set of unselected features for cluster $l$. Selected features are discarded from $U_l$. If $q = 1$, then the total Euclidean distance for each cluster $l$ is the square root of $d_{1'}^l$. If $q > 1$, then $q - 1$ iterations of feature selection are performed for each cluster.

For each selection, an index counter $c$ is initialized to zero. $c$ is used to move among the features in the set $U_l$ and increased by one in each comparison. $f'$ represents the first feature to be compared in $U_l$ and $s'$ represents the second feature to be compared. First, $D_{F_l \cup \{f'\}}^l$ and $D_{F_l \cup \{s'\}}^l$ are calculated. If $D_{F_l \cup \{f'\}}^l < D_{F_l \cup \{s'\}}^l$, $f'$ is selected as the next feature. Otherwise, the next two features in $U_l$ are to be compared. If no more features are to be compared in $U_l$, then $s'$ is selected as the next feature. The selection continues and $D_{F_l}^l$ is updated until $q$ features are selected for a cluster. $obj_F$ is increased by each cluster's total distance, $D_{F_l}^l$.

The algorithm is further explained using a numerical example in the Subsection 5.1.1.1.

#### 5.1.1.1 A Numerical Example for Feature Selection Algorithm

Let's consider a cluster $l$ with three data points where data point 1 is the cluster center. In this example, the number of features, $m$, is four, and the number of features to be selected, $q$, is three. The values of each data point $i$ for each feature $k$, $v_{ik}$ are given in Table 5.1.

As the first step, $d_k^l$ for each feature is calculated, and $d_k^l$ are sorted from smallest to largest in Table 5.2. For instance, $d_1^l$ is $(9 - 6.6)^2 + (2.4 - 6.6)^2$ which is 23.4.

The first feature added to $F_l$ is feature 2 with the smallest $k'$, i.e., feature $1'$. $F_l$ and $U_l$ are updated such that $F_l = \{2\}$ and $U_l = \{1, 3, 4\}$. Then, the total Euclidean distances over a subset of features $\{2, 1\}$ and over a subset of features $\{2, 3\}$ are

Table 5.1: An Example for Feature Selection Algorithm

| **Data Points \Features** | **Feature 1** | **Feature 2** | **Feature 3** | **Feature 4** |
|---|---|---|---|---|
| **Data Point 1** | 6.6 | 5 | 4.5 | 4.5 |
| **Data Point 2** | 2.4 | 3 | 1.55 | 2 |
| **Data Point 3** | 9 | 8.5 | 8.5 | 9 |

Table 5.2: Total Squared Euclidean Distance and Rank For Each Feature

| | **Feature 1** | **Feature 2** | **Feature 3** | **Feature 4** |
|---|---|---|---|---|
| $d_k^l$ | 23.40 | 16.25 | 24.70 | 26.50 |
| $k'$ | $2'$ | $1'$ | $3'$ | $4'$ |

calculated for selecting the second feature. Table 5.3 shows that $D^l_{\{2,1\}}$ is greater than $D^l_{\{2,3\}}$. Therefore, the search continues by comparing the total Euclidean distances over $\{2,3\}$ and over $\{2,4\}$. This time, $D^l_{\{2,3\}}$ is lower than $D^l_{\{2,4\}}$. Therefore, feature 3 is added to $F_l$, i.e., $F_l = \{2,3\}$ and $U_l = \{1,4\}$.

Table 5.3: Total Euclidean Distances for Selecting the Second Feature

| | $D^l_{\{2,1\}}$ | $D^l_{\{2,3\}}$ |
|---|---|---|
| **Iteration 1** | 8.90 | 8.88 |
| | $D^l_{\{2,3\}}$ | $D^l_{\{2,4\}}$ |
| **Iteration 2** | 8.88 | 8.90 |

The total Euclidean distances over feature subsets $\{2,3,1\}$ and $\{2,3,4\}$ are compared for selecting the third feature in Table 5.4. $D^l_{\{2,3,1\}}$ is greater than $D^l_{\{2,3,4\}}$; however, there are no more features to be compared in $U_l$. Therefore, feature 4 is added to $F_l$ and removed from $U_l$, i.e., $F_l = \{2,3,4\}$ and $U_l = \{1\}$.

Table 5.4: Total Euclidean Distances for Selecting the Third Feature

| | $D^l_{\{2,3,1\}}$ | $D^l_{\{2,3,4\}}$ |
|---|---|---|
| **Iteration 1** | 11.34 | 11.32 |

Since three features have been selected so far, the algorithm terminates. The total Euclidean distance for the cluster is 8.88, and the selected subset of relevant features are $F_l = \{2, 3, 4\}$.

In this section, the $CLFS^F$ problem and a heuristic algorithm to solve the $CLFS^F$ problem have been proposed. The next section proposes the $CLFS^A$ problem.

## 5.2 Assignment Problem ($CLFS^A$)

$CLFS^A$ aims to assign each data point to the cluster where the center has the minimum Euclidean distance from the data point according to the relevant features of the cluster. $CLFS^A$ is illustrated with a two-cluster setting in Figure 5.2. It can be seen that the centers, $c_1$ and $c_2$, and the relevant sets of features of clusters, $F_1$ and $F_2$, are already determined, and the problem is determining the partition of data points to clusters shown with a dashed line.

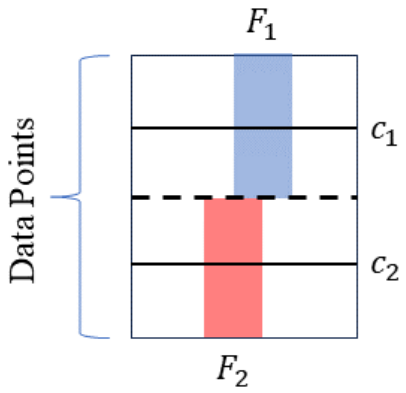


Figure 5.2: Assignment Problem ($CLFS^A$)

Since cluster centers and relevant features of clusters are given, the Euclidean distance of each data point from each cluster center is known in advance. Therefore, the assignment decision can be given for each data point separately. For each data point

$i$, $CLFS^A$ can be formulated as follows:

$$(F^A) \quad \min \quad \sum_{l=1}^{p} \sqrt{\sum_{k \in F_l} (v_{ik} - v_{C_l k})^2} \cdot x_l \tag{5.4}$$

s.t.

$$\sum_{l=1}^{p} x_l = 1 \tag{5.5}$$

$$x_l \geq 0 \qquad \forall l \tag{5.6}$$

$C_l$ is the center of cluster $l$. $F_l$ is the set of relevant features for cluster $l$. $x_l$ is the decision variable that takes the value of 1 if the data point $i$ is assigned to cluster $l$, 0 otherwise. The objective function 5.4 is the Euclidean distance between data point $i$ and the center of the cluster that it is assigned. Constraint 5.5 guarantees that the data point is assigned to exactly one cluster. Constraint 5.6 is the non-negativity constraint for the decision variable $x_l$.

$\sqrt{\sum_{k \in F_l} (v_{ik} - v_{C_l k})^2}$ is a parameter for each data point and each cluster center. The problem for each data point $i$ is determining the cluster center with minimum $\sqrt{\sum_{k \in F_l} (v_{ik} - v_{C_l k})^2}$. It is a trivial problem to solve, and the Algorithm 2 can be used to determine the closest cluster center.

**Algorithm 2** An Algorithm for $CLFS^A$

**Data:** $C_l, F_l$

**Result:** $A_i, obj_A$

Initialize $obj_A = 0$

**for** $i = 1, 2, ..., n$ **do**
    Initialize $min\_dist = \infty$
    **for** $l = 1, 2, ..., p$ **do**
        $d_{il} \leftarrow \sqrt{\sum_{k \in F_l} (v_{ik} - v_{C_l k})^2}$
        **if** $d_{il} < min\_dist$ **then**
            $min\_dist \leftarrow d_{il}$
            $A_i \leftarrow l$
        **end**
    **end**
    $obj_A \leftarrow obj_A + min\_dist$
**end**

$A_i$ represents the cluster index of data point $i$. $obj_A$ is the total Euclidean distance between data points and the center of their clusters. The algorithm starts with initializing $obj_A$ to 0 and then increasing $obj_A$ by the distance obtained from the assignment of each data point. For each data point $i$, a variable called $min\_dist$ is defined and initialized by infinity at the beginning. Then, the Euclidean distance between each data point $i$ and the center of each cluster $l$ is calculated over $F_l$. If the Euclidean distance between data point $i$ and $C_l$, $d_{il}$, is lower than $min\_dist$, then $min\_dist$ is updated as $d_{il}$, and the data point is assigned to cluster $l$ so $A_i$ takes the value of $l$.

The next section introduces the last subproblem for the heuristic method, $CLFS^C$.

### 5.3 Center Selection Problem ($CLFS^C$)

$CLFS^C$ aims to select the center of each cluster from the data points within the cluster when the clusters and relevant features of clusters are given. $CLFS^C$ is illustrated with a two-cluster setting in Figure 5.3. It can be seen that the partition of data points and the relevant sets of features of clusters, $F_1$ and $F_2$, are already determined, and the center for each cluster is shown with dashed lines within clusters.

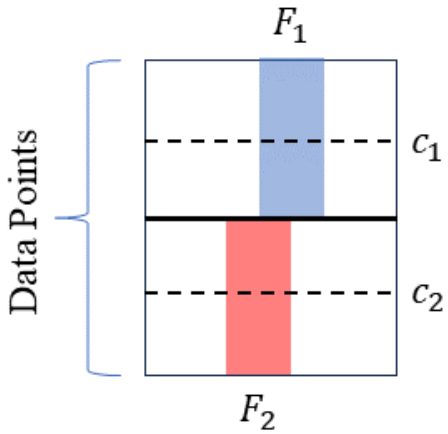


Figure 5.3: Center Selection Problem ($CLFS^C$)

Since the clusters are known in advance, the problem can be solved for each cluster

separately. For each cluster $l$, $CLFS^C$ can be formulated as follows:

$$(F^C) \quad \min \quad \sum_{i \in A_l} \sum_{\substack{j \in A_l \\ j \neq i}} \sqrt{\sum_{k \in F_l} (v_{ik} - v_{jk})^2} \cdot y_j \tag{5.7}$$

s.t.

$$\sum_{j \in A_l} y_j = 1 \tag{5.8}$$

$$y_j \in \{0, 1\} \qquad j \in A_l \tag{5.9}$$

$A_l$ is the given set of data points assigned to cluster $l$. $F_l$ is the given set of relevant features for cluster $l$. $y_j$ is the binary decision variable that takes the value of 1 if the data point $j$ is the center of cluster $l$, 0 otherwise. The objective function 5.7 is the total Euclidean distance between data points within the cluster and the cluster center. Constraint 5.8 states that one center must be selected for the cluster. Constraint 5.9 is the set constraint for the decision variables $y_j$.

$CLFS^C$ problem and $F^C$ are previously discussed for the matheuristic method in Section 4.2 also. In the matheuristic method, $F^C$ is solved by using an optimization solver. In the heuristic method, we aim to solve each subproblem without solving a mathematical programming model to reduce the solution time.

Since the partition of data points to clusters is already known, the $CLFS^C$ problem for each cluster is basically to select the center from the data points within the cluster. The data point with the minimum total Euclidean distance to other data points within the cluster should be selected as the cluster center. It is a trivial problem, and Algorithm 3 can be used to find the center with the minimum total distance to other data points for each cluster.

**Algorithm 3** An Algorithm for $CLFS^C$

**Data:** $A_i, F_l$

**Result:** $C_l, obj_C$

Initialize $obj_C = 0$

**for** $l = 1, 2, ..., p$ **do**
  Initialize $min\_dist = \infty$
  **for** $j = 1, 2, ..., n$ **do**
    **if** $A_j == l$ **then**
      Initialize $c\_dist = 0$
      **for** $i = 1, 2, ..., n$ **do**
        **if** $A_i == l$ **then**
          $c\_dist \leftarrow c\_dist + \sqrt{\sum_{k \in F_l} (v_{ik} - v_{jk})^2}$
        **end**
      **end**
      **if** $c\_dist < min\_dist$ **then**
        $min\_dist \leftarrow c\_dist$
        $C_l \leftarrow j$
      **end**
    **end**
  **end**
  $obj_C \leftarrow obj_C + min\_dist$
**end**

$A_i$ is the cluster index of data point $i$ while $F_l$ is the set of relevant features for cluster $l$. $C_l$ represents the center of cluster $l$. $obj_C$ is the total Euclidean distance between data points and the center of their clusters. The algorithm starts with initializing $obj_C$ to 0 and then increasing $obj_C$ by each selected center's total distance from the data points within its cluster. For each cluster, $l$, a variable called $min\_dist$ is defined and initialized by infinity at the beginning. Then, the total Euclidean distance of each data point $j$ from the other data points within the cluster, $c\_dist$, is calculated over $F_l$. If $c\_dist$ of a data point $j$ is lower than $min\_dist$ of its cluster $l$, then $min\_dist$ is updated as $c\_dist$ and $C_l$ takes the value of $j$.

Until now, the subproblems of the heuristic method have been discussed, and an al-

gorithm has been proposed for each problem. The next section explains how these subproblems are used in an algorithm to solve the CLFS problem.

## 5.4 An Iterative Heuristic Algorithm for CLFS

The heuristic method works based on the idea of solving $CLFS^A$, $CLFS^F$ and $CLFS^C$ iteratively to solve the CLFS problem. $CLFS^A$ and $CLFS^F$ problems form a submodule in the heuristic method. In this submodule, $CLFS^A$ and $CLFS^F$ are solved iteratively for the given centers to obtain a solution of assignments and relevant features. The output of one subproblem serves as the input to the other. Then, the $CLFS^C$ problem is solved to obtain cluster centers for the given assignments and features. The submodule and $CLFS^C$ are also solved iteratively, with the output of one serving as the input to the other to solve the CLFS problem.

Figure 5.4 demonstrates subproblems and their relationships in the heuristic method.

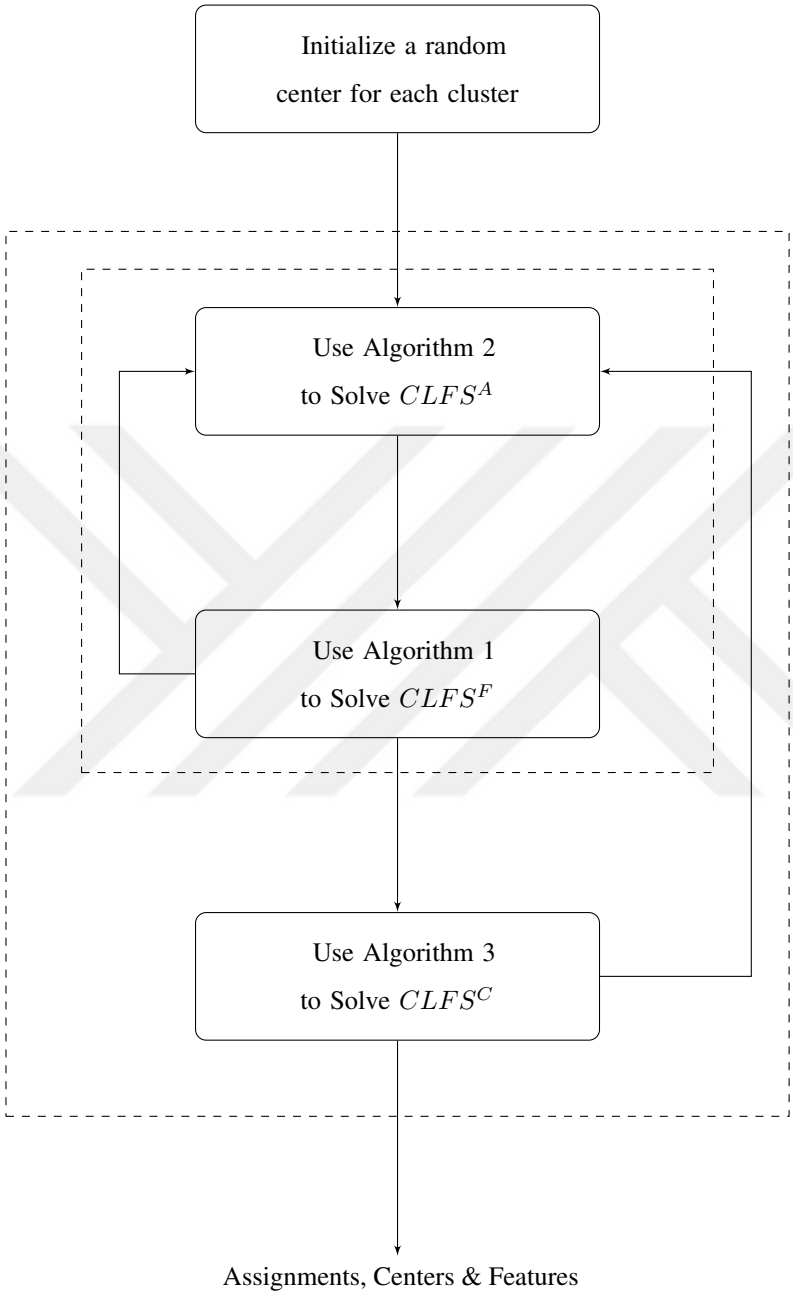


Figure 5.4: Representation of the Iterative Heuristic Method for CLFS

The method begins with the random initialization of cluster centers. The first set of iterations are between $CLFS^A$ and $CLFS^F$. These two subproblems are solved iteratively, with the output of one sub-problem serving as the input to the other. The algorithm solves the $CLFS^A$ problem using all $m$ features at the first iteration. The $CLFS^F$ problem is solved using initialized cluster centers and the obtained assignments. Then, $CLFS^A$ is solved with initialized cluster centers and obtained assignments.

The iterations between $CLFS^A$ and $CLFS^F$ continue until the objective values converge or the iteration limit is exceeded. Since the $CLFS^F$ problem is solved by using a heuristic algorithm, an optimal solution for the $CLFS^F$ problem, and therefore, the convergence of the objective values may not always be encountered. An iteration limit is used to cope with that.

After obtaining assignments and relevant features from $CLFS^A$ and $CLFS^F$ for the given centers, this time, $CLFS^C$ is solved with the given assignments and features to obtain new cluster centers. If the objective value obtained from $CLFS^C$ is not the same as that obtained from $CLFS^A$ and $CLFS^F$, then the method returns to solving $CLFS^A$ using all features for the given centers. These iterations continue until the objective values converge or the iteration limit is exceeded. Convergence is not always guaranteed, and an iteration limit is used to terminate the algorithm.

The clustering solution obtained as the result of the algorithm highly depends on the initial centers. The algorithm is repeated with many different random initializations of cluster centers to reduce the impact of the random selection. The Algorithm 4 defines the heuristic algorithm's pseudocode.

**Algorithm 4** A Heuristic Algorithm for CLFS

**Data:** Data set, $p$, $q$, $R$

**Result:** $A_i, C_l, F_l, obj_H$

Initialize $obj_H = \infty$ and $rep\_count = 0$

**while** $rep\_count < R$ **do**

- $rep\_count = rep\_count + 1$
- Initialize $cnt1 = 0$, $obj_{AF} = 1$ and $obj_C = 0$
- Initialize a random center for each cluster $\rightarrow$ $C_l$
- **while** $|obj_{AF} - obj_C| > \epsilon$ *and* $cnt1 < R$ **do**
  - $cnt1 = cnt1 + 1$
  - Initialize $cnt2 = 0$, $obj_F = 1$ and $obj_A = 0$
  - Initialize $F_l$ with all features $k = 1, 2, .., m$
  - **while** $|obj_F - obj_A| > \epsilon$ *and* $cnt2 < R$ **do**
    - Solve $CLFS^A$ with given $C_l$ and $F_l \rightarrow$ $A_i$, $obj_A$
    - $objAF \leftarrow objA$
    - $cnt2 = cnt2 + 1$
    - **if** $|obj_F - obj_A| > \epsilon$ *and* $cnt2 < R$ **then**
      - Solve $CLFS^F$ with given $C_l$ and $A_i \rightarrow$ $F_l$, $obj_F$
      - $objAF \leftarrow objF$
      - $cnt2 = cnt2 + 1$
    - **else**
      - Break
  - **if** $|obj_{AF} - obj_C| > \epsilon$ *and* $cnt1 < R$ **then**
    - Solve $CLFS^C$ with given $A_i$ and $F_l \rightarrow$ $C_l$, $obj_C$
    - $cnt1 = cnt1 + 1$
  - **else**
    - Break
- **if** $obj_{AF} <= obj_C$ & $obj_{AF} < obj_H$ **then**
  - $obj_H \leftarrow obj_{AF}$
- **else if** $obj_C < obj_{AF}$ & $obj_C < obj_H$ **then**
  - $obj_H \leftarrow obj_C$

The algorithm gets the data set, the number of clusters, $p$, the number of features to be selected for each cluster, $q$, and the iteration limit, $R$, as the inputs. It gives the

assignments, centers, and relevant features of clusters as the output with the objective function, $obj_H$. $A_i$ represents the cluster index of data point $i$; where $C_l$ is the center and $F_l$ is the set of relevant features for cluster $l$.

$obj_H$ is initialized to $\infty$, and the count of random initializations, $rep\_count$, is initialized to zero at the beginning of the algorithm. $rep\_count$ is increased by one for each random initialization, and the repetitions continue until $rep\_count$ gets the value of $R$. In each repetition, a random center for each cluster $l$ is initialized, and the objectives $obj_{AF}$ and $obj_C$ are initialized to different small values.

$cnt1$ and $cnt2$ are used to count the number of iterations. $cnt1$ is the count of iterations between $CLFS^A$ and $CLFS^C$. $cnt2$ is the count of iterations between $CLFS^A$ and $CLFS^F$. $cnt1$ is initialized to zero at each random initialization. $cnt2$ is initialized to zero for each set of iterations between $CLFS^A$ and $CLFS^F$.

First, $CLFS^A$ and $CLFS^F$ problems are solved iteratively to determine the $obj_{AF}$. $CLFS^A$ is initially solved by using all available features since $F_l$ is not determined yet. Then, $CLFS^F$ determines $F_l$ using the given $C_l$ and $A_i$. Next, $CLFS^A$ determines $A_i$ using the given $C_l$ and $F_l$. The iterations between $CLFS^F$ and $CLFS^A$ continue until $obj_F$ and $obj_A$ converge to the same value or the iteration limit is exceeded. As a result, $A_i$, $F_l$, and $obj_{AF}$ are obtained from these iterations.

Second, $CLFS^C$ is solved to determine $C_l$ using the given $A_i$ and $F_l$. If the difference between $obj_C$ and $obj_{AF}$ is smaller than a threshold value, $\epsilon$, then the algorithm returns to calculate $obj_{AF}$. These iterations continue until $obj_{AF}$ and $obj_C$ converge or the iteration limit is exceeded. After $R$ random initialization, $obj_H$ is determined as the minimum of the obtained objectives.

The performance of the heuristic method was experimented on using generated data sets. The next chapter provides the results of the computational experiments on the $F1_{MISOCP}$, $F2_{MISOCP}$, matheuristic and heuristic methods.

# CHAPTER 6

# COMPUTATIONAL RESULTS

In this chapter, we present the results of computational experiments for the methods discussed in Chapter 3, Chapter 4 and Chapter 5. Synthetic data sets were generated for the experiments. We provide information about the generated data sets, the details of experiments, and the computational results related to the proposed methods.

The remainder of the chapter is organized as follows. Section 6.1 describes the experimental design and data sets. Section 6.2 compares the computational results of $F1_{MISOCP}$ and $F2_{MISOCP}$ formulations. Section 6.3 presents the computational results for the matheuristic method and compares it with $F1_{MISOCP}$. Lastly, Section 6.4 evaluates the performance of the iterative heuristic method and compares it with the matheuristic method.

## 6.1 Generated Data Sets

In order to test the computational performance of the proposed solution approaches, we have created 420 random problem instances for the CLFS problem. We ensure that each problem instance is designed to have a certain number of clusters, $p$, for a certain number of data points, $n$. In the data set of each problem instance, each data point has a value for a particular number of features, $m$. For each cluster, $q$ features out of $m$ features define the cluster structure. The remaining $m - q$ features are irrelevant to the cluster structure for each cluster.

Table 6.1 shows the numbers of clusters, data points, features, and relevant features used in the data sets. The number of features, $m$, should be greater than the number

of relevant features, $q$. Therefore, 140 data sets are produced for each number of clusters, $p$, and 60 data sets for each number of data points, $n$.

Table 6.1: Parameters of Generated Data Sets

| | |
|---|---|
| **Number of clusters ($p$)** | $2, 3, 4$ |
| **Number of data points ($n$)** | $40, 50, 80, 100, 200, 500, 1000$ |
| **Number of features ($m$)** | $4, 5, 6, 8, 10, 12$ |
| **Number of relevant features ($q$)** | $2, 3, 4, 6$ |

In generated data sets, it was assumed that each cluster includes almost the same number of data points. For example, there are 20 data points in each cluster for a data set with 40 data points and 2 clusters, while 13, 13, and 14 data points exist in clusters for a data set with 3 clusters and 40 data points.

A data set contains $q$ relevant features and $m - q$ irrelevant features for each cluster. For each cluster, the columns of the data set corresponding to relevant features are randomly generated based on the multivariate normal distribution. On the other hand, the columns of the data set corresponding to irrelevant features are randomly generated based on uniform distribution. The choice of the distributions is based on the idea that the multivariate uniform distribution is a dense distribution suitable for representing a cluster structure. In contrast, the uniform distribution is scattered among data points and thus not ideal to represent a cluster of data points.

The means of multivariate normal distributions for relevant features differ for each cluster in a data set. Table 6.2 provides the parameters of multivariate normal distributions. $\mu$ defines the mean vector for a cluster while $\Sigma$ is the covariance matrix. To further explain the distributions, let's consider a data set with 2 clusters. Each column corresponding to relevant features for the first cluster has a mean value of 0 and a variance of 1. For the second cluster, the mean of each column corresponding to relevant features is 5, and the variance is 1.

The remaining $m - q$ columns corresponding to irrelevant features are randomly generated based on the U[0,5], U[0,10], and U[0,20] for each cluster.

The center of a cluster is determined as the data point with the smallest total distance

to other data points within the cluster.

For each generated data set, the total Euclidean distance between data points and their corresponding cluster center is calculated based on the relevant features of the clusters. We call the total distance according to the created cluster structure the "base objective" and denote it by $Z_B$. The base objective value is used to evaluate the objective values of the methods found for each generated instance.

Table 6.2: Parameters of Multivariate Distributions

| **Clusters Number** | **Distribution Parameters** |
|---|---|
| **Cluster 1** | $(\mu_1, \Sigma_1) = (\mathbf{0}_q, \mathbf{I}_q)$ |
| **Cluster 2** | $(\mu_2, \Sigma_2) = (\mathbf{5}_q, \mathbf{I}_q)$ |
| **Cluster 3** | $(\mu_3, \Sigma_3) = (-\mathbf{7}_q, \mathbf{I}_q)$ |
| **Cluster 4** | $(\mu_4, \Sigma_4) = (\mathbf{11}_q, \mathbf{I}_q)$ |

## 6.2 Computational Results of $F1_{MISOCP}$ and $F2_{MISOCP}$ Models

This section evaluates MISOCP models, $F1_{MISOCP}$ and $F2_{MISOCP}$, proposed in Chapter 3 using generated data sets and compares the results with each other. The MISOCP models were solved using CPLEX 22.1.0.0. The experiments were conducted on a 64-bit Windows 11 Enterprise PC with a 3.60 GHz six-core Intel Xeon E-2246G processor and 16 GB RAM. The results of the experiments were evaluated based on the following performance measures.

**MIP Relative Gap** ($\%Gap_R$): MIP Relative Gap is an optimality measure directly reported by CPLEX. It is the gap between the reported objective $Z$ and the best objective that can be obtained from the remaining nodes, $Z_R$. It is calculated as follows:

$$\%Gap_R = 100 \cdot (Z - Z_R)/Z_R \tag{6.1}$$

**Percent Gap from the Base Objective** ($\%Gap_B$): The objective value of the model is denoted by $Z$. The percent gap between base objective value, $Z_B$, and $Z$ is calculated

as follows:

$$\%Gap_B = 100 \cdot (Z - Z_B)/Z_B \tag{6.2}$$

$\%Gap_B$ may take negative values because there may be clusters of data points that result in a smaller total Euclidean distance than the created clusters. A negative gap means the proposed method found more compact clusters than the created ones for the data set.

**Computation Time**: Computation time is the solve time of the models in CPU seconds. TL indicates that the optimal solution cannot be found within the time limit specified as 1 hour (3600 seconds) in this study.

**Number of Instances the Optimal Solution Was Found** ($\#Opt$): The number of instances where the optimal solution was found. A solution is optimal if $\%Gap_R$ is zero.

**Number of Instances the Base Objective Was Found** ($\#Base$): The number of instances the base objective or a smaller objective value was found.

**Number of Instances the Time Limit Was Exceeded** ($\#TL$): The number of instances where the time limit was exceeded.

Table 6.3-6.5 provide the results of experiments conducted with $F1_{MISOCP}$ and $F2_{MISOCP}$ models on instances with 40 data points. The results of the experiments on the data sets with 50,80 and 100 data points are provided in Appendix B.1-B.9.

Each row of the tables corresponds to a problem instance. The first column shows the number of features, and the second column shows the number of relevant features in the data set. The remaining columns provide $\%Gap_R$, $\%Gap_B$, and computation time for $F1_{MISOCP}$ and $F2_{MISOCP}$ models for each problem instance. The last row of the tables shows the average values of each column. The average computation time is calculated for the data sets without exceeding the time limit. If the time limit is exceeded in all data sets with this setting, then the average computation time is left blank.

A zero value of $\%Gap_R$ shows that the obtained solution is optimal. Table 6.3 demonstrates that both $F1_{MISOCP}$ and $F2_{MISOCP}$ found the optimal solution within 1-hour

Table 6.3: Experimental Results for Data Sets with $n = 40$ and $p = 2$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 0.00 | -1.55 | 67.95 | 0.00 | -1.55 | 126.47 |
| 4 | 3 | 0.00 | 0.00 | 24.13 | 0.00 | 0.00 | 67.27 |
| 5 | 2 | 0.00 | 0.00 | 188.08 | 0.00 | 0.00 | 263.00 |
| 5 | 3 | 0.00 | 0.00 | 259.24 | 0.00 | 0.00 | 262.91 |
| 5 | 4 | 0.00 | 0.00 | 36.44 | 0.00 | 0.00 | 199.97 |
| 6 | 2 | 0.00 | -0.58 | 486.78 | 0.00 | -0.58 | 537.00 |
| 6 | 3 | 0.00 | 0.00 | 412.45 | 0.00 | 0.00 | 717.28 |
| 6 | 4 | 0.00 | 0.00 | 348.89 | 0.00 | 0.00 | 351.80 |
| 8 | 2 | 0.00 | 0.00 | 1407.19 | 0.00 | 0.00 | 3036.78 |
| 8 | 3 | 41.59 | 2.14 | TL | 0.00 | 0.00 | 2971.47 |
| 8 | 4 | 0.00 | 0.00 | 1734.16 | 0.00 | 0.00 | 2293.53 |
| 8 | 6 | 0.00 | -4.06 | 554.75 | 0.00 | -4.06 | 1557.22 |
| 10 | 2 | 91.62 | -1.05 | TL | 89.30 | -1.05 | TL |
| 10 | 3 | 93.10 | 8.98 | TL | 95.08 | 0.28 | TL |
| 10 | 4 | 91.37 | 2.06 | TL | 96.16 | 0.00 | TL |
| 10 | 6 | 85.96 | 1.82 | TL | 72.83 | -0.22 | TL |
| 12 | 2 | 97.96 | 4.38 | TL | 100.00 | 1.98 | TL |
| 12 | 3 | 97.11 | 0.17 | TL | 100.00 | -5.36 | TL |
| 12 | 4 | 99.45 | 5.14 | TL | 100.00 | 7.32 | TL |
| 12 | 6 | 99.11 | 2.37 | TL | 99.08 | 0.43 | TL |
| **Average** | | 39.86 | 0.99 | 501.82 | 37.62 | -0.14 | 1032.06 |

when $m < 10$. The only exception is that $F1_{MISOCP}$ could not find the optimal solution for the instance with $m = 8$, $q = 3$. For this instance, $F2_{MISOCP}$ found the optimal solution. $\%Gap_B$ may take negative values, such as in the problem instance with $m = 4$ and $q = 2$. This shows that the optimal objective value is less than the base objective value.

For the instances with $m = 10, 12$, the limit was exceeded by both $F1_{MISOCP}$ and $F2_{MISOCP}$. The optimal solution could not be found in these instances, but a feasible solution was obtained. $\%Gap_B$ shows that the obtained feasible solution gives a smaller objective value than the base objective in some instances, such as $m = 10$ and $q = 2$. $F1_{MISOCP}$ and $F2_{MISOCP}$ have a $\%Gap_B$ value of -1.05%, although the

optimal solution could not be found within the time limit.

For $n = 40$ and $p = 2$, $F1_{MISOCP}$ found the optimal solution in 11 instances and the base objective or a smaller value in 12 instances. $F1_{MISOCP}$ exceeded the limit in 9 instances. On the other hand, $F2_{MISOCP}$ found the optimal solution in 12 instances and found the base objective or a smaller value in 16 instances. It exceeded the time limit for 8 instances. When the computation times are compared, it can be seen that $F1_{MISOCP}$ has a smaller computation time on average. However, average values of $\%Gap_B$ and $\%Gap_R$ are smaller for $F2_{MISOCP}$.

Table 6.4: Experimental Results for Data Sets with $n = 40$ and $p = 3$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 99.30 | 5.78 | TL | 100.00 | 1.32 | TL |
| 4 | 3 | 0.00 | 0.00 | 1448.92 | 0.00 | 0.00 | 1487.58 |
| 5 | 2 | 100.00 | 0.56 | TL | 100.00 | -4.89 | TL |
| 5 | 3 | 96.88 | 0.00 | TL | 100.00 | 0.46 | TL |
| 5 | 4 | 0.00 | 0.00 | 3427.14 | 90.10 | 0.53 | TL |
| 6 | 2 | 100.00 | 6.16 | TL | 100.00 | -0.65 | TL |
| 6 | 3 | 100.00 | 7.58 | TL | 100.00 | 6.83 | TL |
| 6 | 4 | 97.85 | 6.20 | TL | 100.00 | 3.03 | TL |
| 8 | 2 | 100.00 | -3.78 | TL | 100.00 | 16.96 | TL |
| 8 | 3 | 100.00 | 19.56 | TL | 100.00 | 10.85 | TL |
| 8 | 4 | 100.00 | 9.14 | TL | 100.00 | 13.97 | TL |
| 8 | 6 | 95.76 | 2.30 | TL | 100.00 | 7.23 | TL |
| 10 | 2 | 100.00 | -7.65 | TL | 100.00 | -12.70 | TL |
| 10 | 3 | 100.00 | 16.93 | TL | 100.00 | 21.71 | TL |
| 10 | 4 | 100.00 | 26.18 | TL | 100.00 | 79.50 | TL |
| 10 | 6 | 100.00 | 13.99 | TL | 100.00 | 87.99 | TL |
| 12 | 2 | 100.00 | -9.60 | TL | 100.00 | -19.29 | TL |
| 12 | 3 | 100.00 | 22.57 | TL | 100.00 | 28.12 | TL |
| 12 | 4 | 100.00 | 22.81 | TL | 100.00 | 35.54 | TL |
| 12 | 6 | 100.00 | 12.07 | TL | 100.00 | 55.56 | TL |
| **Average** | | 89.49 | 7.54 | 2438.03 | 94.51 | 16.60 | 1487.58 |

Table 6.4 shows that $F1_{MISOCP}$ found the optimal solution for data sets with $m = 4$, $q = 3$ and $m = 5$, $q = 3$. $F2_{MISOCP}$ found the optimal solution for only the data

set with $m = 4$, $q = 3$. A feasible solution was found in the remaining data sets. In the data sets where the optimal solution could not be found, $F1_{MISOCP}$ found the base objective value in 5 instances, and $F2_{MISOCP}$ found the base objective value in 6 instances. This time, average values of $\%Gap_B$ and $\%Gap_R$ are smaller for $F1_{MISOCP}$.

Table 6.5: Experimental Results for Data Sets with $n = 40$ and $p = 4$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 5.13 | TL | 100.00 | 5.16 | TL |
| 4 | 3 | 96.72 | 4.95 | TL | 100.00 | 14.26 | TL |
| 5 | 2 | 100.00 | 15.83 | TL | 100.00 | 10.10 | TL |
| 5 | 3 | 100.00 | 4.82 | TL | 100.00 | 0.36 | TL |
| 5 | 4 | 95.92 | -1.33 | TL | 100.00 | 11.10 | TL |
| 6 | 2 | 100.00 | 19.19 | TL | 100.00 | 2.08 | TL |
| 6 | 3 | 100.00 | 30.80 | TL | 100.00 | 27.99 | TL |
| 6 | 4 | 100.00 | 7.26 | TL | 100.00 | 66.85 | TL |
| 8 | 2 | 100.00 | -13.26 | TL | 100.00 | -19.38 | TL |
| 8 | 3 | 100.00 | 17.25 | TL | 100.00 | 31.03 | TL |
| 8 | 4 | 100.00 | 22.86 | TL | 100.00 | 88.43 | TL |
| 8 | 6 | 100.00 | 7.45 | TL | 100.00 | 17.06 | TL |
| 10 | 2 | 100.00 | -19.01 | TL | 100.00 | -20.36 | TL |
| 10 | 3 | 100.00 | 33.08 | TL | 100.00 | 46.92 | TL |
| 10 | 4 | 100.00 | 22.68 | TL | 100.00 | 19.31 | TL |
| 10 | 6 | 100.00 | 13.72 | TL | 100.00 | 58.71 | TL |
| 12 | 2 | 100.00 | -43.05 | TL | 100.00 | -30.88 | TL |
| 12 | 3 | 100.00 | 25.92 | TL | 100.00 | 34.56 | TL |
| 12 | 4 | 100.00 | 48.04 | TL | 100.00 | 58.37 | TL |
| 12 | 6 | 100.00 | 80.43 | TL | 100.00 | 75.14 | TL |
| **Average** | | 99.63 | 14.14 | - | 100.00 | 24.84 | - |

Table 6.5 shows that none of $F1_{MISOCP}$ and $F2_{MISOCP}$ can find the optimal solution within the time limit. However, the base objective value was found in 4 data sets by $F1_{MISOCP}$ and in 3 data sets by $F2_{MISOCP}$. $\%Gap_R$ is 100% for $F2_{MISOCP}$ in all data sets and it is 100% for $F1_{MISOCP}$ except the data sets with $m = 4$, $q = 3$ and $m = 5$, $q = 4$. Also, $F1_{MISOCP}$ has a smaller average $\%Gap_B$.

Table 6.3-6.5 also shows that the computation time increases with the number of features, $m$. On the other hand, there is no positive relationship between the number of selected features, $q$, and the computation time for the same $m$.

Table 6.6 summarizes the experimental results according to the number of data points and clusters. In this table, each row corresponds to the results of 20 data sets with $n$ data points and $p$ clusters. The remaining columns provide the $\#Opt$, $\#Base$ and $\#TL$ values for both $F1_{MISOCP}$ and $F2_{MISOCP}$.

Table 6.6: Experimental Results by Number of Data Points and Number of Clusters

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **n** | **p** | $\#Opt$ | $\#Base$ | $\#TL$ | $\#Opt$ | $\#Base$ | $\#TL$ |
| 40 | 2 | 11 | 12 | 9 | 12 | 16 | 8 |
| 40 | 3 | 2 | 6 | 18 | 1 | 6 | 19 |
| 40 | 4 | 0 | 4 | 20 | 0 | 3 | 20 |
| 50 | 2 | 9 | 10 | 11 | 8 | 16 | 12 |
| 50 | 3 | 0 | 0 | 20 | 0 | 5 | 20 |
| 50 | 4 | 0 | 2 | 20 | 0 | 2 | 20 |
| 80 | 2 | 3 | 4 | 17 | 0 | 3 | 20 |
| 80 | 3 | 0 | 2 | 20 | 0 | 0 | 20 |
| 80 | 4 | 0 | 1 | 20 | 0 | 0 | 20 |
| 100 | 2 | 1 | 3 | 19 | 0 | 1 | 20 |
| 100 | 3 | 0 | 0 | 20 | 0 | 0 | 20 |
| 100 | 4 | 0 | 1 | 20 | 0 | 0 | 20 |

Table 6.6 shows a relationship between $\#Opt$ and $\#TL$ such that the summation of $\#Opt$ and $\#TL$ gives the number of instances, 20, for each row. The reason is that if the time limit was exceeded and if a feasible solution is obtained, the feasible solution is not optimal for this data set. $\#Opt$ is always smaller than or equal to $\#Base$ because the feasible solution may result in an objective value smaller than or equal to the base objective value, even though it is not optimal.

Table 6.6 demonstrates that $\#Opt$ and $\#Base$ decrease with the increase in $n$ and $p$ for both $F1_{MISOCP}$ and $F2_{MISOCP}$. On the other hand, $\#TL$ increases if $n$ or $p$ increases. Therefore, it can be said that the number of clusters and data points contribute to the computation time for both $F1_{MISOCP}$ and $F2_{MISOCP}$ besides the

number of features.

When $F1_{MISOCP}$ and $F2_{MISOCP}$ are compared according to Table 6.6, it can be seen that the $\#Opt$ is generally higher and $\#TL$ is lower for $F1_{MISOCP}$. The only exception is the instances with $n = 40$ and $p = 2$. Therefore, it can be said that $F1_{MISOCP}$ is superior to $F2_{MISOCP}$ in terms of computational time for especially larger data sets. None of the models dominates the other with respect to $\#Base$.

To conclude, both $F1_{MISOCP}$ and $F2_{MISOCP}$ can find the optimal solution within the time limit for data sets with lower $n$, $p$, and $m$ values. However, computation time increases as $n$, $p$, and $m$ increase, and the optimal solution cannot be found within the time limit for larger data sets.

The next section presents the results of computational experiments for the matheuristic method proposed in Chapter 4.

## 6.3 Computational Results of the Matheuristic Method

This section evaluates the matheuristic method proposed in Chapter 4 using generated data sets and compares the results with $F1_{MISOCP}$. The algorithm was coded in Python, and the mathematical models within the algorithm were solved using CPLEX 22.1.0.0.

The algorithm was repeated with 100 initializations of cluster centers for each data set to cope with the effect of a random start. The results of the experiments for each data set with 100 initializations were evaluated based on the following performance measures.

**Best Solution Gap ($\%Gap_{Best}$)**: The objective value of the best solution out of 100 initializations is denoted $Z^{Best}$. If an optimal solution is reported, $\%Gap_{Best}$ is the percent gap between $Z^{Best}$ and $Z^*$ where $Z^*$ is the optimal objective value, and it is calculated as follows:

$$\%Gap_{Best} = 100 \cdot (Z^{Best} - Z^*)/Z^* \tag{6.3}$$

If the optimal solution is not known, $\%Gap_{Best}$ is the percent gap between $Z^{Best}$ and

$Z_B$, the base objective value, and it is calculated as follows:

$$\%Gap_{Best} = 100 \cdot (Z^{Best} - Z_B)/Z_B \tag{6.4}$$

The instances where the optimal solution is known are indicated with "*".

**Number of Best Solutions ($\#Best$)**: The number of initializations where the best solution was found out of 100 initializations.

**Worst Solution Gap ($\%Gap_{Worst}$)**: The percent gap between the objective value of the worst solution found, $Z^{Worst}$, and $Z_B$ is calculated as follows:

$$\%Gap_{Worst} = 100 \cdot (Z^{Worst} - Z_M)/Z_B \tag{6.5}$$

It should be noted that the worst solution may be the same as the best solution if all initializations yield the same objective value.

**Number of Worst Solutions ($\#Worst$)**: The number of initializations where the worst solution was found out of 100.

**Average Gap ($\%Gap_{Avg}$)**: The objective value of the solution in each initialization $t$ is denoted as $Z_t$, and the gap between $Z_t$ and $Z_B$ is calculated for each initialization $t$ according to the following equation:

$$\%Gap_t = 100 \cdot (Z_t - Z_B)/Z_B \tag{6.6}$$

The average gap between the solutions and the base objective in a data set is as follows:

$$\%Gap_{Avg} = \sum_{t=1}^{100} \%Gap_t \tag{6.7}$$

**Number of Base Solutions ($\#IBase$)**: The number of initializations where the base or a smaller objective value was found out of 100.

**Total Computation Time (Total Time)**: It is the total computation time (in seconds) for 100 initializations. The computation time of mathematical models within the algorithm was restricted to 1-hour. The total computation time for an instance is restricted to 24 hours. If the 24 hours is exceeded, the initializations obtained until the time limit were considered in evaluations.

Table 6.7-6.9 presents the results of experiments on the matheuristic method and its comparison with $F1_{MISOCP}$ for data sets consisting of 40 data points. The results of the experiments on the data sets with 50, 80 and 100 data points are provided in Appendix C.1-C.7.

Each row of the tables corresponds to a problem instance. The first column shows the number of features, and the second column shows the number of relevant features in the data set. The next two columns show $\%Gap_B$ and computation time for $F1_{MISOCP}$. The remaining columns provide the values of performance measures for the matheuristic method. The last row of the tables shows the average values of each column. The average computation time is calculated for the data sets without exceeding the time limit.

Table 6.7: Experimental Results for Data Sets with $n = 40$ and $p = 2$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -1.55* | 67.95 | 0.00* | 3.70 | 50.43 | 45 | 2 | 90 | 315.69 |
| 4 | 3 | 0.00* | 24.13 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 579.64 |
| 5 | 2 | 0.00* | 188.08 | 0.00* | 9.49 | 110.08 | 86 | 1 | 86 | 382.57 |
| 5 | 3 | 0.00* | 259.24 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 547.09 |
| 5 | 4 | 0.00* | 36.44 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 606.41 |
| 6 | 2 | -0.58* | 486.78 | 0.00* | 4.78 | 86.16 | 56 | 1 | 89 | 438.96 |
| 6 | 3 | 0.00* | 412.45 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 428.64 |
| 6 | 4 | 0.00* | 348.89 | 0.00* | 3.25 | 153.18 | 81 | 1 | 81 | 924.77 |
| 8 | 2 | 0.00* | 1407.19 | 0.00* | 4.77 | 71.31 | 75 | 1 | 75 | 512.95 |
| 8 | 3 | 2.14 | TL | 0.00* | 1.34 | 8.74 | 78 | 5 | 78 | 548.02 |
| 8 | 4 | 0.00* | 1734.16 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 535.54 |
| 8 | 6 | -4.06* | 554.75 | 0.00* | -3.14 | -2.33 | 47 | 53 | 100 | 1487.10 |
| 10 | 2 | -1.05 | TL | -1.05 | 0.86 | 50.13 | 65 | 1 | 65 | 759.45 |
| 10 | 3 | 8.98 | TL | 0.00 | 10.06 | 69.10 | 51 | 1 | 51 | 790.35 |
| 10 | 4 | 2.06 | TL | 0.00 | 2.12 | 93.56 | 97 | 1 | 97 | 794.76 |
| 10 | 6 | 1.82 | TL | -0.22 | -0.21 | -0.19 | 47 | 53 | 100 | 1757.12 |
| 12 | 2 | 4.38 | TL | 0.00 | 6.68 | 82.80 | 67 | 1 | 67 | 777.52 |
| 12 | 3 | 0.17 | TL | -6.49 | -0.99 | 50.91 | 61 | 1 | 64 | 933.20 |
| 12 | 4 | 5.14 | TL | 0.00 | 8.71 | 77.52 | 30 | 1 | 30 | 1430.14 |
| 12 | 6 | 2.37 | TL | 0.00 | 4.13 | 66.14 | 37 | 1 | 37 | 1114.35 |
| **Average** | | 0.99 | 501.82 | -0.39 | 2.78 | 48.38 | 73.80 | 31.20 | 80.50 | 783.21 |

As discussed, $\%Gap_{Best}$ shows the percent gap of the best solution from the optimal objective value if the optimal solution is known. Otherwise, it shows the percent gap of the best solution from the base objective value. The instances where the optimal solution is known are indicated with "*". Therefore, $\%Gap_{Best}$ shows the percent gap

of the best solution from the optimal objective value for these instances.

Table 6.7 shows that the optimal solution is known for the data sets with $m < 10$. It can be seen that $\%Gap_{Best}$ is zero for all the data sets where the optimal solution is known. In the data set with $m = 8$ and $q = 3$, the optimal solution was reported by $F2_{MISOCP}$ while $F1_{MISOCP}$ could not find the optimal solution within the time limit. It can be seen that the matheuristic reports the optimal solution for this data set also. For the data sets where the optimal solution is unknown, the best solution found is always smaller than or equal to the base objective according to $\%Gap_{Best}$. Also, it is observed that the matheuristic reports a solution at least as good as the solution of $F1_{MISOCP}$.

$\%Gap_{Avg}$ represents the percentage gap of an average solution from the base objective value while $\%Gap_{Worst}$ shows the percent gap of worst-case solution. In some instances, all solutions obtained are the same, and the worst solution found is even optimal. For example, in the data set with $m = 4$ and $q = 3$, all solutions obtained are the optimal solutions, i.e., $\%Gap_{Best}$, $\%Gap_{Avg}$ and $\%Gap_{Worst}$ are zero. In some instances, the average and worst solutions yield the base objective value, although they are not optimal. To illustrate, in the data set with $m = 8$ and $q = 6$, $\%Gap_{Avg}$ is -3.14% and $\%Gap_{Worst}$ is -2.33% while the gap between the optimal solution and the base objective is -4.06%.

$\#Best$ demonstrates the frequency of finding the best solution through the initializations. Table 6.7 shows that $\#Best$ is over 50 in most data sets, and the average $\#Best$ is 73.80. $\#IBase$ shows the frequency of finding the base objective value. The average value of $\#IBase$ is 80.50, which shows that the proposed method frequently finds the hidden cluster structure or a more compact one. Although many initializations are performed to cope with the random start, the high values of $\#IBase$ show that the base objective value can also be found with lower numbers of initializations in most data sets. $\#Worst$ demonstrates the frequency of encountering the worst-case scenario through the initializations. Although $\%Gap_{Worst}$ is high for some instances, it can be seen that $\#Worst$ is quite low for these instances.

Table 6.7 demonstrates that the total computation time of the matheuristic method is always less than 1 hour. Although the computation time of $F1_{MISOCP}$ is smaller in

the data sets with less number of features, the matheuristic method can find a solution in data sets where $F1_{MISOCP}$ exceeded the time limit.

Table 6.8: Experimental Results for Data Sets with $n = 40$ and $p = 3$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 5.78 | TL | 0.00 | 12.58 | 68.73 | 44 | 2 | 44 | 433.32 |
| 4 | 3 | 0.00* | 1448.92 | 0.00* | 39.74 | 159.22 | 73 | 1 | 73 | 694.90 |
| 5 | 2 | 0.56 | TL | -5.63 | 6.97 | 83.69 | 42 | 1 | 74 | 576.41 |
| 5 | 3 | 0.00 | TL | 0.00 | 22.18 | 199.09 | 24 | 1 | 24 | 517.83 |
| 5 | 4 | 0.00* | 3427.14 | 0.00* | 40.35 | 148.73 | 68 | 1 | 68 | 1460.23 |
| 6 | 2 | 6.16 | TL | -4.45 | 3.60 | 48.34 | 39 | 1 | 47 | 789.57 |
| 6 | 3 | 7.58 | TL | -5.00 | 6.16 | 102.91 | 20 | 2 | 88 | 672.71 |
| 6 | 4 | 6.20 | TL | 0.00 | 46.74 | 203.43 | 66 | 1 | 66 | 590.45 |
| 8 | 2 | -3.78 | TL | -8.31 | -2.70 | 27.03 | 23 | 1 | 79 | 902.08 |
| 8 | 3 | 19.56 | TL | 0.00 | 12.64 | 56.26 | 53 | 2 | 53 | 1023.57 |
| 8 | 4 | 9.14 | TL | -0.71 | 20.48 | 139.73 | 33 | 1 | 33 | 1298.67 |
| 8 | 6 | 2.30 | TL | 0.00 | 51.31 | 226.98 | 50 | 1 | 50 | 1914.95 |
| 10 | 2 | -7.65 | TL | -24.35 | -18.19 | -3.64 | 15 | 1 | 100 | 1883.48 |
| 10 | 3 | 16.93 | TL | -0.87 | 7.87 | 44.43 | 34 | 1 | 47 | 4411.87 |
| 10 | 4 | 26.18 | TL | 0.00 | 13.39 | 71.27 | 64 | 1 | 64 | 3590.31 |
| 10 | 6 | 13.99 | TL | -3.11 | 20.96 | 109.74 | 10 | 1 | 71 | 9560.39 |
| 12 | 2 | -9.60 | TL | -20.87 | -15.74 | 4.59 | 6 | 1 | 99 | 3837.13 |
| 12 | 3 | 22.57 | TL | -2.79 | 4.88 | 51.83 | 20 | 1 | 44 | 23476.87 |
| 12 | 4 | 22.81 | TL | -2.13 | 10.73 | 73.97 | 5 | 1 | 19 | 12383.55 |
| 12 | 6 | 12.07 | TL | -4.24 | 20.93 | 133.15 | 13 | 1 | 51 | 67886.85 |
| **Average** | | 7.54 | 2438.03 | -4.12 | 15.24 | 97.47 | 35.10 | 1.15 | 59.70 | 6895.26 |

Table 6.8 shows that the optimal solution is known only for the data sets with $m = 4$, $q = 3$ and $m = 5$, $q = 4$. In these instances, the matheuristic method found the optimal solution, i.e., $\%Gap_{Best} = 0$. The percent gap of the best solution from the base objective value is reported as less than or equal to zero for all the data sets where the optimal solution is unknown. Also, it can be observed that the best solution is always better than or equal to the solution obtained by $F1_{MISOCP}$. It can be seen that the $\%Gap_{Avg}$ and $\%Gap_{Worst}$ are less than zero for some instances, such as $m = 10$ and $q = 2$. Although $\%Gap_{Worst}$ is a high average value of 97.47, $\#Worst$ takes the value of one or two in all instances.

The total computation time is less than 1 hour for the data sets with $m < 10$. On the other hand, the total time increases to more than 1 hour with $m = 10$ and exceeds 18 hours with $m = 12$.

Table 6.9: Experimental Results for Data Sets with $n = 40$ and $p = 4$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 5.13 | TL | 0.00 | 24.82 | 111.03 | 21 | 1 | 21 | 559.83 |
| 4 | 3 | 4.95 | TL | -2.38 | 34.39 | 108.81 | 14 | 1 | 37 | 488.19 |
| 5 | 2 | 15.83 | TL | -2.62 | 11.11 | 73.62 | 24 | 1 | 31 | 759.19 |
| 5 | 3 | 4.82 | TL | -6.17 | 17.49 | 98.55 | 21 | 1 | 69 | 738.40 |
| 5 | 4 | -1.33 | TL | -2.40 | 44.80 | 172.37 | 25 | 1 | 36 | 859.22 |
| 6 | 2 | 19.19 | TL | 0.00 | 17.78 | 59.13 | 27 | 1 | 27 | 1375.69 |
| 6 | 3 | 30.80 | TL | 0.00 | 8.01 | 89.83 | 52 | 1 | 52 | 1282.90 |
| 6 | 4 | 7.26 | TL | 0.00 | 22.75 | 121.94 | 78 | 1 | 78 | 1025.26 |
| 8 | 2 | -13.26 | TL | -27.70 | -22.50 | 1.42 | 6 | 1 | 99 | 3969.04 |
| 8 | 3 | 17.25 | TL | -7.91 | 9.24 | 55.31 | 16 | 1 | 46 | 7949.39 |
| 8 | 4 | 22.86 | TL | -1.77 | 27.97 | 111.28 | 60 | 1 | 60 | 3618.11 |
| 8 | 6 | 7.45 | TL | -0.58 | 57.84 | 211.61 | 34 | 1 | 34 | 2273.50 |
| 10 | 2 | -19.01 | TL | -28.12 | -24.39 | -6.21 | 8 | 1 | 100 | 55843.61 |
| 10 | 3 | 33.08 | TL | -3.6 | 7.22 | 29.03 | 10 | 5 | 40 | TL |
| 10 | 4 | 22.68 | TL | -13.03 | 9.07 | 47.18 | 8.7 | 4.35 | 56.52 | TL |
| 10 | 6 | 13.72 | TL | 0.00 | 48.56 | 120.1 | 28 | 4 | 28 | TL |
| 12 | 2 | -43.05 | TL | -46.24 | -41.26 | -34.11 | 9.38 | 3.13 | 100 | TL |
| 12 | 3 | 25.92 | TL | -6.79 | 8.36 | 26.52 | 16.67 | 16.67 | 50 | TL |
| 12 | 4 | 48.04 | TL | 4.98 | 17.81 | 34.05 | 20 | 20 | 0 | TL |
| 12 | 6 | 80.43 | TL | -1.57 | 12.39 | 57.39 | 16.67 | 8.33 | 25 | TL |
| **Average** | | 14.14 | - | -7.3 | 14.57 | 74.44 | 24.77 | 3.72 | 49.48 | 8716.26 |

Table 6.9 demonstrates that the optimal solution is unknown for any data set; therefore, $\%Gap_{Best}$ shows the gap from the base objective value. It can be seen that the best solution finds the base objective value in almost all data sets. The only exception is the data set with $m = 12$ and $q = 4$. For this data set, the time limit (24 hours) is exceeded, and the initializations until the time limit is evaluated. A positive value of $\%Gap_{Best}$ shows that the base objective could not be found within 24 hours.

The total computation time is almost less than 1 hour for the data sets with $m < 10$, except for the instance with $m = 8$ and $q = 3$. For the data sets with $m = 10, 12$, the time limit is exceeded. For these data sets, the initializations until the time limit are evaluated, and the $\#Best$, $\#Worst$, and $\#IBase$ are scaled to 100 initializations for comparison with other data sets. The average total computation time is calculated for the data sets where the time limit was not exceeded.

Table 6.10 summarizes the experimental results according to the number of data

points and the number of clusters. In this table, each row corresponds to the results of data sets with $n$ data points and $p$ clusters. The average values of $\%Gap_{Best}$, $\%Gap_{Avg}$, $\%Gap_{Worst}$, $\#Best$, $\#Worst$, $\#IBase$, and the average total time are presented.

Table 6.10: Experimental Results by Number of Data Points and Number of Clusters

| n | p | Average $\%Gap_{Best}$ | Average $\%Gap_{Avg}$ | Average $\%Gap_{Worst}$ | Average $\#Best$ | Average $\#Worst$ | Average $\#IBase$ | Average Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 40 | 2 | -0.39 | 2.78 | 48.38 | 73.80 | 31.20 | 80.50 | 783.21 |
| 40 | 3 | -4.12 | 15.24 | 97.47 | 35.10 | 1.15 | 59.70 | 6895.26 |
| 40 | 4 | -7.30 | 14.60 | 74.44 | 24.65 | 3.74 | 49.58 | 8716.26 |
| 50 | 2 | -0.09 | 3.43 | 80.01 | 76.50 | 12.45 | 78.65 | 1319.56 |
| 50 | 3 | -1.97 | 16.41 | 87.40 | 36.60 | 1.05 | 55.60 | 8074.00 |
| 50 | 4 | -5.04 | 15.95 | 79.15 | 27.04 | 5.12 | 47.98 | 8793.26 |
| 80 | 2 | -0.66 | 2.13 | 43.24 | 78.60 | 20.00 | 78.60 | 3466.65 |
| 80 | 3 | -3.22 | 13.98 | 83.38 | 40.92 | 2.39 | 56.52 | 10794.51 |
| 80 | 4 | -2.35 | 19.33 | 82.80 | 26.40 | 5.33 | 37.26 | 9117.75 |
| 100 | 2 | -0.25 | 2.85 | 69.28 | 81.91 | 13.86 | 81.91 | 8707.67 |
| 100 | 3 | -1.52 | 11.93 | 81.31 | 42.29 | 1.41 | 56.36 | 12382.35 |
| 100 | 4 | 1.53 | 19.70 | 81.29 | 42.21 | 11.33 | 49.48 | 20874.11 |

Table 6.10 shows that average $\%Gap_{Best}$ is less than zero for all data sets except data sets with $n = 100$ and $p = 4$. $\%Gap_{Avg}$ is almost 16% on average, and average $\#IBase$ is generally over 50% for even 100 data points. Therefore, it can be said that the matheuristic method performs well in finding the hidden clustering structure. Also, the average value of $\#Best$ is higher than $\#Worst$ in each row. This shows that the matheuristic is more likely to converge the best solution rather than the worst solution through initializations. Table 6.10 also shows that the average computation time of the matheuristic method increases with the number of data points and the number of clusters.

To conclude, the matheuristic method can find the base objective value in larger data sets where $F1_{MISOCP}$ cannot find an optimal solution within the time limit. However, the computation time of the matheuristic accelerates also as the data sets grow. This is because the method's subproblems are still solved using mathematical programming. The next section presents the computational results regarding the heuristic method proposed in Chapter 5 and its comparison with the matheuristic approach.

## 6.4 Computational Results of the Iterative Heuristic Method

This section evaluates the computational results of the heuristic method proposed in Chapter 5 on generated data sets. The performance of the heuristic method is compared with the matheuristic method proposed in Chapter 4.

The algorithm was compiled in Python. As in the matheuristic method, the heuristic algorithm is repeated with 100 initializations for each data set. The results of the experiments for each data set were evaluated based on the performance measures proposed in Section 6.3. Also, the following performance measures are used to compare the matheuristic and the heuristic methods:

**Percentage of Initializations that the Matheuristic Found Better ($\%Mth$)**: The percentage of random initializations where the matheuristic method found a smaller objective value than the heuristic method.

**Percentage of Initializations that the Heuristic Found Better ($\%Hrs$)**: The percentage of random initializations where the heuristic method found a smaller objective value than the matheuristic method.

**Percentage of Initializations that the Matheuristic and Heuristic Found the Same Solution ($\%Eql$)**: The percentage of random initializations where the matheuristic and heuristic methods found the same objective value.

Table 6.11-6.13 presents the results of experiments for the data sets with 40 data points. The results of the experiments on the data set with 50, 80, 100, 200, 500, and 1000 data points are provided in Appendix D.1-D.18. The tables for the data sets with $n \leq 100$ data points also include the results of the matheuristic for comparison.

Each table shows the results for the data sets with a certain number of data points, $n$, and a certain number of clusters, $p$, where each row in tables corresponds to a problem instance. The first two columns show the number of features and the number of relevant features for the problem instance. For the data sets with 100 or fewer data points, the performance measures of the matheuristic are presented first. The last set of columns shows the performance measures for the heuristic method. The last row demonstrates the average of each column in the tables.

Table 6.11: Experimental Results for Data Sets with $n = 40$ and $p = 2$

| | | **Matheuristic Method** | | | | | | | **Heuristic Method** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00* | 3.70 | 50.43 | 45 | 2 | 90 | 315.69 | 0.00* | 3.48 | 66.87 | 65 | 1 | 91 | 0.58 |
| 4 | 3 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 579.64 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 0.53 |
| 5 | 2 | 0.00* | 9.49 | 110.08 | 86 | 1 | 86 | 382.57 | 0.00* | 9.37 | 120.86 | 86 | 1 | 86 | 0.61 |
| 5 | 3 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 547.09 | 0.00* | 2.22 | 121.40 | 98 | 1 | 98 | 0.65 |
| 5 | 4 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 606.41 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 0.55 |
| 6 | 2 | 0.00* | 4.78 | 86.16 | 56 | 1 | 89 | 438.96 | 0.00* | 5.45 | 97.90 | 55 | 1 | 77 | 1.25 |
| 6 | 3 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 428.64 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 0.62 |
| 6 | 4 | 0.00* | 3.25 | 153.18 | 81 | 1 | 81 | 924.77 | 0.00* | 1.63 | 11.04 | 83 | 8 | 83 | 0.61 |
| 8 | 2 | 0.00* | 4.77 | 71.31 | 75 | 1 | 75 | 512.95 | 0.00* | 8.33 | 110.27 | 74 | 1 | 74 | 0.73 |
| 8 | 3 | 0.00* | 1.34 | 8.74 | 78 | 5 | 78 | 548.02 | 0.00* | 3.13 | 86.33 | 64 | 1 | 64 | 0.65 |
| 8 | 4 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 535.54 | 0.00* | 5.78 | 156.42 | 96 | 2 | 96 | 0.74 |
| 8 | 6 | 0.00* | -3.14 | -2.33 | 47 | 53 | 100 | 1487.10 | 0.00* | -3.06 | -2.33 | 42 | 58 | 100 | 1.10 |
| 10 | 2 | -1.05 | 0.86 | 50.13 | 65 | 1 | 65 | 759.45 | -1.05 | 2.40 | 69.30 | 29 | 1 | 29 | 0.86 |
| 10 | 3 | 0.00 | 10.06 | 69.10 | 51 | 1 | 51 | 790.35 | 0.00 | 7.62 | 91.92 | 40 | 2 | 40 | 1.01 |
| 10 | 4 | 0.00 | 2.12 | 93.56 | 97 | 1 | 97 | 794.76 | 0.00 | 1.78 | 90.95 | 98 | 1 | 98 | 0.90 |
| 10 | 6 | -0.22 | -0.21 | -0.19 | 100 | 53 | 100 | 1757.12 | -0.22 | -0.21 | -0.19 | 100 | 53 | 100 | 1.71 |
| 12 | 2 | 0.00 | 6.68 | 82.80 | 67 | 1 | 67 | 777.52 | 0.00 | 12.04 | 109.74 | 22 | 1 | 22 | 0.82 |
| 12 | 3 | -6.49 | -0.99 | 50.91 | 61 | 1 | 64 | 933.20 | -6.49 | 5.04 | 68.90 | 40 | 1 | 41 | 1.00 |
| 12 | 4 | 0.00 | 8.71 | 77.52 | 30 | 1 | 30 | 1430.14 | 0.00 | 8.72 | 80.31 | 39 | 1 | 39 | 1.14 |
| 12 | 6 | 0.00 | 4.13 | 66.14 | 37 | 1 | 37 | 1114.35 | 0.00 | 3.74 | 104.74 | 42 | 1 | 42 | 1.35 |
| **Average** | | -0.39 | 2.78 | 48.38 | 73.80 | 31.20 | 80.50 | 783.21 | -0.39 | 3.87 | 69.22 | 68.65 | 21.75 | 74.00 | 0.87 |

Table 6.12: Experimental Results for Data Sets with $n = 40$ and $p = 3$

| | | **Matheuristic Method** | | | | | | | **Heuristic Method** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 12.58 | 68.73 | 44 | 2 | 44 | 433.32 | 0.00 | 38.60 | 131.99 | 19 | 1 | 19 | 0.60 |
| 4 | 3 | 0.00* | 39.74 | 159.22 | 73 | 1 | 73 | 694.90 | 0.00* | 40.23 | 160.01 | 73 | 1 | 73 | 0.65 |
| 5 | 2 | -5.63 | 6.97 | 83.69 | 42 | 1 | 74 | 576.41 | -5.63 | 6.81 | 61.69 | 33 | 1 | 67 | 0.75 |
| 5 | 3 | 0.00* | 22.18 | 199.09 | 24 | 1 | 24 | 517.83 | 0.00* | 45.75 | 203.12 | 12 | 1 | 12 | 0.76 |
| 5 | 4 | 0.00 | 40.35 | 148.73 | 68 | 1 | 68 | 1460.23 | 0.00 | 52.17 | 149.56 | 58 | 2 | 58 | 0.71 |
| 6 | 2 | -4.45 | 3.60 | 48.34 | 39 | 1 | 47 | 789.57 | -4.34 | 9.01 | 67.03 | 3 | 1 | 44 | 2.04 |
| 6 | 3 | -5.00 | 6.16 | 102.91 | 20 | 2 | 88 | 672.71 | -5.00 | 27.64 | 154.75 | 15 | 1 | 66 | 0.91 |
| 6 | 4 | 0.00 | 46.74 | 203.43 | 66 | 1 | 66 | 590.45 | 0.00 | 54.76 | 207.17 | 59 | 1 | 59 | 0.91 |
| 8 | 2 | -8.31 | -2.70 | 27.03 | 23 | 1 | 79 | 902.08 | -8.31 | 2.14 | 39.98 | 18 | 1 | 64 | 1.65 |
| 8 | 3 | 0.00 | 12.64 | 56.26 | 53 | 2 | 53 | 1023.57 | 0.00 | 24.58 | 67.34 | 33 | 1 | 33 | 1.12 |
| 8 | 4 | -0.71 | 20.48 | 139.73 | 33 | 1 | 33 | 1298.67 | -0.71 | 28.86 | 162.06 | 23 | 1 | 23 | 1.25 |
| 8 | 6 | 0.00 | 51.31 | 226.98 | 50 | 1 | 50 | 1914.95 | 0.00 | 54.23 | 226.82 | 49 | 1 | 49 | 1.15 |
| 10 | 2 | -24.35 | -18.19 | -3.64 | 15 | 1 | 100 | 1883.48 | -20.35 | -7.63 | 13.83 | 1 | 1 | 88 | 1.71 |
| 10 | 3 | -0.87 | 7.87 | 44.43 | 34 | 1 | 47 | 4411.87 | -0.87 | 17.60 | 55.04 | 32 | 1 | 35 | 1.10 |
| 10 | 4 | 0.00 | 13.39 | 71.27 | 64 | 1 | 64 | 3590.31 | 0.00 | 33.38 | 90.01 | 11 | 1 | 11 | 1.75 |
| 10 | 6 | -3.11 | 20.96 | 109.74 | 10 | 1 | 71 | 9560.39 | -3.11 | 24.59 | 110.20 | 6 | 1 | 62 | 1.68 |
| 12 | 2 | -20.87 | -15.74 | 4.59 | 6 | 1 | 99 | 3837.13 | -19.93 | -10.35 | 12.30 | 4 | 1 | 93 | 2.00 |
| 12 | 3 | -2.79 | 4.88 | 51.83 | 20 | 1 | 44 | 23476.87 | -2.79 | 6.04 | 53.94 | 24 | 1 | 45 | 2.39 |
| 12 | 4 | -2.13 | 10.73 | 73.97 | 5 | 1 | 19 | 12383.55 | -2.13 | 16.02 | 58.72 | 2 | 1 | 20 | 1.38 |
| 12 | 6 | -4.24 | 20.93 | 133.15 | 13 | 1 | 51 | 67886.85 | -4.24 | 20.33 | 150.83 | 16 | 1 | 50 | 1.98 |
| **Average** | | -4.12 | 15.24 | 97.47 | 35.10 | 1.15 | 59.70 | 6895.26 | -3.87 | 24.24 | 108.82 | 24.55 | 1.05 | 48.55 | 1.32 |

Table 6.13: Experimental Results for Data Sets with $n = 40$ and $p = 4$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 24.82 | 111.03 | 21 | 1 | 21 | 559.83 | 0.00 | 44.22 | 151.30 | 14 | 1 | 14 | 0.90 |
| 4 | 3 | -2.38 | 34.39 | 108.81 | 14 | 1 | 37 | 488.19 | -2.38 | 19.51 | 231.12 | 12 | 1 | 41 | 0.74 |
| 5 | 2 | -2.62 | 11.11 | 73.62 | 24 | 1 | 31 | 759.19 | -2.62 | 15.71 | 57.70 | 5 | 1 | 11 | 1.54 |
| 5 | 3 | -6.17 | 17.49 | 98.55 | 21 | 1 | 69 | 738.40 | -6.17 | 25.31 | 124.85 | 24 | 1 | 62 | 1.01 |
| 5 | 4 | -2.40 | 44.80 | 172.37 | 25 | 1 | 36 | 859.22 | -2.40 | 50.23 | 172.65 | 24 | 1 | 35 | 0.80 |
| 6 | 2 | 0.00 | 17.78 | 59.13 | 27 | 1 | 27 | 1375.69 | 0.00 | 21.17 | 109.40 | 9 | 1 | 9 | 1.29 |
| 6 | 3 | 0.00 | 8.01 | 89.83 | 52 | 1 | 52 | 1282.90 | 0.00 | 27.08 | 215.14 | 38 | 1 | 38 | 0.88 |
| 6 | 4 | 0.00 | 22.75 | 121.94 | 78 | 1 | 78 | 1025.26 | 0.00 | 48.06 | 225.73 | 45 | 1 | 45 | 1.02 |
| 8 | 2 | -27.70 | -22.50 | 1.42 | 6 | 1 | 99 | 3969.04 | -26.78 | -11.94 | 22.80 | 3 | 1 | 75 | 1.60 |
| 8 | 3 | -7.91 | 9.24 | 55.31 | 16 | 1 | 46 | 7949.39 | -7.91 | 20.35 | 83.87 | 7 | 1 | 28 | 1.94 |
| 8 | 4 | -1.77 | 27.97 | 111.28 | 60 | 1 | 60 | 36187.11 | -1.77 | 36.69 | 147.63 | 50 | 1 | 50 | 1.59 |
| 8 | 6 | -0.58 | 57.84 | 211.61 | 34 | 1 | 34 | 2273.50 | -0.58 | 62.78 | 217.12 | 33 | 1 | 33 | 1.59 |
| 10 | 2 | -28.12 | -24.39 | -6.21 | 8 | 1 | 100 | 55843.61 | -27.89 | -16.67 | 5.72 | 1 | 1 | 96 | 2.63 |
| 10 | 3 | -3.6 | 7.22 | 29.03 | 10 | 5 | 40 | TL | -1.29 | 19.38 | 69.44 | 3 | 1 | 14 | 1.55 |
| 10 | 4 | -13.03 | 9.07 | 47.18 | 8.7 | 4.35 | 56.52 | TL | -13.03 | 24.39 | 130.35 | 5 | 1 | 45 | 1.43 |
| 10 | 6 | 0 | 48.56 | 120.1 | 28 | 4 | 28 | TL | 0 | 42.77 | 124.49 | 42 | 1 | 42 | 1.88 |
| 12 | 2 | -46.24 | -41.26 | -34.11 | 9.38 | 3.13 | 100 | TL | -45.64 | -30.43 | -10.85 | 1 | 1 | 100 | 3.05 |
| 12 | 3 | -6.79 | 8.36 | 26.52 | 16.67 | 16.67 | 50 | TL | -5.14 | 14.08 | 54.7 | 6 | 1 | 19 | 2.26 |
| 12 | 4 | 4.98 | 17.81 | 34.05 | 20 | 20 | 0 | TL | 0 | 27.28 | 92.88 | 5 | 1 | 5 | 1.93 |
| 12 | 6 | -1.57 | 12.39 | 57.39 | 16.67 | 8.33 | 25 | TL | -3.56 | 28.42 | 168.09 | 10 | 1 | 30 | 2.13 |
| **Average** | | -7.3 | 14.57 | 74.44 | 24.77 | 3.72 | 49.48 | 8716.26 | -7.36 | 23.42 | 119.71 | 16.85 | 1.00 | 39.60 | 1.59 |

Table 6.11 shows that $\%Gap_{Best}$ is zero in all data sets where the optimal solution is known for both the matheuristic and the heuristic method. For the data sets where the optimal solution is unknown, the matheuristic and heuristic find the same best objective value, which is smaller than or equal to the base objective. $\%Gap_{Avg}$ and $\%Gap_{Worst}$ are lower for the matheuristic on average, although there are instances where the heuristic finds a lower $\%Gap_{Avg}$ and $\%Gap_{Worst}$ such as the data set with $m = 6$ and $q = 4$.

$\#Best$ and $\#IBase$ are higher for the matheuristic on average. However, the difference is about 5%, so it can be said that the difference is not significant. The matheuristic and heuristic show a similar pattern in $\#Worst$. It can be seen that the heuristic method takes significantly less computation time than the matheuristic method.

Table 6.12 demonstrates that the optimal solution is known for the data sets where $m = 4$, $p = 3$ and $m = 5$, $p = 3$. In these data sets, both matheuristic and heuristic found the optimal solution. On the other data sets, both methods found an objective value which smaller than or equal to the base objective value. However, the matheuristic found a better solution than the heuristic method in a few data sets, e.g., $m = 6$ and $q = 2$. $\#Best$ and $\#IBase$ are higher for the matheuristic on average while $\#Worst$ is very close to each other. The total computation time of the matheuristic method exceeds 18 hours with $m = 12$. On the other hand, the total computation for the heuristic method is only up to 2 seconds.

Table 6.13 shows that the time limit is exceeded for the matheuristic method in data sets where $m > 8$. The matheuristic method mostly finds the base objective value in these data sets. However, the matheuristic could not find the base solution in the data set with $m = 12$ and $q = 4$. In this data set, the heuristic method found a better solution than the matheuristic method. Also, it can be seen that the heuristic method lasts only a few seconds, even in the data sets where $m > 8$. Therefore, it can be said that the heuristic method becomes superior to the matheuristic method as the data sets grow.

Table 6.11-6.13 show that $\%Gap_{Best}$ is the same for matheuristic and heuristic methods in most instances. However, the matheuristic found a better best solution than the heuristic method in a few data sets, and the average $\#Best$ is higher for the

matheuristic method. Moreover, it can be said that the matheuristic method performs better in terms of the average and worst solutions. The average $\#IBase$ is also higher for the matheuristic method. However, there are data sets where the time limit was exceeded by the matheuristic method, while the heuristic method can find the base solution in a few seconds. The heuristic method shows a much better performance in terms of the computation time.

The heuristic method was also experimented on with data sets consisting of 200, 500, and 1000 data points. It is observed that it can yield the base objective value in almost all data sets. $F1_{MISOCP}$ and $F2_{MISOCP}$ models and the matheuristic method do not apply to data sets of these sizes in reasonable times.

Table 6.14 summarizes the experimental results according to the number of data points and number of clusters. The average values of $\%Gap_{Best}$, $\%Gap_{Avg}$, $\%Gap_{Worst}$, $\#Best$, $\#Worst$, $\#IBase$, and total time are presented for the heuristic method.

Table 6.14: Experimental Results by Number of Data Points and Number of Clusters

| n | p | Average $\%Gap_{Best}$ | Average $\%Gap_{Avg}$ | Average $\%Gap_{Worst}$ | Average $\#Best$ | Average $\#Worst$ | Average $\#IBase$ | Average Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 40 | 2 | -0.7 | 3.9 | 69.2 | 68.7 | 21.8 | 74.0 | 0.87 |
| 40 | 3 | -3.9 | 24.2 | 108.8 | 24.6 | 1.1 | 48.6 | 1.32 |
| 40 | 4 | -7.4 | 23.4 | 119.7 | 16.9 | 1.0 | 39.6 | 1.59 |
| 50 | 2 | -0.3 | 4.6 | 74.7 | 69.6 | 18.0 | 72.3 | 1.18 |
| 50 | 3 | -1.7 | 25.5 | 104.4 | 29.0 | 1.1 | 41.7 | 1.53 |
| 50 | 4 | -4.6 | 24.4 | 120.4 | 19.3 | 1.1 | 38.1 | 2.48 |
| 80 | 2 | -0.7 | 2.3 | 53.8 | 72.5 | 14.8 | 76.1 | 2.30 |
| 80 | 3 | -3.0 | 20.7 | 94.4 | 33.3 | 1.2 | 47.6 | 3.97 |
| 80 | 4 | -2.3 | 27.7 | 113.8 | 19.0 | 1.1 | 31.1 | 4.36 |
| 100 | 2 | -0.3 | 3.3 | 81.0 | 77.4 | 17.4 | 77.4 | 3.42 |
| 100 | 3 | -1.3 | 21.6 | 101.0 | 32.2 | 1.1 | 40.9 | 5.10 |
| 100 | 4 | -3.7 | 24.2 | 118.6 | 26.4 | 1.0 | 42.7 | 6.34 |
| 200 | 2 | -0.4 | 3.8 | 45.9 | 68.4 | 27.3 | 70.7 | 10.87 |
| 200 | 3 | -0.7 | 23.4 | 103.7 | 39.9 | 1.2 | 47.8 | 18.02 |
| 200 | 4 | -3.2 | 22.8 | 109.2 | 25.5 | 1.0 | 41.8 | 20.61 |
| 500 | 2 | -0.2 | 5.1 | 48.1 | 65.9 | 21.3 | 67.1 | 90.82 |
| 500 | 3 | -0.7 | 21.6 | 110.5 | 40.2 | 1.0 | 45.8 | 95.57 |
| 500 | 4 | -1.5 | 26.3 | 114.7 | 32.1 | 1.0 | 38.1 | 105.68 |
| 1000 | 2 | -0.2 | 4.8 | 36.4 | 61.9 | 25.9 | 63.7 | 368.23 |
| 1000 | 3 | -1.1 | 21.2 | 114.7 | 51.3 | 1.0 | 61.9 | 392.15 |
| 1000 | 4 | -1.1 | 25.4 | 117.5 | 33.7 | 1.0 | 43.5 | 394.44 |

Table 6.14 demonstrates that average $\%Gap_{Best}$ is smaller than zero for all values

of $n$ and $p$. $\%Gap_{Avg}$ is about 17% and $\#IBase$ is about 52% on average. These show that the heuristic method successfully finds the optimal or base solution through initializations. Average $\#Best$ is higher than average $\#Worst$ in each row. This shows that the matheuristic is more likely to converge the best solution than the worst solution on average. Although there is an increase in the computation time as the data set grows, this increase is negligibly small, and the average computation time is less than 7 minutes, even for the largest data sets.

The matheuristic and heuristic methods experimented with the same initializations of cluster centers. Therefore, the performance of the solutions obtained by the matheuristic and heuristic methods can be compared for each initialization. Table 6.15 demonstrates the results of performance measures $\%Mth$, $\%Hrs$, and $\%Eql$ by the number of data points and the number of clusters.

Table 6.15: Comparison of Experimental Results For the Matheuristic and Heuristic Methods

| **n** | **p** | $\%Eql$ | $\%Mth$ | $\%Hrs$ |
|---|---|---|---|---|
| 40 | 2 | 75.5 | 15.7 | 8.8 |
| 40 | 3 | 34.2 | 46.8 | 19.1 |
| 40 | 4 | 26.2 | 55.7 | 18.1 |
| 50 | 2 | 79.5 | 14.7 | 5.9 |
| 50 | 3 | 42.8 | 41.7 | 15.5 |
| 50 | 4 | 22.8 | 53.5 | 23.8 |
| 80 | 2 | 82.2 | 11.5 | 6.4 |
| 80 | 3 | 37.6 | 41.4 | 21.0 |
| 80 | 4 | 23.8 | 50.7 | 25.4 |
| 100 | 2 | 83.5 | 11.0 | 5.4 |
| 100 | 3 | 39.5 | 44.9 | 15.5 |
| 100 | 4 | 28.7 | 45.3 | 26.0 |
| **Average** | | 48.1 | 36.0 | 15.9 |

$\%Eql$ shows that the matheuristic and heuristic methods found the same objective in most initializations. On the other hand, $\%Mth$ is greater than $\%Hrs$ in each row of the table. This shows that the matheuristic method found better objective values than the heuristic method on average for the initializations where the heuristic and matheuristic found different solutions.

In summary, solving $F1_{MISOCP}$ and $F2_{MISOCP}$ models can only be useful for the smallest data sets, and the computation times of the models are not reasonable for larger data sets. Although the solution time of the problem is significantly decreased with the matheuristic method for one initialization, the total computation time of multiple initializations is still not reasonable for larger data sets. The iterative heuristic method can solve the CLFS problem in almost less than 10 minutes for even 1000 data points. Also, it can find the base objective value for almost all data sets with 100 initializations according to conducted experiments.

# CHAPTER 7

# CONCLUSION

Clustering is a widely studied unsupervised learning method that groups data points according to their similarities. It enables organizations to gain insights from their data for informed decision-making. Partitioning clustering aims to group data points into a predetermined number of clusters to optimize a criterion. In this thesis, we consider a partitioning clustering problem where each data point is assigned to one cluster, and a cluster center is selected from the data points within each cluster.

High-dimensional data causes significant challenges for clustering methods. Data sets may contain irrelevant or redundant features, which mask the hidden clustering structure and reduce the learning performance and computational efficiency. Feature selection is the most used technique to cope with redundant and irrelevant features in clustering problems. Most feature selection methods proposed for clustering perform global feature selection. These methods select a common relevant set of features for all clusters. However, the set of relevant features for clustering may vary across clusters. In this thesis, we focus on clustering with localized feature selection, where a relevant set of features is selected for each cluster separately.

Our problem is forming $p$ clusters by assigning each data point to a cluster, selecting a cluster center from the data points within the cluster, and selecting $q$ relevant features for each cluster. The features selected may differ for each cluster, and a feature can be chosen for multiple or no clusters. The objective is to minimize the total Euclidean distance from data points to their corresponding cluster center according to relevant features. We called this problem Clustering with Localized Feature Selection (CLFS) in this thesis.

In the literature, many mathematical programming-based exact and heuristic solution approaches were developed for the partitioning clustering problem. On the other hand, the number of studies proposing mathematical programming-based approaches to solve the clustering problem with localized feature selection is limited. This thesis is the first study to focus on a clustering problem with localized feature selection where the partitioning criterion is defined based on Euclidean distance.

We have provided two mathematical programming formulations for the CLFS: a center-based formulation ($F1$) and a cluster-based formulation ($F2$). In the center-based formulation, clusters are represented by cluster centers, while clusters are represented by indices in the cluster-based formulation. $F1$ and $F2$ decide the assignments of data points to clusters, cluster centers, and relevant features for each cluster simultaneously. We show that $F1$ and $F2$ can be reformulated as MISOCP models.

The proposed $F1_{MISOCP}$ and $F2_{MISOCP}$ models were tested on generated data sets. The experiments show that $F1_{MISOCP}$ performs better in computation time. Also, it is seen that both models' computation time rapidly increases with the number of data points, features, and clusters. Therefore, a matheuristic and a heuristic method were developed to solve the CLFS problem for larger data sets.

The matheuristic method decomposes the three groups of decisions into two subproblems: $CLFS^{AF}$ and $CLFS^{C}$. $CLFS^{AF}$ aims to determine assignments of data points to clusters and relevant features when the cluster centers are given. $CLFS^{C}$ aims to select a cluster center for each cluster when the data points within the cluster and relevant features are given. Two MISOCP formulations are proposed for the $CLFS^{AF}$ problem, and an integer programming model is proposed for the $CLFS^{C}$ problem. The mathematical models for subproblems are solved iteratively to find a solution for the CLFS problem.

The matheuristic method is experimented with generated data sets. It is observed that the matheuristic method finds the optimal solution for all the data sets where the optimal solution is known. It finds the base objective value in almost all data sets where the optimal solution is unknown. However, the computation time of the matheuristic accelerates also as the data sets grow. This is because the method's subproblems are still solved using mathematical programming.

The heuristic method decomposes three groups of decisions in the CLFS problem into subproblems such that each subproblem gives one group of decisions. The first subproblem $CLFS^A$ aims to assign each data point to a cluster when the cluster center and relevant features are given for each cluster. The second subproblem $CLFS^F$ aims to determine relevant features for each cluster when each cluster is given. The third subproblem $CLFS^C$ aims to select a cluster center for each cluster when the data points within the cluster and relevant features are given. Optimization algorithms are used to solve $CLFS^A$ and $CLFS^C$ problems. A heuristic method is proposed for the $CLFS^F$ problem. The subproblems are solved iteratively to find a solution for the CLFS problem. The heuristic method is also tested with generated data sets, and the results are compared with the matheuristic method's.

The results show that the matheuristic method is better than the heuristic method for finding the base objective value in a random initialization of cluster centers. However, the iterative heuristic method outperforms the matheuristic method in terms of computation time when the methods are performed with many initializations. According to experiments, the heuristic method can find the optimal solution for the data sets where the optimal solution was reported. Also, it finds the base objective value in a significantly short computation time for almost all data sets up to 1000 data points.

This thesis contributes literature in several ways. First, the clustering problem with localized feature selection with Euclidean distance clustering criterion is addressed for the first time in the literature. Second, it proposes two exact solutions approaches to the problem. Third, it is the first study to propose a matheuristic method that combines mathematical programming with a heuristic approach for the clustering problem with feature selection. Fourth, it proposes a heuristic method that demonstrates a strong performance in solving large-scale CLFS problems.

The proposed mathematical models can also be modified for the clustering problem with global feature selection. To our knowledge, no study provides a mathematical programming formulation for the clustering problem with global feature selection where the optimization criterion is based on Euclidean distance.

A future research direction can be clustering with localized feature selection problems where the total number of features to be selected is given, but the number of

features selected for each cluster will be decided. This adds additional complexity to the problem by requiring the decision on the number of features for each cluster. Another future research direction can be clustering with localized feature selection problems where the cluster center can be any point in the feature space instead of restricting to one of the data points within the clusters. The proposed methods were experimented on data sets that only included continuous features and spherical clusters. For further analysis, the proposed methods can be experimented on different data sets with arbitrary-shaped clusters or categorical features.

# REFERENCES


Alelyani, S., Tang, J., & Liu, H. (2018). Feature selection for clustering: A review. *Data Clustering*, 29–60.

Benati, S., & Garcıa, S. (2014). A mixed integer linear model for clustering with variable selection. *Computers & operations research*, *43*, 280–285.

Benati, S., Garcıa, S., & Puerto, J. (2018). Mixed integer linear programming and heuristic methods for feature selection in clustering. *Journal of the Operational Research Society*, *69*(9), 1379–1395.

Brusco, M. J., & Cradit, J. D. (2001). A variable-selection heuristic for k-means clustering. *Psychometrika*, *66*, 249–270.

Gnägi, M., & Baumann, P. (2021). A matheuristic for large-scale capacitated clustering. *Computers & operations research*, *132*, 105304.

Guha, S., Rastogi, R., & Shim, K. (1998). Cure: An efficient clustering algorithm for large databases. *ACM Sigmod record*, *27*(2), 73–84.

Guha, S., Rastogi, R., & Shim, K. (2000). Rock: A robust clustering algorithm for categorical attributes. *Information systems*, *25*(5), 345–366.

Hancer, E., Xue, B., & Zhang, M. (2020). A survey on feature selection approaches for clustering. *Artificial Intelligence Review*, *53*(6), 4519–4545.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, *31*(8), 651–666.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM computing surveys (CSUR)*, *31*(3), 264–323.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

Kriegel, H.-P., Kröger, P., & Zimek, A. (2009). Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *Acm transactions on knowledge discovery from data (tkdd)*, *3*(1), 1–58.

Li, Y., Dong, M., & Hua, J. (2008a). Localized feature selection for clustering. *Pattern Recognition Letters*, *29*(1), 10–18.

Li, Y., Dong, M., & Hua, J. (2008b). Simultaneous localized feature selection and model detection for gaussian mixtures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(5), 953–960.

MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(14), 281–297.

Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European journal of operational research*, *187*(3), 1429–1448.

Ólafsson, S., & Yang, J. (2005). Intelligent partitioning for feature selection. *INFORMS Journal on Computing*, *17*(3), 339–355.

Öz, S. Ö. (2019). *Mixed integer programming and heuristics approaches for clustering with cluster-based feature selection* (Master's thesis). Middle East Technical University.

Park, H.-S., & Jun, C.-H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, *36*(2), 3336–3341.

Rao, M. (1971). Cluster analysis and mathematical programming. *Journal of the American statistical association*, *66*(335), 622–626.

Rokach, L., & Maimon, O. (2005). Clustering methods. *Data mining and knowledge discovery handbook*, 321–352.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681.

Şener, İ. A. (2022). *A memetic algorithm for clustering with cluster based feature selection* (Master's thesis). Middle East Technical University.

Steinley, D., & Brusco, M. J. (2008). A new variable weighting and selection procedure for k-means cluster analysis. *Multivariate Behavioral Research*, *43*(1), 77–108.

Vinod, H. D. (1969). Integer programming and the theory of grouping. *Journal of the American Statistical association*, *64*(326), 506–519.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, *16*(3), 645–678.

Yang, J., & Olafsson, S. (2006). Optimization-based feature selection with adaptive instance sampling. *Computers & Operations Research*, *33*(11), 3088–3106.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). Birch: An efficient data clustering method for very large databases. *ACM sigmod record*, *25*(2), 103–114.

## Appendix A

## A MATHEURISTIC ALGORITHM FOR CLFS

**Algorithm 5** A Matheuristic Algorithm for CLFS

**Data:** Data set, $p$, $q$, $R$ and $TL$

**Result:** $A_l, C_l, F_l, obj_M$

Initialize $obj_M = \infty$, $rep\_count = 0$ and $total$

**while** $rep\_count < R$ *and* $total\_time < TL$ **do**

Update $total\_time$

$rep\_count = rep\_count + 1$

Initialize $obj_{AF} = 1$ and $obj_C = 0$

Initialize a random center for each cluster $\rightarrow$ $C_l$

**while** $|obj_{AF} - obj_C| > \epsilon$ **do**

Solve $CLFS^{AF}$ with $C_l \rightarrow$ $A_l$,$F_l$, $obj_{AF}$

**if** $|obj_{AF} - obj_C| > \epsilon$ **then**

Solve $CLFS^{C}$ with $A_l$ and $F_l \rightarrow$ $C_l$, $obj_C$

**end**

**else**

Break

**end**

**end**

**if** $obj_{AF} <= obj_C$ & $obj_{AF} < obj_M$ **then**

$obj_M \leftarrow obj_{AF}$

**end**

**else if** $obj_C < obj_{AF}$ & $obj_C < obj_M$ **then**

$obj_M \leftarrow obj_C$

**end**

**end**

# Appendix B

## EXPERIMENTAL RESULTS OF $F1\_MISOCP$ AND $F2\_MISOCP$ MODELS

This appendix provides the experimental results of $F1\_MISOCP$ and $F2\_MISOCP$ models for datasets with 50,80,100 data points. TL shows that the 1 hour time limit is exceeded for the problem instance. The performance measures used in the tables are as follows:

- $\%Gap_R$: MIP Relative Gap
- $\%Gap_B$: Percent Gap from the Base Objective
- Time: Computation Time (seconds)

Table B.1: Experimental Results for Data Sets with $n = 50$ and $p = 2$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 0.00 | -2.60 | 210.66 | 0.00 | -2.60 | 520.09 |
| 4 | 3 | 0.00 | 0.00 | 81.16 | 0.00 | 0.00 | 279.09 |
| 5 | 2 | 0.00 | 0.00 | 475.13 | 0.00 | 0.00 | 1053.27 |
| 5 | 3 | 0.00 | 0.00 | 328.31 | 0.00 | 0.00 | 1406.84 |
| 5 | 4 | 0.00 | 0.00 | 135.75 | 0.00 | 0.00 | 483.50 |
| 6 | 2 | 0.00 | -1.49 | 1570.86 | 0.00 | -1.49 | 2100.16 |
| 6 | 3 | 0.00 | 0.00 | 1467.08 | 0.00 | 0.00 | 3142.70 |
| 6 | 4 | 0.00 | 0.00 | 688.67 | 0.00 | 0.00 | 1608.95 |
| 8 | 2 | 94.94 | -1.80 | TL | 100.00 | -1.80 | TL |
| 8 | 3 | 95.50 | 2.62 | TL | 100.00 | 0.00 | TL |
| 8 | 4 | 95.11 | 0.58 | TL | 100.00 | 0.00 | TL |
| 8 | 6 | 0.00 | 0.00 | 1662.14 | 69.01 | 0.00 | TL |
| 10 | 2 | 99.35 | 2.61 | TL | 100.00 | 1.92 | TL |
| 10 | 3 | 100.00 | 2.28 | TL | 100.00 | 0.00 | TL |
| 10 | 4 | 100.00 | 6.27 | TL | 100.00 | 0.00 | TL |
| 10 | 6 | 99.23 | 7.28 | TL | 100.00 | 0.00 | TL |
| 12 | 2 | 100.00 | 8.25 | TL | 100.00 | 0.00 | TL |
| 12 | 3 | 99.84 | 11.45 | TL | 100.00 | 24.91 | TL |
| 12 | 4 | 100.00 | 24.32 | TL | 100.00 | 18.62 | TL |
| 12 | 6 | 100.00 | 14.60 | TL | 100.00 | 3.47 | TL |
| Average | | 54.20 | 3.72 | 735.53 | 58.45 | 2.15 | 1324.33 |

Table B.2: Experimental Results for Data Sets with $n = 50$ and $p = 3$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 4.92 | TL | 100.00 | -2.00 | TL |
| 4 | 3 | 97.70 | 4.56 | TL | 100.00 | 1.03 | TL |
| 5 | 2 | 100.00 | 0.70 | TL | 100.00 | 14.26 | TL |
| 5 | 3 | 100.00 | 4.05 | TL | 100.00 | -0.71 | TL |
| 5 | 4 | 97.93 | 7.62 | TL | 100.00 | 0.39 | TL |
| 6 | 2 | 100.00 | 16.95 | TL | 100.00 | 0.00 | TL |
| 6 | 3 | 100.00 | 8.95 | TL | 100.00 | 9.00 | TL |
| 6 | 4 | 100.00 | 4.14 | TL | 100.00 | 0.40 | TL |
| 8 | 2 | 100.00 | 6.41 | TL | 100.00 | -4.49 | TL |
| 8 | 3 | 100.00 | 10.60 | TL | 100.00 | 8.54 | TL |
| 8 | 4 | 100.00 | 18.91 | TL | 100.00 | 47.60 | TL |
| 8 | 6 | 100.00 | 12.97 | TL | 100.00 | 4.82 | TL |
| 10 | 2 | 100.00 | 6.96 | TL | 100.00 | -2.55 | TL |
| 10 | 3 | 100.00 | 9.37 | TL | 100.00 | 29.84 | TL |
| 10 | 4 | 100.00 | 11.03 | TL | 100.00 | 68.15 | TL |
| 10 | 6 | 100.00 | 10.55 | TL | 100.00 | 94.86 | TL |
| 12 | 2 | 100.00 | 9.95 | TL | 100.00 | 17.25 | TL |
| 12 | 3 | 100.00 | 24.85 | TL | 100.00 | 58.88 | TL |
| 12 | 4 | 100.00 | 21.61 | TL | 100.00 | 34.90 | TL |
| 12 | 6 | 100.00 | 3.25 | TL | 100.00 | 62.34 | TL |
| Average | | 99.78 | 9.92 | - | 100.00 | 22.13 | - |

Table B.3: Experimental Results for Data Sets with $n = 50$ and $p = 4$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 11.22 | TL | 100.00 | 6.68 | TL |
| 4 | 3 | 100.00 | 7.88 | TL | 100.00 | 65.60 | TL |
| 5 | 2 | 100.00 | 38.03 | TL | 100.00 | 28.76 | TL |
| 5 | 3 | 100.00 | 38.19 | TL | 100.00 | 63.97 | TL |
| 5 | 4 | 99.98 | 18.69 | TL | 100.00 | 8.81 | TL |
| 6 | 2 | 100.00 | 15.59 | TL | 100.00 | 15.44 | TL |
| 6 | 3 | 100.00 | 27.63 | TL | 100.00 | 68.27 | TL |
| 6 | 4 | 100.00 | 8.85 | TL | 100.00 | 85.49 | TL |
| 8 | 2 | 100.00 | -13.58 | TL | 100.00 | -8.76 | TL |
| 8 | 3 | 100.00 | 24.77 | TL | 100.00 | 41.40 | TL |
| 8 | 4 | 100.00 | 35.15 | TL | 100.00 | 134.14 | TL |
| 8 | 6 | 100.00 | 13.66 | TL | 100.00 | 17.61 | TL |
| 10 | 2 | 100.00 | 2.69 | TL | 100.00 | -12.51 | TL |
| 10 | 3 | 100.00 | 22.02 | TL | 100.00 | 33.17 | TL |
| 10 | 4 | 100.00 | 27.64 | TL | 100.00 | 96.13 | TL |
| 10 | 6 | 100.00 | 34.91 | TL | 100.00 | 44.56 | TL |
| 12 | 2 | 100.00 | -2.74 | TL | 100.00 | 4.78 | TL |
| 12 | 3 | 100.00 | 47.49 | TL | 100.00 | 87.46 | TL |
| 12 | 4 | 100.00 | 14.37 | TL | 100.00 | 98.68 | TL |
| 12 | 6 | 100.00 | 71.66 | TL | 100.00 | 88.14 | TL |
| Average | | 100.00 | 22.21 | - | 100.00 | 48.39 | - |

Table B.4: Experimental Results for Data Sets with $n = 80$ and $p = 2$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 0.00 | -0.06 | 2798.70 | 97.49 | 0.00 | TL |
| 4 | 3 | 0.00 | 0.00 | 795.52 | 61.57 | 0.00 | TL |
| 5 | 2 | 99.90 | 1.13 | TL | 100.00 | 4.39 | TL |
| 5 | 3 | 100.00 | 0.13 | TL | 100.00 | 4.88 | TL |
| 5 | 4 | 0.00 | 0.00 | 1070.89 | 100.00 | 0.00 | TL |
| 6 | 2 | 100.00 | -3.52 | TL | 100.00 | 19.27 | TL |
| 6 | 3 | 100.00 | 2.67 | TL | 100.00 | 5.27 | TL |
| 6 | 4 | 100.00 | 2.99 | TL | 100.00 | 4.98 | TL |
| 8 | 2 | 100.00 | 0.76 | TL | 100.00 | 1.89 | TL |
| 8 | 3 | 100.00 | 9.31 | TL | 100.00 | 37.33 | TL |
| 8 | 4 | 100.00 | 0.45 | TL | 100.00 | 9.19 | TL |
| 8 | 6 | 99.94 | 1.63 | TL | 100.00 | 15.23 | TL |
| 10 | 2 | 100.00 | 4.49 | TL | 100.00 | 16.43 | TL |
| 10 | 3 | 100.00 | 4.85 | TL | 100.00 | 37.40 | TL |
| 10 | 4 | 100.00 | 2.65 | TL | 100.00 | 11.66 | TL |
| 10 | 6 | 100.00 | 6.41 | TL | 100.00 | 107.08 | TL |
| 12 | 2 | 100.00 | 2.32 | TL | 100.00 | 54.89 | TL |
| 12 | 3 | 100.00 | 11.11 | TL | 100.00 | 11.13 | TL |
| 12 | 4 | 100.00 | 10.96 | TL | 100.00 | 67.23 | TL |
| 12 | 6 | 100.00 | 14.92 | TL | 100.00 | 51.75 | TL |
| Average | | 84.99 | 3.66 | 1555.04 | 97.95 | 23.00 | - |

Table B.5: Experimental Results for Data Sets with $n = 80$ and $p = 3$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | -1.64 | TL | 100.00 | 6.27 | TL |
| 4 | 3 | 100.00 | 18.55 | TL | 100.00 | 15.56 | TL |
| 5 | 2 | 100.00 | 17.73 | TL | 100.00 | 37.75 | TL |
| 5 | 3 | 100.00 | 16.92 | TL | 100.00 | 51.48 | TL |
| 5 | 4 | 100.00 | 6.20 | TL | 100.00 | 6.65 | TL |
| 6 | 2 | 100.00 | 8.69 | TL | 100.00 | 24.54 | TL |
| 6 | 3 | 100.00 | 17.29 | TL | 100.00 | 78.61 | TL |
| 6 | 4 | 100.00 | 6.91 | TL | 100.00 | 105.13 | TL |
| 8 | 2 | 100.00 | 16.45 | TL | 100.00 | 13.77 | TL |
| 8 | 3 | 100.00 | 11.72 | TL | 100.00 | 76.02 | TL |
| 8 | 4 | 100.00 | 20.29 | TL | 100.00 | 148.53 | TL |
| 8 | 6 | 100.00 | 16.30 | TL | 100.00 | 235.71 | TL |
| 10 | 2 | 100.00 | 16.05 | TL | 100.00 | 23.55 | TL |
| 10 | 3 | 100.00 | 26.68 | TL | 100.00 | 54.64 | TL |
| 10 | 4 | 100.00 | 36.48 | TL | 100.00 | 123.63 | TL |
| 10 | 6 | 100.00 | 38.14 | TL | 100.00 | 190.67 | TL |
| 12 | 2 | 100.00 | -7.15 | TL | 100.00 | 0.27 | TL |
| 12 | 3 | 100.00 | 29.93 | TL | 100.00 | 72.22 | TL |
| 12 | 4 | 100.00 | 46.23 | TL | 100.00 | 116.87 | TL |
| 12 | 6 | 100.00 | 43.10 | TL | 100.00 | 132.30 | TL |
| Average | | 100.00 | 19.24 | - | 100.00 | 75.71 | - |

Table B.6: Experimental Results for Data Sets with $n = 80$ and $p = 4$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 15.90 | TL | 100.00 | 57.24 | TL |
| 4 | 3 | 100.00 | 28.77 | TL | 100.00 | 93.70 | TL |
| 5 | 2 | 100.00 | 4.64 | TL | 100.00 | 61.21 | TL |
| 5 | 3 | 100.00 | 49.38 | TL | 100.00 | 95.62 | TL |
| 5 | 4 | 100.00 | 13.07 | TL | 100.00 | 68.86 | TL |
| 6 | 2 | 100.00 | 25.38 | TL | 100.00 | 64.73 | TL |
| 6 | 3 | 100.00 | 47.73 | TL | 100.00 | 103.35 | TL |
| 6 | 4 | 100.00 | 58.38 | TL | 100.00 | 159.01 | TL |
| 8 | 2 | 100.00 | 22.87 | TL | 100.00 | 54.05 | TL |
| 8 | 3 | 100.00 | 56.30 | TL | 100.00 | 86.69 | TL |
| 8 | 4 | 100.00 | 30.55 | TL | 100.00 | 129.51 | TL |
| 8 | 6 | 100.00 | 32.64 | TL | 100.00 | 227.14 | TL |
| 10 | 2 | 100.00 | 18.83 | TL | 100.00 | 21.34 | TL |
| 10 | 3 | 100.00 | 18.00 | TL | 100.00 | 134.26 | TL |
| 10 | 4 | 100.00 | 76.04 | TL | 100.00 | 171.36 | TL |
| 10 | 6 | 100.00 | 29.55 | TL | 100.00 | 256.03 | TL |
| 12 | 2 | 100.00 | -2.36 | TL | 100.00 | 26.06 | TL |
| 12 | 3 | 100.00 | 53.37 | TL | 100.00 | 84.19 | TL |
| 12 | 4 | 100.00 | 62.59 | TL | 100.00 | 186.91 | TL |
| 12 | 6 | 100.00 | 31.44 | TL | 100.00 | 232.56 | TL |
| Average | | 100.00 | 33.65 | - | 100.00 | 115.69 | - |

Table B.7: Experimental Results for Data Sets with $n = 100$ and $p = 2$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 0.14 | TL | 100.00 | 12.83 | TL |
| 4 | 3 | 0.00 | 0.00 | 2394.64 | 100.00 | 0.00 | TL |
| 5 | 2 | 100.00 | 2.59 | TL | 100.00 | 11.31 | TL |
| 5 | 3 | 100.00 | 1.12 | TL | 100.00 | 3.96 | TL |
| 5 | 4 | 98.06 | 0.00 | TL | 100.00 | 3.38 | TL |
| 6 | 2 | 100.00 | 3.22 | TL | 100.00 | 50.77 | TL |
| 6 | 3 | 100.00 | 4.38 | TL | 100.00 | 23.26 | TL |
| 6 | 4 | 100.00 | 2.19 | TL | 100.00 | 16.28 | TL |
| 8 | 2 | 100.00 | -1.20 | TL | 100.00 | 35.37 | TL |
| 8 | 3 | 100.00 | 6.99 | TL | 100.00 | 11.28 | TL |
| 8 | 4 | 100.00 | 6.62 | TL | 100.00 | 90.30 | TL |
| 8 | 6 | 100.00 | 5.40 | TL | 100.00 | 7.67 | TL |
| 10 | 2 | 100.00 | 3.01 | TL | 100.00 | 113.64 | TL |
| 10 | 3 | 100.00 | 0.85 | TL | 100.00 | 100.35 | TL |
| 10 | 4 | 100.00 | 9.28 | TL | 100.00 | 80.02 | TL |
| 10 | 6 | 100.00 | 7.32 | TL | 100.00 | 164.07 | TL |
| 12 | 2 | 100.00 | 24.89 | TL | 100.00 | 21.92 | TL |
| 12 | 3 | 100.00 | 20.17 | TL | 100.00 | 94.93 | TL |
| 12 | 4 | 100.00 | 12.65 | TL | 100.00 | 112.75 | TL |
| 12 | 6 | 100.00 | 19.66 | TL | 100.00 | 113.47 | TL |
| Average | | 94.90 | 6.46 | 2394.64 | 100.00 | 53.38 | - |

Table B.8: Experimental Results for Data Sets with $n = 100$ and $p = 3$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 6.83 | TL | 100.00 | 37.39 | TL |
| 4 | 3 | 100.00 | 12.34 | TL | 100.00 | 21.20 | TL |
| 5 | 2 | 100.00 | 22.72 | TL | 100.00 | 42.65 | TL |
| 5 | 3 | 100.00 | 18.16 | TL | 100.00 | 52.70 | TL |
| 5 | 4 | 100.00 | 8.32 | TL | 100.00 | 117.82 | TL |
| 6 | 2 | 100.00 | 8.46 | TL | 100.00 | 10.11 | TL |
| 6 | 3 | 100.00 | 13.07 | TL | 100.00 | 132.28 | TL |
| 6 | 4 | 100.00 | 11.37 | TL | 100.00 | 94.87 | TL |
| 8 | 2 | 100.00 | 36.35 | TL | 100.00 | 100.37 | TL |
| 8 | 3 | 100.00 | 23.41 | TL | 100.00 | 45.00 | TL |
| 8 | 4 | 100.00 | 15.50 | TL | 100.00 | 163.60 | TL |
| 8 | 6 | 100.00 | 7.64 | TL | 100.00 | 269.24 | TL |
| 10 | 2 | 100.00 | 11.92 | TL | 100.00 | 51.15 | TL |
| 10 | 3 | 100.00 | 48.55 | TL | 100.00 | 131.43 | TL |
| 10 | 4 | 100.00 | 71.01 | TL | 100.00 | 168.96 | TL |
| 10 | 6 | 100.00 | 16.22 | TL | 100.00 | 202.05 | TL |
| 12 | 2 | 100.00 | 9.84 | TL | 100.00 | 31.30 | TL |
| 12 | 3 | 100.00 | 37.00 | TL | 100.00 | 89.14 | TL |
| 12 | 4 | 100.00 | 28.51 | TL | 100.00 | 118.19 | TL |
| 12 | 6 | 100.00 | 59.01 | TL | 100.00 | 173.38 | TL |
| Average | | 100.00 | 23.31 | - | 100.00 | 102.64 | - |

Table B.9: Experimental Results for Data Sets with $n = 100$ and $p = 4$

| | | $F1_{MISOCP}$ | | | $F2_{MISOCP}$ | | |
|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_R$ | $\%Gap_B$ | Time (sec) | $\%Gap_R$ | $\%Gap_B$ | Time (sec) |
| 4 | 2 | 100.00 | 22.47 | TL | 100.00 | 104.41 | TL |
| 4 | 3 | 100.00 | 13.69 | TL | 100.00 | 26.87 | TL |
| 5 | 2 | 100.00 | 44.28 | TL | 100.00 | 75.96 | TL |
| 5 | 3 | 100.00 | 21.79 | TL | 100.00 | 166.55 | TL |
| 5 | 4 | 100.00 | 9.50 | TL | 100.00 | 262.39 | TL |
| 6 | 2 | 100.00 | 43.83 | TL | 100.00 | 76.94 | TL |
| 6 | 3 | 100.00 | 26.41 | TL | 100.00 | 173.17 | TL |
| 6 | 4 | 100.00 | 74.36 | TL | 100.00 | 319.20 | TL |
| 8 | 2 | 100.00 | 2.98 | TL | 100.00 | 37.44 | TL |
| 8 | 3 | 100.00 | 34.03 | TL | 100.00 | 169.68 | TL |
| 8 | 4 | 100.00 | 39.80 | TL | 100.00 | 308.91 | TL |
| 8 | 6 | 100.00 | 13.52 | TL | 100.00 | 90.72 | TL |
| 10 | 2 | 100.00 | -1.68 | TL | 100.00 | 25.87 | TL |
| 10 | 3 | 100.00 | 41.69 | TL | 100.00 | 150.55 | TL |
| 10 | 4 | 100.00 | 78.29 | TL | 100.00 | 148.13 | TL |
| 10 | 6 | 100.00 | 22.23 | TL | 100.00 | 281.16 | TL |
| 12 | 2 | 100.00 | 22.41 | TL | 100.00 | 54.24 | TL |
| 12 | 3 | 100.00 | 34.95 | TL | 100.00 | 63.14 | TL |
| 12 | 4 | 100.00 | 39.20 | TL | 100.00 | 182.25 | TL |
| 12 | 6 | 100.00 | 64.39 | TL | 100.00 | 239.54 | TL |
| Average | | 100.00 | 32.41 | - | 100.00 | 147.86 | - |

# Appendix C

## EXPERIMENTAL RESULTS OF MATHEURISTIC METHOD

This appendix provides the experimental result of the matheuristic method for datasets with 50,80,100 data points and its comparison with $F1\_MISOCP$. TL shows that the 24-hour time limit is exceeded for the problem instance. The performance measures used in the evaluations are as follows:

- $\%Gap_B$: Percent Gap from the Base Objective
- Time: Computation Time (seconds)
- $\%Gap_{Best}$: Best Solution Gap
- $\%Gap_{Avg}$: Average Gap
- $\%Gap_{Worst}$: Worst Solution Gap
- $\#Best$: Number of Best Solutions
- $\#Worst$: Number of Worst Solutions
- $\#IBase$: Number of Base Solutions
- Total Time: Total Computation Time of Initializations

Table C.1: Experimental Results for Data Sets with $n = 50$ and $p = 2$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -2.60* | 210.66 | 0.00* | -1.59 | 49.12 | 98 | 1 | 98 | 530.00 |
| 4 | 3 | 0.00* | 81.16 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 976.14 |
| 5 | 2 | 0.00* | 475.13 | 0.00* | 1.50 | 53.86 | 97 | 2 | 97 | 595.76 |
| 5 | 3 | 0.00* | 328.31 | 0.00* | 10.30 | 146.17 | 82 | 3 | 82 | 808.38 |
| 5 | 4 | 0.00* | 135.75 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 981.56 |
| 6 | 2 | -1.49* | 1570.86 | 0.00* | 2.02 | 60.64 | 94 | 1 | 94 | 644.74 |
| 6 | 3 | 0.00* | 1467.08 | 0.00* | 4.08 | 143.62 | 81 | 1 | 81 | 653.97 |
| 6 | 4 | 0.00* | 688.67 | 0.00* | 0.46 | 1.90 | 76 | 24 | 76 | 1282.67 |
| 8 | 2 | -1.80 | TL | -1.80 | 4.13 | 46.70 | 20 | 1 | 63 | 824.59 |
| 8 | 3 | 2.62 | TL | 0.00 | 5.92 | 95.23 | 89 | 2 | 89 | 899.62 |
| 8 | 4 | 0.58 | TL | 0.00 | 1.64 | 141.29 | 89 | 1 | 89 | 898.06 |
| 8 | 6 | 0.00* | 1662.14 | 0.00* | 1.36 | 136.36 | 99 | 1 | 99 | 2047.01 |
| 10 | 2 | 2.61 | TL | 0.00 | 4.28 | 113.02 | 25 | 1 | 25 | 1249.78 |
| 10 | 3 | 2.28 | TL | 0.00 | 2.16 | 80.24 | 97 | 1 | 97 | 1370.20 |
| 10 | 4 | 6.27 | TL | 0.00 | 2.96 | 75.85 | 75 | 1 | 75 | 1280.58 |
| 10 | 6 | 7.28 | TL | 0.00 | 0.54 | 5.19 | 25 | 4 | 25 | 2444.07 |
| 12 | 2 | 8.25 | TL | 0.00 | 7.62 | 98.71 | 60 | 1 | 60 | 1233.93 |
| 12 | 3 | 11.45 | TL | 0.00 | 7.72 | 128.61 | 85 | 2 | 85 | 1597.53 |
| 12 | 4 | 24.32 | TL | 0.00 | 10.62 | 97.27 | 51 | 1 | 51 | 1817.75 |
| 12 | 6 | 14.60 | TL | 0.00 | 2.93 | 126.34 | 87 | 1 | 87 | 4254.86 |
| **Average** | | 3.72 | 735.53 | -0.09 | 3.43 | 80.01 | 76.50 | 12.45 | 78.65 | 1319.56 |

Table C.2: Experimental Results for Data Sets with $n = 50$ and $p = 3$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 4.92 | TL | -2.35 | 3.92 | 36.07 | 32 | 1 | 46 | 644.27 |
| 4 | 3 | 4.56 | TL | 0.00 | 48.72 | 160.74 | 50 | 1 | 50 | 898.29 |
| 5 | 2 | 0.70 | TL | -0.46 | 9.50 | 58.43 | 61 | 1 | 61 | 776.31 |
| 5 | 3 | 4.05 | TL | -0.93 | 28.20 | 157.04 | 70 | 2 | 70 | 741.70 |
| 5 | 4 | 7.62 | TL | 0.00 | 31.91 | 121.06 | 59 | 1 | 59 | 1407.98 |
| 6 | 2 | 16.95 | TL | 0.00 | 11.66 | 64.84 | 16 | 1 | 16 | 921.23 |
| 6 | 3 | 8.95 | TL | -5.15 | 5.27 | 81.31 | 61 | 1 | 81 | 886.28 |
| 6 | 4 | 4.14 | TL | 0.00 | 45.87 | 186.43 | 70 | 1 | 70 | 1296.32 |
| 8 | 2 | 6.41 | TL | -4.72 | 0.13 | 22.56 | 37 | 1 | 72 | 1473.05 |
| 8 | 3 | 10.60 | TL | 0.00 | 5.72 | 66.31 | 33 | 1 | 33 | 1700.66 |
| 8 | 4 | 18.91 | TL | 0.00 | 18.64 | 138.59 | 60 | 1 | 60 | 2267.17 |
| 8 | 6 | 12.97 | TL | -0.48 | 46.79 | 182.46 | 16 | 1 | 68 | 2689.75 |
| 10 | 2 | 6.96 | TL | -9.08 | -4.11 | 19.68 | 11 | 1 | 87 | 3086.65 |
| 10 | 3 | 9.37 | TL | -1.86 | 7.64 | 30.69 | 15 | 1 | 15 | 5933.65 |
| 10 | 4 | 11.03 | TL | 0.00 | 5.58 | 53.32 | 68 | 1 | 68 | 4690.36 |
| 10 | 6 | 10.55 | TL | -0.67 | 36.54 | 155.94 | 1 | 1 | 38 | 12671.17 |
| 12 | 2 | 9.95 | TL | -8.40 | -1.08 | 20.09 | 16 | 1 | 63 | 7065.49 |
| 12 | 3 | 24.85 | TL | 0.00 | 8.83 | 47.42 | 33 | 1 | 33 | 31305.55 |
| 12 | 4 | 21.61 | TL | -3.91 | 4.24 | 54.38 | 16 | 1 | 68 | 30657.01 |
| 12 | 6 | 3.25 | TL | -1.28 | 14.14 | 90.58 | 7 | 1 | 54 | 50367.02 |
| **Average** | | 9.92 | - | -1.97 | 16.41 | 87.40 | 36.60 | 1.05 | 55.60 | 8074.00 |

Table C.3: Experimental Results for Data Sets with $n = 50$ and $p = 4$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 11.22 | TL | 0.00 | 10.38 | 98.85 | 56 | 1 | 56 | 762.08 |
| 4 | 3 | 7.88 | TL | 0.00 | 37.34 | 104.49 | 32 | 1 | 32 | 768.36 |
| 5 | 2 | 38.03 | TL | 0.00 | 14.10 | 89.12 | 28 | 1 | 28 | 973.93 |
| 5 | 3 | 38.19 | TL | 0.00 | 33.78 | 150.93 | 27 | 2 | 27 | 1029.39 |
| 5 | 4 | 18.69 | TL | 0.00 | 41.39 | 199.00 | 47 | 2 | 47 | 1202.25 |
| 6 | 2 | 15.59 | TL | -6.37 | 1.14 | 22.24 | 18 | 1 | 50 | 1906.32 |
| 6 | 3 | 27.63 | TL | -4.57 | 6.73 | 63.64 | 25 | 1 | 70 | 2020.01 |
| 6 | 4 | 8.85 | TL | -1.15 | 34.72 | 106.81 | 45 | 1 | 54 | 1446.35 |
| 8 | 2 | -13.58 | TL | -25.07 | -17.88 | 2.33 | 3 | 1 | 98 | 7451.57 |
| 8 | 3 | 24.77 | TL | -1.27 | 14.97 | 61.52 | 3 | 1 | 14 | 11551.66 |
| 8 | 4 | 35.15 | TL | -1.78 | 29.27 | 118.34 | 9 | 1 | 17 | 42526.62 |
| 8 | 6 | 13.66 | TL | -2.58 | 38.29 | 213.33 | 50 | 1 | 53 | 3932.10 |
| 10 | 2 | 2.69 | TL | -18.61 | -10.06 | 12.57 | 14 | 1 | 91 | 38741.72 |
| 10 | 3 | 22.02 | TL | -0.82 | 10.99 | 29.45 | 28.57 | 14.29 | 28.57 | TL |
| 10 | 4 | 27.64 | TL | 0.00 | 23.17 | 52.96 | 13.33 | 6.67 | 13.33 | TL |
| 10 | 6 | 34.91 | TL | -1.36 | 38.13 | 124.55 | 33.33 | 5.56 | 44.44 | TL |
| 12 | 2 | -2.74 | TL | -26.25 | -19.86 | 1.14 | 40 | 6.67 | 93.33 | TL |
| 12 | 3 | 47.49 | TL | 0.00 | 16.12 | 33.74 | 20 | 20 | 20 | TL |
| 12 | 4 | 14.37 | TL | -9.17 | -2.95 | 4.95 | 20 | 20 | 80 | TL |
| 12 | 6 | 71.66 | TL | -1.81 | 19.30 | 93.12 | 28.57 | 14.29 | 42.86 | TL |
| **Average** | | 22.21 | - | -5.04 | 15.95 | 79.15 | 27.04 | 5.12 | 47.98 | 8793.26 |

Table C.4: Experimental Results for Data Sets with $n = 80$ and $p = 2$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -0.06* | 2798.70 | 0.00* | 0.36 | 4.54 | 91 | 9 | 91 | 1410.29 |
| 4 | 3 | 0.00* | 795.52 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 2094.17 |
| 5 | 2 | 1.13 | TL | 0.00 | 3.10 | 101.47 | 89 | 1 | 89 | 4214.06 |
| 5 | 3 | 0.13 | TL | 0.00 | 5.30 | 42.25 | 81 | 10 | 81 | 1995.21 |
| 5 | 4 | 0.00* | 1070.89 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 2507.57 |
| 6 | 2 | -3.52 | TL | -3.52 | -1.33 | 57.21 | 87 | 1 | 87 | 2230.52 |
| 6 | 3 | 2.67 | TL | 0.00 | 5.57 | 128.50 | 83 | 2 | 83 | 1693.99 |
| 6 | 4 | 2.99 | TL | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 2773.77 |
| 8 | 2 | 0.76 | TL | -2.16 | -0.08 | 43.01 | 86 | 1 | 86 | 3221.80 |
| 8 | 3 | 9.31 | TL | 0.00 | 3.94 | 92.80 | 95 | 2 | 95 | 2484.92 |
| 8 | 4 | 0.45 | TL | -1.37 | 1.20 | 16.41 | 83 | 11 | 83 | 2677.23 |
| 8 | 6 | 1.63 | TL | 0.00 | 1.10 | 2.38 | 40 | 17 | 40 | 4057.72 |
| 10 | 2 | 4.49 | TL | 0.00 | 7.29 | 75.23 | 81 | 1 | 81 | 3183.46 |
| 10 | 3 | 4.85 | TL | 0.00 | 2.11 | 18.18 | 75 | 1 | 75 | 3143.44 |
| 10 | 4 | 2.65 | TL | 0.00 | 1.18 | 3.58 | 67 | 33 | 67 | 3524.75 |
| 10 | 6 | 6.41 | TL | 0.00 | 1.80 | 6.48 | 38 | 7 | 38 | 6185.46 |
| 12 | 2 | 2.32 | TL | -6.11 | -2.38 | 42.09 | 82 | 1 | 82 | 4506.86 |
| 12 | 3 | 11.11 | TL | 0.00 | 3.55 | 71.29 | 57 | 1 | 57 | 4293.24 |
| 12 | 4 | 10.96 | TL | 0.00 | 4.40 | 71.61 | 81 | 1 | 81 | 5402.25 |
| 12 | 6 | 14.92 | TL | 0.00 | 5.50 | 87.87 | 56 | 1 | 56 | 7732.20 |
| **Average** | | 3.66 | 1555.04 | -0.66 | 2.13 | 43.24 | 78.60 | 20.00 | 78.60 | 3466.65 |

Table C.5: Experimental Results for Data Sets with $n = 80$ and $p = 3$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -1.64 | TL | -3.35 | 4.22 | 32.99 | 42 | 1 | 51 | 1597.55 |
| 4 | 3 | 18.55 | TL | 0.00 | 40.01 | 167.67 | 75 | 1 | 75 | 1785.24 |
| 5 | 2 | 17.73 | TL | -2.54 | 3.37 | 36.67 | 36 | 1 | 41 | 2070.11 |
| 5 | 3 | 16.92 | TL | 0.00 | 24.57 | 167.63 | 81 | 2 | 81 | 1780.79 |
| 5 | 4 | 6.20 | TL | -0.67 | 51.23 | 157.46 | 16 | 3 | 39 | 2295.31 |
| 6 | 2 | 8.69 | TL | -1.93 | 6.88 | 42.23 | 57 | 1 | 57 | 2719.82 |
| 6 | 3 | 17.29 | TL | -2.47 | 2.15 | 84.57 | 55 | 1 | 70 | 2657.54 |
| 6 | 4 | 6.91 | TL | 0.00 | 36.97 | 166.22 | 74 | 1 | 74 | 1986.08 |
| 8 | 2 | 16.45 | TL | -14.21 | -6.62 | 11.64 | 15 | 1 | 86 | 3776.64 |
| 8 | 3 | 11.72 | TL | 0.00 | 5.16 | 27.97 | 33 | 1 | 33 | 4966.71 |
| 8 | 4 | 20.29 | TL | 0.00 | 20.52 | 157.47 | 18 | 1 | 18 | 6768.68 |
| 8 | 6 | 16.30 | TL | 0.00 | 49.79 | 179.76 | 54 | 1 | 54 | 16156.22 |
| 10 | 2 | 16.05 | TL | -10.43 | -5.26 | 8.15 | 14 | 1 | 91 | 7920.38 |
| 10 | 3 | 26.68 | TL | -1.19 | 5.52 | 42.04 | 57 | 1 | 60 | 18825.84 |
| 10 | 4 | 36.48 | TL | 0.00 | 11.83 | 64.88 | 52 | 1 | 52 | 17464.07 |
| 10 | 6 | 38.14 | TL | 0.00 | 22.75 | 151.47 | 33.33 | 1.08 | 33.33 | TL |
| 12 | 2 | -7.15 | TL | -24.70 | -19.08 | -7.32 | 8 | 1 | 100 | 21375.43 |
| 12 | 3 | 29.93 | TL | -2.57 | 1.24 | 61.09 | 58 | 1 | 75 | 69360.34 |
| 12 | 4 | 46.23 | TL | -0.24 | 10.26 | 38.41 | 25 | 25 | 25 | TL |
| 12 | 6 | 43.10 | TL | 0.00 | 14.16 | 76.62 | 15 | 1.67 | 15 | TL |
| **Average** | | 19.24 | - | -3.22 | 13.98 | 83.38 | 40.92 | 2.39 | 56.52 | 10794.51 |

Table C.6: Experimental Results for Data Sets with $n = 80$ and $p = 4$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 15.90 | TL | 0.00 | 16.58 | 137.03 | 45 | 1 | 45 | 1981.24 |
| 4 | 3 | 28.77 | TL | 0.00 | 46.79 | 115.93 | 35 | 1 | 35 | 1614.88 |
| 5 | 2 | 4.64 | TL | -2.90 | 10.37 | 41.83 | 38 | 1 | 38 | 2856.21 |
| 5 | 3 | 49.38 | TL | 0.00 | 16.75 | 141.84 | 82 | 1 | 82 | 2570.39 |
| 5 | 4 | 13.07 | TL | 0.00 | 65.94 | 216.98 | 19 | 1 | 19 | 2278.20 |
| 6 | 2 | 25.38 | TL | -2.33 | 11.36 | 51.11 | 15 | 1 | 32 | 5661.03 |
| 6 | 3 | 47.73 | TL | 0.00 | 0.73 | 53.89 | 90 | 1 | 90 | 5220.70 |
| 6 | 4 | 58.38 | TL | 0.00 | 47.37 | 155.83 | 9 | 1 | 9 | 3490.49 |
| 8 | 2 | 22.87 | TL | -9.38 | -0.49 | 18.91 | 16 | 1 | 52 | 24027.89 |
| 8 | 3 | 56.30 | TL | 0.00 | 13.40 | 56.50 | 7 | 1 | 7 | 32590.25 |
| 8 | 4 | 30.55 | TL | 0.00 | 11.27 | 105.29 | 1.04 | 1.04 | 1.04 | TL |
| 8 | 6 | 32.64 | TL | 0.00 | 62.13 | 235.65 | 34 | 1 | 34 | 18004.02 |
| 10 | 2 | 18.83 | TL | -12.47 | -4.52 | 5.49 | 1.92 | 1.92 | 84.62 | TL |
| 10 | 3 | 18.00 | TL | 0.46 | 9.59 | 30.76 | 22.22 | 11.11 | 0 | TL |
| 10 | 4 | 76.04 | TL | 0.00 | 8.59 | 41.49 | 25 | 6.25 | 25 | TL |
| 10 | 6 | 29.55 | TL | 0.00 | 39.89 | 125.05 | 18.75 | 6.25 | 18.75 | TL |
| 12 | 2 | -2.36 | TL | -23.86 | -18.07 | -11.87 | 10 | 10 | 100 | TL |
| 12 | 3 | 53.37 | TL | 2.16 | 6.49 | 9.24 | 25 | 25 | 0 | TL |
| 12 | 4 | 62.59 | TL | 4.29 | 31.11 | 47.47 | 25 | 25 | 0 | TL |
| 12 | 6 | 31.44 | TL | -2.89 | 11.24 | 77.65 | 9.09 | 9.09 | 72.73 | TL |
| **Average** | | 33.65 | - | -2.35 | 19.33 | 82.80 | 26.40 | 5.33 | 37.26 | 9117.75 |

Table C.7: Experimental Results for Data Sets with $n = 100$ and $p = 2$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.14 | TL | -0.52 | 2.64 | 157.70 | 98 | 2 | 98 | 2346.44 |
| 4 | 3 | 0.00* | 2394.64 | 0.00* | 3.28 | 163.97 | 98 | 2 | 98 | 3062.48 |
| 5 | 2 | 2.59 | TL | 0.00 | 2.73 | 48.51 | 63 | 2 | 63 | 2392.99 |
| 5 | 3 | 1.12 | TL | 0.00 | 3.67 | 32.79 | 77 | 8 | 77 | 3306.48 |
| 5 | 4 | 0.00 | TL | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 3647.99 |
| 6 | 2 | 3.22 | TL | 0.00 | 1.25 | 55.11 | 80 | 2 | 80 | 3141.05 |
| 6 | 3 | 4.38 | TL | 0.00 | 5.93 | 118.69 | 95 | 5 | 95 | 3252.96 |
| 6 | 4 | 2.19 | TL | 0.00 | 2.52 | 7.37 | 56 | 24 | 56 | 4242.78 |
| 8 | 2 | -1.20 | TL | -1.20 | 2.48 | 55.78 | 93 | 1 | 93 | 5000.91 |
| 8 | 3 | 6.99 | TL | 0.00 | 0.33 | 1.64 | 80 | 20 | 80 | 3898.12 |
| 8 | 4 | 6.62 | TL | 0.00 | 4.59 | 132.81 | 85 | 2 | 85 | 4025.98 |
| 8 | 6 | 5.40 | TL | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 5265.75 |
| 10 | 2 | 3.01 | TL | -3.35 | -0.23 | 62.38 | 91 | 1 | 91 | 6536.77 |
| 10 | 3 | 0.85 | TL | 0.00 | 1.72 | 73.45 | 95 | 1 | 95 | 5325.46 |
| 10 | 4 | 9.28 | TL | 0.00 | 5.02 | 77.66 | 74 | 1 | 74 | 5709.09 |
| 10 | 6 | 7.32 | TL | 0.00 | 2.10 | 158.09 | 87 | 1 | 87 | 20570.94 |
| 12 | 2 | 24.89 | TL | 0.00 | 4.79 | 83.99 | 68 | 1 | 68 | 17839.78 |
| 12 | 3 | 20.17 | TL | -0.02 | 5.35 | 55.27 | 83 | 2 | 83 | 11646.74 |
| 12 | 4 | 12.65 | TL | 0.00 | 7.28 | 90.48 | 52 | 1 | 52 | 54233.10 |
| 12 | 6 | 19.66 | TL | 0.00 | 1.48 | 9.88 | 63.29 | 1.27 | 63.29 | TL |
| **Average** | | 6.46 | 2394.64 | -0.25 | 2.85 | 69.28 | 81.91 | 13.86 | 81.91 | 8707.67 |

Table C.8: Experimental Results for Data Sets with $n = 100$ and $p = 3$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 6.83 | TL | 0.00 | 8.52 | 39.68 | 26 | 1 | 26 | 2428.52 |
| 4 | 3 | 12.34 | TL | 0.00 | 32.36 | 169.33 | 76 | 1 | 76 | 3869.32 |
| 5 | 2 | 22.72 | TL | -0.17 | 6.36 | 52.27 | 24 | 1 | 50 | 2910.87 |
| 5 | 3 | 18.16 | TL | 0.00 | 31.57 | 185.00 | 58 | 1 | 58 | 4549.14 |
| 5 | 4 | 8.32 | TL | 0.00 | 25.82 | 110.35 | 68 | 1 | 68 | 5369.06 |
| 6 | 2 | 8.46 | TL | -3.85 | 2.91 | 38.99 | 64 | 1 | 64 | 3522.81 |
| 6 | 3 | 13.07 | TL | 0.00 | 5.84 | 121.62 | 93 | 1 | 93 | 3400.41 |
| 6 | 4 | 11.37 | TL | 0.00 | 35.71 | 192.61 | 76 | 1 | 76 | 2936.83 |
| 8 | 2 | 36.35 | TL | -1.35 | 4.66 | 31.42 | 32 | 1 | 32 | 7146.38 |
| 8 | 3 | 23.41 | TL | 0.00 | 7.97 | 36.50 | 26 | 1 | 26 | 7570.05 |
| 8 | 4 | 15.50 | TL | 0.00 | 11.65 | 105.32 | 54 | 3 | 54 | 17541.69 |
| 8 | 6 | 7.64 | TL | -0.50 | 21.90 | 102.94 | 12 | 1 | 78 | 30186.60 |
| 10 | 2 | 11.92 | TL | -6.18 | -0.26 | 13.72 | 10 | 2 | 56 | 13243.00 |
| 10 | 3 | 48.55 | TL | -0.44 | 5.35 | 35.87 | 62 | 1 | 62 | 29645.93 |
| 10 | 4 | 71.01 | TL | 0.00 | 10.26 | 80.80 | 57 | 1 | 57 | 25424.00 |
| 10 | 6 | 16.22 | TL | 0.00 | 20.68 | 163.04 | 56.16 | 1.37 | 56.16 | TL |
| 12 | 2 | 9.84 | TL | -16.08 | -10.33 | 1.10 | 8 | 1 | 98 | 38372.99 |
| 12 | 3 | 37.00 | TL | -1.00 | 5.12 | 36.15 | 25.93 | 1.23 | 44.44 | TL |
| 12 | 4 | 28.51 | TL | -0.78 | 5.42 | 39.28 | 4.65 | 2.33 | 39.53 | TL |
| 12 | 6 | 59.01 | TL | 0.00 | 7.15 | 70.23 | 13.04 | 4.35 | 13.04 | TL |
| **Average** | | 23.31 | - | -1.52 | 11.93 | 81.31 | 42.29 | 1.41 | 56.36 | 12382.35 |

Table C.9: Experimental Results for Data Sets with $n = 100$ and $p = 4$

| | | $F1_{MISOCP}$ | | **Matheuristic** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_B$ | Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 22.47 | TL | 0.00 | 10.96 | 112.55 | 45 | 1 | 45 | 3050.79 |
| 4 | 3 | 13.69 | TL | 0.00 | 40.22 | 116.52 | 18 | 1 | 18 | 2761.42 |
| 5 | 2 | 44.28 | TL | -0.29 | 8.56 | 76.18 | 57 | 1 | 57 | 4179.29 |
| 5 | 3 | 21.79 | TL | 0.00 | 20.93 | 129.00 | 67 | 1 | 67 | 3863.16 |
| 5 | 4 | 9.50 | TL | 0.00 | 61.99 | 220.36 | 46 | 1 | 46 | 3827.98 |
| 6 | 2 | 43.83 | TL | -1.93 | 5.51 | 47.48 | 73 | 1 | 73 | 7638.35 |
| 6 | 3 | 26.41 | TL | -1.44 | 6.49 | 89.81 | 41 | 1 | 56 | 7288.35 |
| 6 | 4 | 74.36 | TL | 0.00 | 49.30 | 145.90 | 60 | 1 | 60 | 10757.87 |
| 8 | 2 | 2.98 | TL | -20.62 | -17.66 | 3.58 | 6 | 1 | 99 | 38102.88 |
| 8 | 3 | 34.03 | TL | 0.00 | 14.35 | 63.93 | 39 | 1 | 39 | 65181.06 |
| 8 | 4 | 39.80 | TL | 0.00 | 23.16 | 116.63 | 70.77 | 1.54 | 70.77 | TL |
| 8 | 6 | 13.52 | TL | 0.00 | 35.19 | 231.51 | 41 | 1 | 41 | 82964.06 |
| 10 | 2 | -1.68 | TL | -24.00 | -18.40 | -2.60 | 11.11 | 1.85 | 100 | TL |
| 10 | 3 | 41.69 | TL | -0.03 | 6.78 | 15.15 | 12.5 | 12.5 | 12.5 | TL |
| 10 | 4 | 78.29 | TL | -2.36 | 13.81 | 41.79 | 14.29 | 7.14 | 42.86 | TL |
| 10 | 6 | 22.23 | TL | 0.00 | 20.56 | 82.95 | 62.5 | 12.5 | 62.5 | TL |
| 12 | 2 | 22.41 | TL | -24.43 | -19.57 | -9.83 | 10 | 10 | 100 | TL |
| 12 | 3 | 34.95 | TL | 7.72 | 9.49 | 11.26 | 50 | 50 | 0 | TL |
| 12 | 4 | 39.20 | TL | 6.20 | 30.58 | 41.81 | 20 | 20 | 0 | TL |
| 12 | 6 | 64.39 | TL | 91.79 | 91.79 | 91.79 | 100 | 100 | 0 | TL |
| **Average** | | 32.41 | - | 1.53 | 19.70 | 81.29 | 42.21 | 11.33 | 49.48 | 20874.11 |

## Appendix D

## EXPERIMENTAL RESULTS OF THE ITERATIVE HEURISTIC METHOD

This appendix provides the experimental result of the iterative heuristic method for datasets with 50,80,100,200,500, and 1000 data points and compares it with the matheuristic method. The performance measures used in the tables are:

- $\%Gap_{Best}$: Best Solution Gap
- $\%Gap_{Avg}$: Average Gap
- $\%Gap_{Worst}$: Worst Solution Gap
- $\#Best$: Number of Best Solutions
- $\#Worst$: Number of Worst Solutions
- $\#IBase$: Number of Base Solutions
- Total Time: Total Computation Time for 100 Initializations

Table D.1: Experimental Results for Datasets with $n = 50$ and $p = 2$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00* | -1.59 | 49.12 | 98 | 1 | 98 | 530.00 | 0.00* | -1.77 | 47.06 | 95 | 1 | 95 | 0.93 |
| 4 | 3 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 976.14 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 0.75 |
| 5 | 2 | 0.00* | 1.50 | 53.86 | 97 | 2 | 97 | 595.76 | 0.00* | 1.50 | 53.86 | 97 | 2 | 97 | 0.71 |
| 5 | 3 | 0.00* | 10.30 | 146.17 | 82 | 3 | 82 | 808.38 | 0.00* | 10.74 | 174.24 | 81 | 1 | 81 | 0.73 |
| 5 | 4 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 981.56 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 0.82 |
| 6 | 2 | 0.00* | 2.02 | 60.64 | 94 | 1 | 94 | 644.74 | 0.00* | 7.66 | 67.17 | 69 | 1 | 69 | 0.81 |
| 6 | 3 | 0.00* | 4.08 | 143.62 | 81 | 1 | 81 | 653.97 | 0.00* | 4.61 | 148.04 | 75 | 1 | 75 | 0.86 |
| 6 | 4 | 0.00* | 0.46 | 1.90 | 76 | 24 | 76 | 1282.67 | 0.00* | 0.53 | 1.90 | 72 | 28 | 72 | 0.85 |
| 8 | 2 | -1.80 | 4.13 | 46.70 | 20 | 1 | 63 | 824.59 | -1.80 | 3.31 | 56.58 | 13 | 1 | 67 | 1.52 |
| 8 | 3 | 0.00 | 5.92 | 95.23 | 89 | 2 | 89 | 899.62 | 0.00 | 4.65 | 119.11 | 92 | 1 | 92 | 1.06 |
| 8 | 4 | 0.00 | 1.64 | 141.29 | 89 | 1 | 89 | 898.06 | 0.00 | 0.23 | 2.26 | 90 | 10 | 90 | 1.08 |
| 8 | 6 | 0.00* | 1.36 | 136.36 | 99 | 1 | 99 | 2047.01 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 1.89 |
| 10 | 2 | 0.00 | 4.28 | 113.02 | 25 | 1 | 25 | 1249.78 | 0.00 | 5.93 | 110.56 | 32 | 1 | 32 | 1.18 |
| 10 | 3 | 0.00 | 2.16 | 80.24 | 97 | 1 | 97 | 1370.20 | 0.00 | 11.20 | 96.81 | 23 | 1 | 23 | 1.60 |
| 10 | 4 | 0.00 | 2.96 | 75.85 | 75 | 1 | 75 | 1280.58 | 0.00 | 5.52 | 84.85 | 61 | 3 | 61 | 1.34 |
| 10 | 6 | 0.00 | 0.54 | 5.19 | 25 | 4 | 25 | 2444.07 | 0.00 | 0.58 | 5.19 | 20 | 5 | 20 | 1.74 |
| 12 | 2 | 0.00 | 7.62 | 98.71 | 60 | 1 | 60 | 1233.93 | 0.00 | 8.41 | 136.22 | 57 | 1 | 57 | 1.16 |
| 12 | 3 | 0.00 | 7.72 | 128.61 | 85 | 2 | 85 | 1597.53 | 0.00 | 9.15 | 147.09 | 81 | 1 | 81 | 1.33 |
| 12 | 4 | 0.00 | 10.62 | 97.27 | 51 | 1 | 51 | 1817.75 | 0.00 | 12.46 | 102.91 | 53 | 1 | 53 | 1.11 |
| 12 | 6 | 0.00 | 2.93 | 126.34 | 87 | 1 | 87 | 4254.86 | 0.00 | 6.45 | 139.47 | 80 | 1 | 80 | 2.15 |
| **Average** | | -0.09 | 3.43 | 80.01 | 76.50 | 12.45 | 78.65 | 1319.56 | -0.09 | 4.56 | 74.67 | 69.55 | 18.00 | 72.25 | 1.18 |

Table D.2: Experimental Results for Datasets with $n = 50$ and $p = 3$

| | | **Matheuristic Method** | | | | | | | **Heuristic Method** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -2.35 | 3.92 | 36.07 | 32 | 1 | 46 | 644.27 | -2.35 | 12.11 | 68.08 | 12 | 2 | 20 | 0.92 |
| 4 | 3 | 0.00 | 48.72 | 160.74 | 50 | 1 | 50 | 898.29 | 0.00 | 58.14 | 163.86 | 49 | 2 | 49 | 0.81 |
| 5 | 2 | -0.46 | 9.50 | 58.43 | 61 | 1 | 61 | 776.31 | -0.46 | 8.94 | 72.78 | 65 | 1 | 65 | 1.11 |
| 5 | 3 | -0.93 | 28.20 | 157.04 | 70 | 2 | 70 | 741.70 | -0.93 | 41.81 | 160.66 | 62 | 1 | 62 | 0.98 |
| 5 | 4 | 0.00 | 31.91 | 121.06 | 59 | 1 | 59 | 1407.98 | 0.00 | 33.36 | 122.62 | 57 | 1 | 57 | 0.96 |
| 6 | 2 | 0.00 | 11.66 | 64.84 | 16 | 1 | 16 | 921.23 | 0.00 | 11.67 | 67.50 | 5 | 1 | 5 | 1.39 |
| 6 | 3 | -5.15 | 5.27 | 81.31 | 61 | 1 | 81 | 886.28 | -5.15 | 15.79 | 96.98 | 60 | 1 | 77 | 1.11 |
| 6 | 4 | 0.00 | 45.87 | 186.43 | 70 | 1 | 70 | 1296.32 | 0.00 | 49.77 | 187.05 | 68 | 1 | 68 | 1.04 |
| 8 | 2 | -4.72 | 0.13 | 22.56 | 37 | 1 | 72 | 1473.05 | -4.49 | 4.29 | 46.60 | 17 | 1 | 54 | 2.32 |
| 8 | 3 | 0.00 | 5.72 | 66.31 | 33 | 1 | 33 | 1700.66 | 0.00 | 25.66 | 83.35 | 16 | 1 | 16 | 1.31 |
| 8 | 4 | 0.00 | 18.64 | 138.59 | 60 | 1 | 60 | 2267.17 | 0.00 | 32.34 | 153.94 | 42 | 1 | 42 | 1.54 |
| 8 | 6 | -0.48 | 46.79 | 182.46 | 16 | 1 | 68 | 2689.75 | -0.48 | 57.74 | 182.46 | 11 | 1 | 61 | 1.77 |
| 10 | 2 | -9.08 | -4.11 | 19.68 | 11 | 1 | 87 | 3086.65 | -5.56 | 2.67 | 26.13 | 3 | 1 | 48 | 2.28 |
| 10 | 3 | -1.86 | 7.64 | 30.69 | 15 | 1 | 15 | 5933.65 | -1.86 | 12.27 | 55.21 | 6 | 1 | 6 | 1.58 |
| 10 | 4 | 0.00 | 5.58 | 53.32 | 68 | 1 | 68 | 4690.36 | 0.00 | 28.75 | 137.21 | 43 | 1 | 43 | 1.49 |
| 10 | 6 | -0.67 | 36.54 | 155.94 | 1 | 1 | 38 | 12671.17 | -0.67 | 43.68 | 153.41 | 1 | 1 | 35 | 1.86 |
| 12 | 2 | -8.40 | -1.08 | 20.09 | 16 | 1 | 63 | 7065.49 | -6.34 | 16.46 | 46.82 | 7 | 1 | 9 | 2.53 |
| 12 | 3 | 0.00 | 8.83 | 47.42 | 33 | 1 | 33 | 31305.55 | 0.00 | 15.54 | 79.28 | 41 | 1 | 41 | 1.58 |
| 12 | 4 | -3.91 | 4.24 | 54.38 | 16 | 1 | 68 | 30657.01 | -3.91 | 16.62 | 56.90 | 4 | 1 | 32 | 1.77 |
| 12 | 6 | -1.28 | 14.14 | 90.58 | 7 | 1 | 54 | 50367.02 | -1.28 | 21.98 | 126.68 | 11 | 1 | 43 | 2.33 |
| **Average** | | -1.97 | 16.41 | 87.40 | 36.60 | 1.05 | 55.60 | 8074.00 | -1.67 | 25.48 | 104.38 | 29.00 | 1.10 | 41.65 | 1.53 |

Table D.3: Experimental Results for Datasets with $n = 50$ and $p = 4$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 10.38 | 98.85 | 56 | 1 | 56 | 762.08 | 0.00 | 22.01 | 113.88 | 60 | 1 | 60 | 1.30 |
| 4 | 3 | 0.00 | 37.34 | 104.49 | 32 | 1 | 32 | 768.36 | 0.00 | 37.11 | 222.13 | 31 | 2 | 31 | 1.11 |
| 5 | 2 | 0.00 | 14.10 | 89.12 | 28 | 1 | 28 | 973.93 | 0.00 | 14.27 | 57.79 | 4 | 1 | 4 | 3.13 |
| 5 | 3 | 0.00 | 33.78 | 150.93 | 27 | 2 | 27 | 1029.39 | 0.00 | 48.21 | 190.31 | 28 | 1 | 28 | 1.18 |
| 5 | 4 | 0.00 | 41.39 | 199.00 | 47 | 2 | 47 | 1202.25 | 0.00 | 50.65 | 200.72 | 44 | 1 | 44 | 1.18 |
| 6 | 2 | -6.37 | 1.14 | 22.24 | 18 | 1 | 50 | 1906.32 | -6.37 | 4.68 | 48.99 | 17 | 1 | 45 | 2.69 |
| 6 | 3 | -4.57 | 6.73 | 63.64 | 25 | 1 | 70 | 2020.01 | -4.57 | 12.98 | 122.45 | 28 | 1 | 67 | 1.43 |
| 6 | 4 | -1.15 | 34.72 | 106.81 | 45 | 1 | 54 | 1446.35 | -1.15 | 37.35 | 205.38 | 47 | 1 | 56 | 2.34 |
| 8 | 2 | -25.07 | -17.88 | 2.33 | 3 | 1 | 98 | 7451.57 | -23.66 | -13.53 | 13.93 | 8 | 1 | 94 | 2.89 |
| 8 | 3 | -1.27 | 14.97 | 61.52 | 3 | 1 | 14 | 11551.66 | -1.27 | 23.62 | 77.00 | 1 | 1 | 7 | 1.92 |
| 8 | 4 | -1.78 | 29.27 | 118.34 | 9 | 1 | 17 | 42526.62 | -1.78 | 36.64 | 177.98 | 7 | 1 | 19 | 1.87 |
| 8 | 6 | -2.58 | 38.29 | 213.33 | 50 | 1 | 53 | 3932.10 | -2.58 | 33.80 | 217.07 | 58 | 1 | 62 | 3.28 |
| 10 | 2 | -18.61 | -10.06 | 12.57 | 14 | 1 | 91 | 38741.72 | -18.49 | -7.04 | 17.57 | 10 | 1 | 84 | 4.07 |
| 10 | 3 | -0.82 | 10.99 | 29.45 | 28.57 | 14.29 | 28.57 | TL | 3.24 | 17.97 | 55.56 | 13 | 1 | 0 | 2.85 |
| 10 | 4 | 0.00 | 23.17 | 52.96 | 13.33 | 6.67 | 13.33 | TL | 0.00 | 38.49 | 140.43 | 9 | 1 | 9 | 2.53 |
| 10 | 6 | -1.36 | 38.13 | 124.55 | 33.33 | 5.56 | 44.44 | TL | -1.36 | 45.56 | 177.41 | 9 | 1 | 30 | 2.56 |
| 12 | 2 | -26.25 | -19.86 | 1.14 | 40 | 7 | 93 | TL | -22.43 | -3.71 | 38.77 | 1 | 1 | 67 | 3.72 |
| 12 | 3 | 0.00 | 16.12 | 33.74 | 20 | 20 | 20 | TL | 0.00 | 29.91 | 107.76 | 1 | 1 | 1 | 4.29 |
| 12 | 4 | -9.17 | -2.95 | 4.95 | 20 | 20 | 80 | TL | -9.17 | 22.59 | 68.72 | 2 | 1 | 27 | 2.47 |
| 12 | 6 | -1.81 | 19.30 | 93.12 | 28.57 | 14.29 | 42.86 | TL | -1.81 | 36.67 | 153.17 | 7 | 1 | 27 | 2.89 |
| **Average** | | -5.04 | 15.95 | 79.15 | 27.04 | 5.12 | 47.98 | 8793.26 | -4.57 | 24.41 | 120.35 | 19.25 | 1.05 | 38.10 | 2.48 |

Table D.4: Experimental Results for Datasets with $n = 80$ and $p = 2$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00* | 0.36 | 4.54 | 91 | 9 | 91 | 1410.29 | 0.00* | 0.31 | 4.54 | 92 | 8 | 92 | 1.28 |
| 4 | 3 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 2094.17 | 0.00* | 2.42 | 241.60 | 99 | 1 | 99 | 1.34 |
| 5 | 2 | 0.00 | 3.10 | 101.47 | 89 | 1 | 89 | 4214.06 | 0.00 | 1.30 | 14.42 | 91 | 9 | 91 | 1.47 |
| 5 | 3 | 0.00 | 5.30 | 42.25 | 81 | 10 | 81 | 1995.21 | 0.00 | 5.06 | 42.25 | 83 | 10 | 83 | 2.72 |
| 5 | 4 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 2507.57 | 0.00* | 0.00 | 0.00 | 100 | 100 | 100 | 1.59 |
| 6 | 2 | -3.52 | -1.33 | 57.21 | 87 | 1 | 87 | 2230.52 | -3.52 | -2.89 | 12.04 | 96 | 4 | 96 | 2.09 |
| 6 | 3 | 0.00 | 5.57 | 128.50 | 83 | 2 | 83 | 1693.99 | 0.00 | 4.94 | 137.05 | 87 | 1 | 87 | 1.53 |
| 6 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 2773.77 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 2.17 |
| 8 | 2 | -2.16 | -0.08 | 43.01 | 86 | 1 | 86 | 3221.80 | -2.13 | 2.25 | 55.20 | 1 | 1 | 72 | 3.15 |
| 8 | 3 | 0.00 | 3.94 | 92.80 | 95 | 2 | 95 | 2484.92 | 0.00 | 3.02 | 93.23 | 96 | 1 | 96 | 2.10 |
| 8 | 4 | -1.37 | 1.20 | 16.41 | 83 | 11 | 83 | 2677.23 | -1.37 | 0.86 | 12.11 | 82 | 15 | 82 | 2.32 |
| 8 | 6 | 0.00 | 1.10 | 2.38 | 40 | 17 | 40 | 4057.72 | 0.00 | 0.96 | 2.38 | 48 | 15 | 48 | 2.62 |
| 10 | 2 | 0.00 | 7.29 | 75.23 | 81 | 1 | 81 | 3183.46 | 0.00 | 5.79 | 79.64 | 85 | 1 | 85 | 2.06 |
| 10 | 3 | 0.00 | 2.11 | 18.18 | 75 | 1 | 75 | 3143.44 | 0.00 | 4.95 | 15.14 | 47 | 15 | 47 | 2.83 |
| 10 | 4 | 0.00 | 1.18 | 3.58 | 67 | 33 | 67 | 3524.75 | 0.00 | 2.72 | 96.34 | 64 | 1 | 64 | 2.53 |
| 10 | 6 | 0.00 | 1.80 | 6.48 | 38 | 7 | 38 | 6185.46 | 0.00 | 1.86 | 6.48 | 33 | 6 | 33 | 2.91 |
| 12 | 2 | -6.11 | -2.38 | 42.09 | 82 | 1 | 82 | 4506.86 | -6.11 | -2.82 | 6.48 | 61 | 5 | 61 | 2.74 |
| 12 | 3 | 0.00 | 3.55 | 71.29 | 57 | 1 | 57 | 4293.24 | 0.00 | 4.54 | 85.52 | 52 | 1 | 52 | 2.11 |
| 12 | 4 | 0.00 | 4.40 | 71.61 | 81 | 1 | 81 | 5402.25 | 0.00 | 4.18 | 81.61 | 78 | 1 | 78 | 2.80 |
| 12 | 6 | 0.00 | 5.50 | 87.87 | 56 | 1 | 56 | 7732.20 | 0.00 | 6.42 | 90.49 | 55 | 1 | 55 | 3.74 |
| **Average** | | -0.66 | 2.13 | 43.24 | 78.60 | 20.00 | 78.60 | 3466.65 | -0.66 | 2.29 | 53.83 | 72.50 | 14.80 | 76.05 | 2.30 |

Table D.5: Experimental Results for Datasets with $n = 80$ and $p = 3$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -3.35 | 4.22 | 32.99 | 42 | 1 | 51 | 1597.55 | -3.35 | 14.61 | 76.63 | 22 | 1 | 25 | 1.81 |
| 4 | 3 | 0.00 | 40.01 | 167.67 | 75 | 1 | 75 | 1785.24 | 0.00 | 54.44 | 174.00 | 66 | 2 | 66 | 1.89 |
| 5 | 2 | -2.54 | 3.37 | 36.67 | 36 | 1 | 41 | 2070.11 | -2.54 | 8.80 | 60.47 | 40 | 2 | 41 | 2.35 |
| 5 | 3 | 0.00 | 24.57 | 167.63 | 81 | 2 | 81 | 1780.79 | 0.00 | 34.34 | 169.57 | 75 | 1 | 75 | 1.93 |
| 5 | 4 | -0.67 | 51.23 | 157.46 | 16 | 3 | 39 | 2295.31 | -0.67 | 49.69 | 157.53 | 15 | 2 | 38 | 1.89 |
| 6 | 2 | -1.93 | 6.88 | 42.23 | 57 | 1 | 57 | 2719.82 | -1.93 | 9.82 | 45.84 | 20 | 1 | 24 | 2.43 |
| 6 | 3 | -2.47 | 2.15 | 84.57 | 55 | 1 | 70 | 2657.54 | -2.47 | 13.90 | 115.23 | 50 | 1 | 61 | 2.24 |
| 6 | 4 | 0.00 | 36.97 | 166.22 | 74 | 1 | 74 | 1986.08 | 0.00 | 36.96 | 167.64 | 75 | 1 | 75 | 2.05 |
| 8 | 2 | -14.21 | -6.62 | 11.64 | 15 | 1 | 86 | 3776.64 | -14.21 | -5.97 | 10.24 | 19 | 1 | 70 | 4.87 |
| 8 | 3 | 0.00 | 5.16 | 27.97 | 33 | 1 | 33 | 4966.71 | 0.00 | 8.79 | 63.81 | 45 | 1 | 45 | 10.09 |
| 8 | 4 | 0.00 | 20.52 | 157.47 | 18 | 1 | 18 | 6768.68 | 0.00 | 31.81 | 134.96 | 17 | 1 | 17 | 2.61 |
| 8 | 6 | 0.00 | 49.79 | 179.76 | 54 | 1 | 54 | 16156.22 | 0.00 | 54.48 | 179.49 | 50 | 1 | 50 | 3.40 |
| 10 | 2 | -10.43 | -5.26 | 8.15 | 14 | 1 | 91 | 7920.38 | -9.83 | -1.84 | 12.22 | 2 | 1 | 67 | 7.99 |
| 10 | 3 | -1.19 | 5.52 | 42.04 | 57 | 1 | 60 | 18825.84 | -0.69 | 12.97 | 58.49 | 45 | 1 | 49 | 3.91 |
| 10 | 4 | 0.00 | 11.83 | 64.88 | 52 | 1 | 52 | 17464.07 | 0.00 | 25.08 | 90.37 | 49 | 1 | 49 | 2.88 |
| 10 | 6 | 0.00 | 22.75 | 151.47 | 33.33 | 1.08 | 33.33 | TL | 0.00 | 38.31 | 151.55 | 17 | 1 | 17 | 3.89 |
| 12 | 2 | -24.70 | -19.08 | -7.32 | 8 | 1 | 100 | 21375.43 | -20.77 | -14.22 | 0.31 | 1 | 1 | 99 | 6.74 |
| 12 | 3 | -2.57 | 1.24 | 61.09 | 58 | 1 | 75 | 69360.34 | -2.57 | 7.94 | 66.02 | 32 | 1 | 58 | 8.69 |
| 12 | 4 | -0.24 | 10.26 | 38.41 | 25 | 25 | 25 | TL | -0.24 | 17.67 | 77.13 | 10 | 1 | 10 | 3.61 |
| 12 | 6 | 0.00 | 14.16 | 76.62 | 15 | 2 | 15 | TL | 0.00 | 16.10 | 76.63 | 16 | 1 | 16 | 4.09 |
| **Average** | | -3.22 | 13.98 | 83.38 | 40.92 | 2.39 | 56.52 | 10794.51 | -2.96 | 20.68 | 94.41 | 33.30 | 1.15 | 47.60 | 3.97 |

Table D.6: Experimental Results for Datasets with $n = 80$ and $p = 4$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 16.58 | 137.03 | 45 | 1 | 45 | 1981.24 | 0.00 | 16.39 | 127.48 | 58 | 1 | 58 | 2.61 |
| 4 | 3 | 0.00 | 46.79 | 115.93 | 35 | 1 | 35 | 1614.88 | 0.00 | 44.63 | 117.77 | 33 | 2 | 33 | 2.29 |
| 5 | 2 | -2.90 | 10.37 | 41.83 | 38 | 1 | 38 | 2856.21 | -2.90 | 19.18 | 58.06 | 16 | 1 | 16 | 3.89 |
| 5 | 3 | 0.00 | 16.75 | 141.84 | 82 | 1 | 82 | 2570.39 | 0.00 | 44.25 | 237.55 | 63 | 1 | 63 | 2.24 |
| 5 | 4 | 0.00 | 65.94 | 216.98 | 19 | 1 | 19 | 2278.20 | 0.00 | 63.89 | 218.46 | 23 | 1 | 23 | 2.29 |
| 6 | 2 | -2.33 | 11.36 | 51.11 | 15 | 1 | 32 | 5661.03 | -2.33 | 9.72 | 76.02 | 11 | 1 | 32 | 4.17 |
| 6 | 3 | 0.00 | 0.73 | 53.89 | 90 | 1 | 90 | 5220.70 | 0.00 | 16.75 | 94.02 | 72 | 1 | 72 | 2.66 |
| 6 | 4 | 0.00 | 47.37 | 155.83 | 9 | 1 | 9 | 3490.49 | 0.00 | 50.88 | 216.13 | 14 | 1 | 14 | 2.64 |
| 8 | 2 | -9.38 | -0.49 | 18.91 | 16 | 1 | 52 | 24027.89 | -9.22 | 4.01 | 23.10 | 3 | 1 | 22 | 7.44 |
| 8 | 3 | 0.00 | 13.40 | 56.50 | 7 | 1 | 7 | 32590.25 | 0.00 | 22.91 | 86.71 | 10 | 1 | 10 | 4.13 |
| 8 | 4 | 0.00 | 11.27 | 105.29 | 1.04 | 1.04 | 1.04 | TL | -0.15 | 31.38 | 112.09 | 1 | 1 | 3 | 3.11 |
| 8 | 6 | 0.00 | 62.13 | 235.65 | 34 | 1 | 34 | 18004.02 | 0.00 | 70.25 | 235.46 | 32 | 1 | 32 | 3.43 |
| 10 | 2 | -12.47 | -4.52 | 5.49 | 1.92 | 1.92 | 84.62 | TL | -12.31 | -1.28 | 16.47 | 2 | 1 | 64 | 8.56 |
| 10 | 3 | 0.46 | 9.59 | 30.76 | 22.22 | 11.11 | 0 | TL | 0.71 | 15.96 | 81.26 | 1 | 1 | 0 | 4.55 |
| 10 | 4 | 0.00 | 8.59 | 41.49 | 25 | 6 | 25 | TL | 0.00 | 25.60 | 81.49 | 4 | 1 | 4 | 4.36 |
| 10 | 6 | 0.00 | 39.89 | 125.05 | 18.75 | 6.25 | 18.75 | TL | 0.00 | 57.18 | 212.46 | 20 | 1 | 20 | 4.34 |
| 12 | 2 | -23.86 | -18.07 | -11.87 | 10 | 10 | 100 | TL | -17.11 | -8.00 | 18.05 | 1 | 1 | 95 | 8.58 |
| 12 | 3 | 2.16 | 6.49 | 9.24 | 25 | 25 | 0 | TL | 0.00 | 18.57 | 53.35 | 5 | 1 | 5 | 5.54 |
| 12 | 4 | 4.29 | 31.11 | 47.47 | 25 | 25 | 0 | TL | 0.00 | 21.34 | 75.40 | 1 | 1 | 1 | 4.76 |
| 12 | 6 | -2.89 | 11.24 | 77.65 | 9.09 | 9.09 | 72.73 | TL | -2.89 | 30.24 | 134.07 | 10 | 1 | 54 | 5.54 |
| **Average** | | -2.35 | 19.33 | 82.80 | 26.40 | 5.33 | 37.26 | 9117.75 | -2.31 | 27.69 | 113.77 | 19.00 | 1.05 | 31.05 | 4.36 |

Table D.7: Experimental Results for Datasets with $n = 100$ and $p = 2$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | -0.52 | 2.64 | 157.70 | 98 | 2 | 98 | 2346.44 | -0.52 | -0.52 | -0.52 | 100 | 100 | 100 | 2.00 |
| 4 | 3 | 0.00* | 3.28 | 163.97 | 98 | 2 | 98 | 3062.48 | 0.00* | 1.64 | 163.97 | 99 | 1 | 99 | 1.95 |
| 5 | 2 | 0.00 | 2.73 | 48.51 | 63 | 2 | 63 | 2392.99 | 0.00 | 3.19 | 48.51 | 57 | 2 | 57 | 1.80 |
| 5 | 3 | 0.00 | 3.67 | 32.79 | 77 | 8 | 77 | 3306.48 | 0.00 | 2.03 | 112.59 | 86 | 1 | 86 | 2.68 |
| 5 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 3647.99 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 2.42 |
| 6 | 2 | 0.00 | 1.25 | 55.11 | 80 | 2 | 80 | 3141.05 | 0.00 | 1.05 | 9.08 | 73 | 10 | 73 | 2.34 |
| 6 | 3 | 0.00 | 5.93 | 118.69 | 95 | 5 | 95 | 3252.96 | 0.00 | 1.72 | 172.00 | 99 | 1 | 99 | 2.78 |
| 6 | 4 | 0.00 | 2.52 | 7.37 | 56 | 24 | 56 | 4242.78 | 0.00 | 3.91 | 113.08 | 53 | 1 | 53 | 2.48 |
| 8 | 2 | -1.20 | 2.48 | 55.78 | 93 | 1 | 93 | 5000.91 | -1.20 | 2.30 | 80.11 | 77 | 1 | 77 | 3.26 |
| 8 | 3 | 0.00 | 0.33 | 1.64 | 80 | 20 | 80 | 3898.12 | 0.00 | 0.34 | 1.64 | 79 | 21 | 79 | 3.06 |
| 8 | 4 | 0.00 | 4.59 | 132.81 | 85 | 2 | 85 | 4025.98 | 0.00 | 7.61 | 160.25 | 84 | 1 | 84 | 3.14 |
| 8 | 6 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 5265.75 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 4.06 |
| 10 | 2 | -3.35 | -0.23 | 62.38 | 91 | 1 | 91 | 6536.77 | -3.35 | -1.12 | 69.53 | 93 | 1 | 93 | 3.88 |
| 10 | 3 | 0.00 | 1.72 | 73.45 | 95 | 1 | 95 | 5325.46 | 0.00 | 3.38 | 91.23 | 84 | 1 | 84 | 3.56 |
| 10 | 4 | 0.00 | 5.02 | 77.66 | 74 | 1 | 74 | 5709.09 | 0.00 | 8.23 | 103.00 | 68 | 1 | 68 | 3.47 |
| 10 | 6 | 0.00 | 2.10 | 158.09 | 87 | 1 | 87 | 20570.94 | 0.00 | 1.98 | 150.29 | 88 | 1 | 88 | 4.56 |
| 12 | 2 | 0.00 | 4.79 | 83.99 | 68 | 1 | 68 | 17839.78 | 0.00 | 10.87 | 52.74 | 9 | 1 | 9 | 7.75 |
| 12 | 3 | -0.02 | 5.35 | 55.27 | 83 | 2 | 83 | 11646.74 | -0.02 | 8.65 | 83.64 | 86 | 1 | 86 | 4.38 |
| 12 | 4 | 0.00 | 7.28 | 90.48 | 52 | 1 | 52 | 54233.10 | 0.00 | 7.80 | 104.58 | 58 | 1 | 58 | 4.08 |
| 12 | 6 | 0.00 | 1.52 | 9.88 | 63 | 1 | 63 | 109036.17 | 0.00 | 3.87 | 105.14 | 55 | 1 | 55 | 4.71 |
| **Average** | | -0.25 | 2.85 | 69.28 | 81.90 | 13.86 | 81.91 | 8707.67 | -0.25 | 3.35 | 81.04 | 77.40 | 17.35 | 77.40 | 3.42 |

Table D.8: Experimental Results for Datasets with $n = 100$ and $p = 3$

| | | Matheuristic Method | | | | | | | Heuristic Method | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 8.52 | 39.68 | 26 | 1 | 26 | 2428.52 | 0.00 | 26.80 | 82.51 | 11 | 1 | 11 | 3.41 |
| 4 | 3 | 0.00 | 32.36 | 169.33 | 76 | 1 | 76 | 3869.32 | 0.00 | 51.13 | 170.53 | 62 | 3 | 62 | 2.35 |
| 5 | 2 | -0.17 | 6.36 | 52.27 | 24 | 1 | 50 | 2910.87 | -0.17 | 9.73 | 56.53 | 24 | 1 | 28 | 3.21 |
| 5 | 3 | 0.00 | 31.57 | 185.00 | 58 | 1 | 58 | 4549.14 | 0.00 | 40.64 | 273.74 | 55 | 1 | 55 | 2.67 |
| 5 | 4 | 0.00 | 25.82 | 110.35 | 68 | 1 | 68 | 5369.06 | 0.00 | 26.79 | 109.67 | 62 | 1 | 62 | 3.01 |
| 6 | 2 | -3.85 | 2.91 | 38.99 | 64 | 1 | 64 | 3522.81 | -3.85 | 3.31 | 52.32 | 60 | 1 | 60 | 3.88 |
| 6 | 3 | 0.00 | 5.84 | 121.62 | 93 | 1 | 93 | 3400.41 | 0.00 | 30.15 | 127.68 | 74 | 1 | 74 | 2.89 |
| 6 | 4 | 0.00 | 35.71 | 192.61 | 76 | 1 | 76 | 2936.83 | 0.00 | 49.14 | 193.83 | 68 | 1 | 68 | 4.72 |
| 8 | 2 | -1.35 | 4.66 | 31.42 | 32 | 1 | 32 | 7146.38 | 1.62 | 5.62 | 41.53 | 36 | 1 | 0 | 9.23 |
| 8 | 3 | 0.00 | 7.97 | 36.50 | 26 | 1 | 26 | 7570.05 | 0.00 | 15.00 | 46.09 | 7 | 1 | 7 | 5.14 |
| 8 | 4 | 0.00 | 11.65 | 105.32 | 54 | 3 | 54 | 17541.69 | 0.00 | 21.39 | 134.96 | 51 | 1 | 51 | 4.72 |
| 8 | 6 | -0.50 | 21.90 | 102.94 | 12 | 1 | 78 | 30186.60 | -0.50 | 24.94 | 103.67 | 13 | 1 | 75 | 4.50 |
| 10 | 2 | -6.18 | -0.26 | 13.72 | 10 | 2 | 56 | 13243.00 | -3.97 | 8.15 | 39.49 | 6 | 1 | 16 | 10.09 |
| 10 | 3 | -0.44 | 5.35 | 35.87 | 62 | 1 | 62 | 29645.93 | -0.44 | 20.34 | 59.33 | 7 | 1 | 19 | 5.70 |
| 10 | 4 | 0.00 | 10.26 | 80.80 | 57 | 1 | 57 | 25424.00 | 0.00 | 26.84 | 97.71 | 40 | 1 | 40 | 4.18 |
| 10 | 6 | 0.00 | 20.68 | 163.04 | 56.16 | 1.37 | 56.16 | TL | 0.00 | 30.69 | 165.07 | 43 | 1 | 43 | 4.82 |
| 12 | 2 | -16.08 | -10.33 | 1.10 | 8 | 1 | 98 | 38372.99 | -16.08 | -5.75 | 14.67 | 2 | 1 | 85 | 9.96 |
| 12 | 3 | -1.00 | 5.12 | 36.15 | 25.93 | 1.23 | 44.44 | TL | -1.00 | 11.88 | 73.06 | 5 | 1 | 19 | 5.81 |
| 12 | 4 | -0.78 | 5.42 | 39.28 | 4.65 | 2.33 | 39.53 | TL | -0.78 | 12.10 | 45.38 | 4 | 1 | 28 | 4.95 |
| 12 | 6 | 0.00 | 7.15 | 70.23 | 13.04 | 4.35 | 13.04 | TL | 0.00 | 23.01 | 132.32 | 14 | 1 | 14 | 6.72 |
| **Average** | | -1.52 | 11.93 | 81.31 | 42.29 | 1.41 | 56.36 | 12382.35 | -1.26 | 21.59 | 101.00 | 32.20 | 1.10 | 40.85 | 5.10 |

Table D.9: Experimental Results for Datasets with $n = 100$ and $p = 4$

| | | **Matheuristic Method** | | | | | | | **Heuristic Method** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time |
| 4 | 2 | 0.00 | 10.96 | 112.55 | 45 | 1 | 45 | 3050.79 | 0.00 | 28.51 | 124.56 | 34 | 1 | 34 | 2.99 |
| 4 | 3 | 0.00 | 40.22 | 116.52 | 18 | 1 | 18 | 2761.42 | 0.00 | 43.89 | 117.85 | 15 | 1 | 15 | 2.80 |
| 5 | 2 | -0.29 | 8.56 | 76.18 | 57 | 1 | 57 | 4179.29 | -0.29 | 21.89 | 102.13 | 44 | 1 | 44 | 3.77 |
| 5 | 3 | 0.00 | 20.93 | 129.00 | 67 | 1 | 67 | 3863.16 | 0.00 | 40.35 | 199.24 | 48 | 1 | 48 | 3.63 |
| 5 | 4 | 0.00 | 61.99 | 220.36 | 46 | 1 | 46 | 3827.98 | 0.00 | 55.60 | 220.75 | 52 | 1 | 52 | 4.01 |
| 6 | 2 | -1.93 | 5.51 | 47.48 | 73 | 1 | 73 | 7638.35 | -1.93 | 11.21 | 94.93 | 68 | 1 | 68 | 6.42 |
| 6 | 3 | -1.44 | 6.49 | 89.81 | 41 | 1 | 56 | 7288.35 | -1.44 | 22.07 | 150.07 | 27 | 1 | 37 | 4.10 |
| 6 | 4 | 0.00 | 49.30 | 145.90 | 60 | 1 | 60 | 10757.87 | 0.00 | 52.75 | 146.63 | 57 | 1 | 57 | 3.63 |
| 8 | 2 | -20.62 | -17.66 | 3.58 | 6 | 1 | 99 | 38102.88 | -20.50 | -14.42 | 12.58 | 2 | 1 | 98 | 11.55 |
| 8 | 3 | 0.00 | 14.35 | 63.93 | 39 | 1 | 39 | 65181.06 | 0.00 | 26.20 | 100.43 | 19 | 1 | 19 | 7.10 |
| 8 | 4 | 0.00 | 23.16 | 116.63 | 70.77 | 1.54 | 70.77 | TL | 0.00 | 43.73 | 139.03 | 50 | 1 | 50 | 3.94 |
| 8 | 6 | 0.00 | 35.19 | 231.51 | 41 | 1 | 41 | 82964.06 | 0.00 | 37.49 | 233.87 | 37 | 1 | 37 | 6.75 |
| 10 | 2 | -24.00 | -18.40 | -2.60 | 11.11 | 1.85 | 100 | TL | -23.33 | -15.97 | 7.34 | 1 | 1 | 98 | 11.30 |
| 10 | 3 | -0.03 | 6.78 | 15.15 | 12.50 | 12.50 | 12.50 | TL | 0.90 | 15.99 | 53.43 | 3 | 1 | 0 | 7.76 |
| 10 | 4 | -2.36 | 13.81 | 41.79 | 14.29 | 7.14 | 42.86 | TL | -2.36 | 19.68 | 118.77 | 27 | 1 | 46 | 5.55 |
| 10 | 6 | 0.00 | 20.60 | 82.95 | 62.5 | 12.5 | 62.5 | TL | 0.00 | 53.65 | 279.29 | 38 | 1 | 38 | 5.99 |
| 12 | 2 | -24.43 | -19.57 | -9.83 | 10 | 10 | 100 | TL | -22.57 | -14.24 | 12.33 | 1 | 1 | 94 | 13.30 |
| 12 | 3 | 7.72 | 9.49 | 11.26 | 50 | 50 | 0 | TL | 0.00 | 12.08 | 41.15 | 1 | 1 | 1 | 8.13 |
| 12 | 4 | 6.20 | 30.58 | 41.81 | 20 | 20 | 0 | TL | 0.00 | 17.17 | 74.17 | 3 | 1 | 3 | 6.49 |
| 12 | 6 | 91.79 | 91.79 | 91.79 | 100 | 100 | 0 | TL | -2.37 | 25.64 | 144.17 | 1 | 1 | 14 | 7.68 |
| **Average** | | 1.53 | 19.70 | 81.29 | 42.21 | 11.33 | 49.48 | 20874.11 | -3.69 | 24.16 | 118.64 | 26.40 | 1.00 | 42.65 | 6.34 |

Table D.10: Experimental Results for Data Sets with $n = 200$ and $p = 2$

| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.39 | 3.33 | 17.34 | 79 | 21 | 79 | 7.52 |
| 4 | 3 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 7.51 |
| 5 | 2 | -1.87 | 4.55 | 79.32 | 1 | 1 | 18 | 11.05 |
| 5 | 3 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 7.88 |
| 5 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 7.73 |
| 6 | 2 | 0.00 | 5.26 | 122.65 | 89 | 1 | 89 | 8.37 |
| 6 | 3 | 0.00 | 3.07 | 46.33 | 73 | 2 | 73 | 9.31 |
| 6 | 4 | 0.00 | 0.28 | 5.52 | 95 | 5 | 95 | 8.96 |
| 8 | 2 | -1.73 | 6.75 | 61.32 | 4 | 2 | 13 | 10.40 |
| 8 | 3 | -0.60 | 5.36 | 75.46 | 66 | 1 | 66 | 11.00 |
| 8 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 9.48 |
| 8 | 6 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 13.71 |
| 10 | 2 | -2.37 | 10.57 | 55.76 | 26 | 1 | 26 | 13.45 |
| 10 | 3 | -0.31 | 7.93 | 82.88 | 65 | 1 | 85 | 12.26 |
| 10 | 4 | 0.00 | 5.92 | 95.22 | 83 | 2 | 83 | 11.33 |
| 10 | 6 | 0.00 | 1.25 | 8.30 | 57 | 5 | 57 | 13.68 |
| 12 | 2 | -1.18 | 7.98 | 64.05 | 30 | 1 | 30 | 12.29 |
| 12 | 3 | -0.13 | 6.13 | 78.35 | 72 | 1 | 72 | 14.13 |
| 12 | 4 | 0.00 | 4.67 | 119.89 | 86 | 1 | 86 | 11.79 |
| 12 | 6 | 0.00 | 2.69 | 5.68 | 41 | 1 | 41 | 15.61 |
| **Average** | | -0.43 | 3.79 | 45.90 | 68.35 | 27.30 | 70.65 | 10.87 |

Table D.11: Experimental Results for Data Sets with $n = 200$ and $p = 3$

| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 0.00 | 26.79 | 108.00 | 58 | 1 | 58 | 10.60 |
| 4 | 3 | 0.00 | 58.35 | 163.02 | 58 | 5 | 58 | 9.34 |
| 5 | 2 | -1.18 | 10.62 | 115.37 | 8 | 1 | 49 | 16.11 |
| 5 | 3 | 0.00 | 42.22 | 178.49 | 68 | 1 | 68 | 9.69 |
| 5 | 4 | 0.00 | 50.02 | 156.00 | 48 | 1 | 48 | 15.97 |
| 6 | 2 | -1.53 | 3.75 | 60.12 | 65 | 1 | 65 | 13.02 |
| 6 | 3 | 0.00 | 29.10 | 136.89 | 63 | 1 | 63 | 9.78 |
| 6 | 4 | 0.00 | 17.03 | 178.90 | 90 | 1 | 90 | 9.66 |
| 8 | 2 | -1.56 | 8.46 | 29.13 | 1 | 1 | 5 | 32.04 |
| 8 | 3 | 0.00 | 23.78 | 62.77 | 22 | 1 | 22 | 15.05 |
| 8 | 4 | 0.00 | 22.45 | 139.48 | 59 | 1 | 59 | 13.83 |
| 8 | 6 | -0.30 | 54.62 | 176.71 | 10 | 1 | 39 | 13.45 |
| 10 | 2 | -2.66 | 2.77 | 10.56 | 2 | 1 | 25 | 43.34 |
| 10 | 3 | -0.17 | 11.10 | 55.60 | 30 | 1 | 30 | 15.66 |
| 10 | 4 | 0.00 | 16.61 | 66.00 | 61 | 1 | 61 | 13.86 |
| 10 | 6 | 0.00 | 39.86 | 161.93 | 45 | 1 | 45 | 14.28 |
| 12 | 2 | -7.48 | -0.83 | 11.72 | 1 | 1 | 62 | 48.21 |
| 12 | 3 | -0.05 | 6.99 | 57.92 | 18 | 1 | 18 | 26.16 |
| 12 | 4 | 0.00 | 17.12 | 70.41 | 40 | 1 | 40 | 13.80 |
| 12 | 6 | 0.00 | 26.39 | 135.13 | 51 | 1 | 51 | 16.59 |
| **Average** | | -0.75 | 23.36 | 103.71 | 39.90 | 1.20 | 47.80 | 18.02 |

Table D.12: Experimental Results for Data Sets with $n = 200$ and $p = 4$

| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 0.00 | 19.08 | 132.59 | 11 | 1 | 11 | 18.46 |
| 4 | 3 | 0.00 | 43.01 | 119.07 | 43 | 1 | 43 | 9.75 |
| 5 | 2 | 0.44 | 6.53 | 47.03 | 20 | 1 | 0 | 29.54 |
| 5 | 3 | 0.00 | 33.79 | 209.17 | 39 | 1 | 39 | 10.66 |
| 5 | 4 | 0.00 | 66.50 | 226.13 | 26 | 1 | 26 | 9.91 |
| 6 | 2 | -1.98 | 6.21 | 46.10 | 37 | 1 | 70 | 15.79 |
| 6 | 3 | -0.12 | 23.51 | 148.97 | 41 | 1 | 41 | 12.06 |
| 6 | 4 | 0.00 | 41.81 | 145.66 | 53 | 1 | 53 | 13.05 |
| 8 | 2 | -16.42 | -10.90 | 8.82 | 2 | 1 | 98 | 33.09 |
| 8 | 3 | 0.00 | 22.96 | 78.16 | 9 | 1 | 9 | 23.91 |
| 8 | 4 | 0.00 | 25.03 | 123.14 | 20 | 1 | 20 | 13.55 |
| 8 | 6 | 0.00 | 57.04 | 217.56 | 30 | 1 | 30 | 14.90 |
| 10 | 2 | -21.68 | -16.68 | 4.28 | 1 | 1 | 98 | 39.46 |
| 10 | 3 | 0.96 | 17.06 | 51.65 | 1 | 1 | 0 | 22.52 |
| 10 | 4 | 0.00 | 29.28 | 142.43 | 42 | 1 | 42 | 15.89 |
| 10 | 6 | 0.00 | 52.21 | 245.65 | 35 | 1 | 35 | 14.70 |
| 12 | 2 | -25.01 | -17.25 | -4.57 | 1 | 1 | 100 | 47.77 |
| 12 | 3 | 0.00 | 9.09 | 53.72 | 63 | 1 | 63 | 27.04 |
| 12 | 4 | 0.00 | 16.92 | 54.86 | 34 | 1 | 34 | 18.86 |
| 12 | 6 | -0.93 | 30.72 | 134.07 | 2 | 1 | 23 | 21.31 |
| **Average** | | -3.24 | 22.80 | 109.22 | 25.50 | 1.00 | 41.75 | 20.61 |

Table D.13: Experimental Results for Data Sets with $n = 500$ and $p = 2$

| m | q | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 53.10 |
| 4 | 3 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 39.44 |
| 5 | 2 | 0.00 | 2.28 | 27.39 | 87 | 3 | 87 | 59.05 |
| 5 | 3 | 0.00 | 3.71 | 137.64 | 97 | 1 | 97 | 51.69 |
| 5 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 43.64 |
| 6 | 2 | -0.26 | 6.92 | 94.55 | 51 | 1 | 51 | 73.28 |
| 6 | 3 | 0.00 | 4.98 | 52.38 | 78 | 1 | 78 | 64.52 |
| 6 | 4 | 0.00 | 1.97 | 196.52 | 99 | 1 | 99 | 45.16 |
| 8 | 2 | -2.22 | 15.19 | 32.26 | 4 | 1 | 6 | 95.77 |
| 8 | 3 | 0.00 | 2.76 | 107.13 | 97 | 1 | 97 | 57.65 |
| 8 | 4 | 0.00 | 2.40 | 13.08 | 70 | 2 | 70 | 54.47 |
| 8 | 6 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 57.39 |
| 10 | 2 | -1.05 | 10.49 | 27.80 | 2 | 1 | 21 | 325.02 |
| 10 | 3 | 0.00 | 6.63 | 53.75 | 55 | 1 | 55 | 80.71 |
| 10 | 4 | 0.00 | 6.17 | 78.20 | 58 | 1 | 58 | 69.45 |
| 10 | 6 | 0.00 | 0.54 | 6.82 | 92 | 4 | 92 | 71.91 |
| 12 | 2 | 0.00 | 17.76 | 33.85 | 1 | 1 | 1 | 305.02 |
| 12 | 3 | -0.42 | 12.94 | 35.90 | 1 | 1 | 5 | 110.41 |
| 12 | 4 | 0.00 | 6.68 | 58.59 | 39 | 1 | 39 | 67.54 |
| 12 | 6 | 0.00 | 0.65 | 5.67 | 86 | 5 | 86 | 91.27 |
| **Average** | | -0.20 | 5.10 | 48.08 | 65.85 | 21.30 | 67.10 | 90.82 |

Table D.14: Experimental Results for Data Sets with $n = 500$ and $p = 3$

| m | q | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.10 | 10.73 | 177.18 | 91 | 1 | 91 | 61.70 |
| 4 | 3 | 0.00 | 52.34 | 158.65 | 49 | 1 | 49 | 55.28 |
| 5 | 2 | -3.61 | 3.19 | 45.83 | 56 | 1 | 81 | 99.01 |
| 5 | 3 | 0.00 | 36.29 | 268.23 | 70 | 1 | 70 | 51.32 |
| 5 | 4 | 0.00 | 51.34 | 161.72 | 60 | 1 | 60 | 59.83 |
| 6 | 2 | -0.87 | 5.55 | 47.28 | 73 | 1 | 73 | 90.78 |
| 6 | 3 | 0.00 | 18.14 | 111.02 | 55 | 1 | 55 | 60.49 |
| 6 | 4 | 0.00 | 35.03 | 175.20 | 69 | 1 | 69 | 56.59 |
| 8 | 2 | -3.30 | 7.25 | 43.81 | 1 | 1 | 6 | 149.33 |
| 8 | 3 | 0.00 | 17.79 | 80.02 | 50 | 1 | 50 | 95.58 |
| 8 | 4 | 0.00 | 28.40 | 154.27 | 59 | 1 | 59 | 66.06 |
| 8 | 6 | 0.00 | 47.28 | 191.44 | 22 | 1 | 22 | 69.27 |
| 10 | 2 | -8.51 | -4.60 | 10.09 | 4 | 1 | 84 | 191.54 |
| 10 | 3 | -1.27 | 6.87 | 54.94 | 5 | 1 | 6 | 97.36 |
| 10 | 4 | 0.00 | 21.24 | 135.76 | 36 | 1 | 36 | 63.78 |
| 10 | 6 | 0.00 | 31.99 | 151.05 | 37 | 1 | 37 | 75.97 |
| 12 | 2 | -1.22 | 6.50 | 29.09 | 1 | 1 | 2 | 285.09 |
| 12 | 3 | 4.33 | 21.28 | 43.75 | 1 | 1 | 0 | 106.50 |
| 12 | 4 | 0.00 | 17.02 | 67.53 | 14 | 1 | 14 | 93.28 |
| 12 | 6 | 0.00 | 18.94 | 104.12 | 51 | 1 | 51 | 82.54 |
| **Average** | | -0.73 | 21.63 | 110.55 | 40.20 | 1.00 | 45.75 | 95.57 |

Table D.15: Experimental Results for Data Sets with $n = 500$ and $p = 4$

| m | q | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.04 | 22.25 | 84.44 | 20 | 1 | 20 | 93.00 |
| 4 | 3 | 0.00 | 37.42 | 183.95 | 43 | 1 | 43 | 58.51 |
| 5 | 2 | -2.03 | 8.30 | 89.59 | 77 | 1 | 78 | 79.99 |
| 5 | 3 | 0.00 | 31.77 | 211.58 | 62 | 1 | 62 | 57.88 |
| 5 | 4 | 0.00 | 50.30 | 232.00 | 56 | 1 | 56 | 54.11 |
| 6 | 2 | -2.18 | 4.33 | 38.64 | 48 | 1 | 48 | 101.62 |
| 6 | 3 | 0.00 | 39.86 | 139.01 | 56 | 1 | 56 | 55.41 |
| 6 | 4 | 0.00 | 50.02 | 208.16 | 47 | 1 | 47 | 58.37 |
| 8 | 2 | -1.51 | 4.68 | 20.16 | 2 | 1 | 10 | 205.52 |
| 8 | 3 | 0.00 | 21.58 | 79.60 | 31 | 1 | 31 | 156.14 |
| 8 | 4 | 0.00 | 26.79 | 101.49 | 37 | 1 | 37 | 68.33 |
| 8 | 6 | 0.00 | 58.75 | 219.57 | 53 | 1 | 53 | 59.72 |
| 10 | 2 | -2.84 | 3.91 | 11.90 | 1 | 1 | 13 | 225.91 |
| 10 | 3 | -0.63 | 19.75 | 73.19 | 20 | 1 | 20 | 115.82 |
| 10 | 4 | 0.00 | 26.41 | 99.46 | 30 | 1 | 30 | 77.61 |
| 10 | 6 | 0.00 | 41.70 | 190.02 | 16 | 1 | 16 | 72.97 |
| 12 | 2 | -19.70 | -9.70 | -2.92 | 1 | 1 | 100 | 245.83 |
| 12 | 3 | -0.19 | 27.09 | 46.26 | 3 | 1 | 3 | 135.95 |
| 12 | 4 | 0.00 | 24.03 | 110.23 | 27 | 1 | 27 | 100.71 |
| 12 | 6 | 0.00 | 37.22 | 158.03 | 12 | 1 | 12 | 90.24 |
| **Average** | | -1.46 | 26.32 | 114.72 | 32.10 | 1.00 | 38.10 | 105.68 |

Table D.16: Experimental Results for Data Sets with $n = 1000$ and $p = 2$

| m | q | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.08 | 4.97 | 90.72 | 85 | 1 | 85 | 325.14 |
| 4 | 3 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 163.92 |
| 5 | 2 | -0.22 | 0.44 | 65.51 | 99 | 1 | 99 | 449.23 |
| 5 | 3 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 205.13 |
| 5 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 171.56 |
| 6 | 2 | 0.00 | 5.60 | 29.10 | 65 | 1 | 65 | 239.08 |
| 6 | 3 | 0.00 | 4.86 | 78.17 | 57 | 1 | 57 | 255.82 |
| 6 | 4 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 178.22 |
| 8 | 2 | -1.85 | 14.12 | 38.14 | 10 | 1 | 12 | 433.94 |
| 8 | 3 | -0.03 | 6.92 | 48.42 | 46 | 1 | 46 | 262.81 |
| 8 | 4 | 0.00 | 1.06 | 106.10 | 99 | 1 | 99 | 225.68 |
| 8 | 6 | 0.00 | 0.00 | 0.00 | 100 | 100 | 100 | 219.81 |
| 10 | 2 | -0.95 | 7.43 | 20.79 | 14 | 1 | 46 | 1164.38 |
| 10 | 3 | 0.00 | 9.44 | 72.70 | 29 | 1 | 29 | 300.12 |
| 10 | 4 | 0.00 | 5.50 | 66.23 | 44 | 1 | 44 | 296.96 |
| 10 | 6 | 0.00 | 0.97 | 4.86 | 80 | 4 | 80 | 296.83 |
| 12 | 2 | 0.00 | 14.88 | 20.90 | 7 | 1 | 7 | 1067.70 |
| 12 | 3 | -0.20 | 12.88 | 33.17 | 2 | 1 | 4 | 481.24 |
| 12 | 4 | 0.00 | 4.11 | 44.73 | 47 | 1 | 47 | 291.65 |
| 12 | 6 | 0.00 | 2.30 | 9.30 | 54 | 1 | 54 | 335.34 |
| **Average** | | -0.17 | 4.77 | 36.44 | 61.90 | 25.90 | 63.70 | 368.23 |

Table D.17: Experimental Results for Data Sets with $n = 1000$ and $p = 3$

| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.14 | 15.53 | 116.59 | 84 | 1 | 84 | 324.36 |
| 4 | 3 | 0.00 | 46.44 | 245.60 | 73 | 1 | 73 | 204.29 |
| 5 | 2 | -1.20 | 7.67 | 54.28 | 64 | 1 | 64 | 382.15 |
| 5 | 3 | 0.00 | 41.32 | 185.41 | 68 | 1 | 68 | 225.68 |
| 5 | 4 | 0.00 | 56.63 | 229.73 | 70 | 1 | 70 | 212.74 |
| 6 | 2 | -1.20 | 2.73 | 48.10 | 84 | 1 | 84 | 429.46 |
| 6 | 3 | -0.07 | 18.48 | 137.66 | 68 | 1 | 68 | 291.35 |
| 6 | 4 | 0.00 | 38.26 | 178.36 | 56 | 1 | 56 | 228.00 |
| 8 | 2 | -1.30 | 5.48 | 42.38 | 18 | 1 | 31 | 797.30 |
| 8 | 3 | 0.00 | 13.40 | 77.18 | 80 | 1 | 80 | 256.69 |
| 8 | 4 | 0.00 | 14.46 | 171.59 | 64 | 1 | 64 | 274.32 |
| 8 | 6 | 0.00 | 51.34 | 219.49 | 64 | 1 | 64 | 236.28 |
| 10 | 2 | -9.71 | -7.87 | 8.39 | 3 | 1 | 97 | 910.12 |
| 10 | 3 | -1.00 | 11.35 | 59.73 | 11 | 1 | 40 | 354.12 |
| 10 | 4 | 0.00 | 20.55 | 130.78 | 38 | 1 | 38 | 241.05 |
| 10 | 6 | 0.00 | 45.42 | 168.10 | 47 | 1 | 47 | 278.44 |
| 12 | 2 | -7.24 | -2.42 | 3.98 | 1 | 1 | 77 | 1088.62 |
| 12 | 3 | -0.66 | 11.10 | 37.61 | 13 | 1 | 13 | 493.96 |
| 12 | 4 | 0.00 | 14.06 | 91.09 | 67 | 1 | 67 | 276.62 |
| 12 | 6 | 0.00 | 19.18 | 88.08 | 52 | 1 | 52 | 337.47 |
| **Average** | | -1.13 | 21.16 | 114.71 | 51.25 | 1.00 | 61.85 | 392.15 |

Table D.18: Experimental Results for Data Sets with $n = 1000$ and $p = 4$

| **m** | **q** | $\%Gap_{Best}$ | $\%Gap_{Avg}$ | $\%Gap_{Worst}$ | $\#Best$ | $\#Worst$ | $\#IBase$ | Total Time (sec) |
|---|---|---|---|---|---|---|---|---|
| 4 | 2 | -0.05 | 12.25 | 130.12 | 63 | 1 | 63 | 312.60 |
| 4 | 3 | 0.00 | 43.89 | 264.74 | 70 | 1 | 70 | 197.32 |
| 5 | 2 | -0.50 | 2.46 | 60.59 | 31 | 1 | 87 | 300.57 |
| 5 | 3 | 0.00 | 36.07 | 138.03 | 66 | 1 | 66 | 205.70 |
| 5 | 4 | 0.00 | 64.79 | 245.32 | 54 | 1 | 54 | 219.28 |
| 6 | 2 | -1.25 | 3.67 | 42.92 | 47 | 1 | 75 | 427.20 |
| 6 | 3 | -0.03 | 30.63 | 204.18 | 52 | 1 | 52 | 212.68 |
| 6 | 4 | 0.00 | 36.30 | 116.82 | 48 | 1 | 48 | 211.62 |
| 8 | 2 | -0.76 | 10.83 | 35.78 | 9 | 1 | 9 | 716.83 |
| 8 | 3 | 0.00 | 14.84 | 90.26 | 47 | 1 | 47 | 412.88 |
| 8 | 4 | 0.00 | 28.97 | 145.88 | 52 | 1 | 52 | 290.74 |
| 8 | 6 | 0.00 | 59.05 | 227.21 | 53 | 1 | 53 | 220.38 |
| 10 | 2 | -0.23 | 5.13 | 12.90 | 1 | 1 | 8 | 870.64 |
| 10 | 3 | -0.69 | 18.17 | 52.47 | 1 | 1 | 8 | 529.14 |
| 10 | 4 | 0.00 | 27.15 | 97.58 | 6 | 1 | 6 | 267.24 |
| 10 | 6 | 0.00 | 44.67 | 210.98 | 27 | 1 | 27 | 254.16 |
| 12 | 2 | -21.21 | -13.30 | -4.71 | 1 | 1 | 100 | 950.37 |
| 12 | 3 | 3.55 | 22.00 | 46.96 | 1 | 1 | 0 | 642.74 |
| 12 | 4 | 0.00 | 20.38 | 114.62 | 26 | 1 | 26 | 369.28 |
| 12 | 6 | 0.00 | 39.71 | 117.05 | 18 | 1 | 18 | 277.49 |
| **Average** | | -1.06 | 25.38 | 117.48 | 33.65 | 1.00 | 43.45 | 394.44 |