# Advances on logistics and transportation problems on complex networks: Evaluation and conclusions

**1 Enero 2017 ---- 29 Junio 2021**

JUSTO PUERTO

FUENGIROLA 23-26 JUNIO, 2021


CITIUS
imus

Contenido:

1. 5 Minutos sobre conos

2. 5 minutos sobre polinomios, el problema de los momentos y la optimización global

3. 5 minutos sobre optimización con orden

4. 5 minutos sobre redes y ciencia de datos

5. 5 minutos sobre aplicaciones

# 5 minutos sobre conos: "pros and cons"

## Conic Programming and Location

A general conic program can be written as:

$$\begin{array}{ll} \min & c^t x \\ \text{s.a:} & Ax - b \geq_K 0, \end{array}$$

where $K$ is a closed, convex, pointed cone, with nonempty interior. Besides, $a \geq_K b \equiv a - b \in K$.

$$\begin{array}{ll} \min & c^t x \\ \text{s.a:} & Ax - b \geq 0, \\ & x \geq 0. \end{array}$$

In this case $K = \mathbb{R}^n_+$.

**Remarkable examples**

1. $K = \mathbb{R}^n_+$. Linear programming.
2. $L^n := \{x \in \mathbb{R}^n : \sqrt{\sum_{i=1}^{n-1} x_i^2} \leq x_n\} \quad \equiv \quad \|(x_1, \ldots, x_{n-1})\|_2 \leq x_n$. Second order cone (SOCP) or Lorentz cone.
3. $K = \mathbb{S}^m_+$ : $A \in S^m_+$ iff $A \in \mathbb{R}^{m \times m}$, $A^t = A$, $x^t A x \geq 0$, $\forall\, x \in \mathbb{R}^m$.
4. $K = \mathscr{C}^n := \{A \in \mathbb{S}^n : x^t A x \geq 0;\ x \geq 0\}$.

Let

$$\mathscr{A}X = \begin{pmatrix} A_1 \bullet X \\ A_2 \bullet X \\ \cdots \\ A_m \bullet X \end{pmatrix}, \qquad \begin{array}{ll} (CLP)\ \min & C \bullet X \\ \text{s.t.} & \mathscr{A}X = b,\ X \in K. \end{array}$$

**Theorem (Farkas Lemma for conic linear programming)**

*It holds:*

- *Consider the set $\mathscr{F}_p = \{X : \mathscr{A}X = b, X \in K\}$. Suppose that there exists a vector $\dot{y}$ such that $-\dot{y}\mathscr{A} \in$ int $K^*$. Then,*
  1. *The set $C := \{\mathscr{A}X \in \mathbb{R}^m : X \in K\}$ is a closed convex set;*
  2. *$\mathscr{F}_p$ is not empty if and only if $\{y : -y^t\mathscr{A} \in K^*,\ y^t b > 0\}$ is empty.*
- *Consider the set $\mathscr{F}_d = \{y : C^t - y^t\mathscr{A} \in K^*\}$. Suppose that there exists a vector $\dot{X} \in$ int $K$ such that $\mathscr{A}\dot{X} = 0$. Then,*
  1. *$C := \{S - y^t\mathscr{A} : S \in K^*\}$ is a closed convex set;*
  2. *$\mathscr{F}_d$ is not empty if and only if the set $\{X : \mathscr{A}X = 0,\ X \in K,\ C \bullet X < 0\}$ is empty.*

# Resultados esperables.

**Problem (P) and its dual**

$$\begin{aligned} &\min \langle a_0, x\rangle \\ &\text{s.t. } \langle a_p, x\rangle = b_p,\ p = 1,\dots,m \\ &\quad x \in K. \end{aligned}$$

$$\begin{aligned} &\max \sum_{p=1}^{m} b_p y_p \\ &\text{s.t. } \sum_{p=1}^{m} a_p y_p + s = a_0, \\ &\quad s \in K^*. \end{aligned}$$

**Duality theorems.**

- $0 \le \langle s, x\rangle = \langle a_0, x\rangle - \sum_{p=1}^{m} b_p y_p, \quad \forall$ feasible $(x, y, s)$.
- If exists an interior solution $(x, y, s)$ $(x \in K,\ s \in K^*)$ then

$$0 = \langle s, x\rangle = \langle a_0, x\rangle - \sum_{p=1}^{m} b_p y_p, \quad \forall \text{ optimal } (x^*, y^*, s^*).$$

Linear optimization over convex cone $K$

**Self-concordance theory.**

- Polynomial-time interior point algorithms for general LOP over cones: Self concordance barrier function in the interior of feasible region. When does it exist?
- Polynomial-time primal-dual interior point methods for LOP over **symmetric cones**. Practical and powerful.



# DOS SUBPROYECTOS:

**Periodo de ejecución:**

**1 Enero 2017 ---- 29 Junio 2021**

**EQUIPO US:**

| Equ. Investigación | Equipo Trabajo |
|---|---|
| Justo Puerto (IP1) | |
| Fernando López (IP2) | M.A. Pozo |
| V. Blanco | F. Ricca |
| M.J. Gago | A. Scozzari |
| M.I. Hartillo | J. Wesolowski |
| Y. Hinojosa | M.Leal |
| F. Perea Rojas-Marcos | D. Ponce |
| B. Salamanca | A. Japón |
| J.M.Ucha | M. Rodríguez-Madrena |
| E. Conde | C. Valverde |
| | F. Temprano |

**EQUIPO UCA:**

| Equ. Investigación | Equipo Trabajo |
|---|---|
| A.M. Rodríguez-Chia (IP) | S. Nickel |
| C. Valero | J. Kalcsics |
| A. Castaño | M.Baldomero |
| I. Espejo | L. Martínez |
| | J.M. Muñoz |
| | |
| | |

# PRINCIPALES LÍNEAS DE INVESTIGACIÓN

- LOCATION ANALYSIS:  11
- OPTIMIZATION WITH ORDER (DOMP): 8
- DATA ANALYSIS (SVM, CLUSTER, OUTLIERS, REGRESSION, FEATURE SELECTION,...): 13
- TRANSPORTATION: 8
- COOPERATIVE GAMES: 5
- MATHEMATICAL PROGRAMMING: 7
- PORFOLIO OPTIMIZATION: 3
- ORDER STATISTICS: 3
- COMBINATORIAL OPTIMIZATION (SPAN TREES, PARTITIONING, RANK AGGREG., REGRET ...): 4
- APPLICATIONS: 11
- SCHEDULING: 2
- FUZZY: 2
- REGRET: 3

# DATOS GLOBALES PUBLICACIONES:

| Categorias | | SEDES | | Trabajos US | |
|---|---|---|---|---|---|
| | | | | | |
| DOMP | 8 | UCA | 14 + 6 | | |
| Datos | 13 | US | 54 + 6 | US | 41 |
| Location | 11 | UCA/US | 6 | UGR | 7 |
| O.C. | 4 | | | UPV | 3 |
| Transporte | 8 | | | US/UGR | 7 |
| Math Prog | 7 | | | US/UPV | 2 |
| Applications | 9 | | | | |
| Games | 5 | | | | |
| Partitioning | 2 | Cuartiles | | | |
| Fuzzy | 2 | Q1 | 46 | | |
| Order stat | 3 | Q2 | 20 | | |
| Portfolio | 3 | Q3 | 8 | | |
| Scheduling | 2 | Q4 | 6 | | |
| Regret | 3 | | | | |
| | | | | | |
| Total | 80 | | | | |

# Location Analysis: 11 publicaciones

Math. Program., Ser. A (2017) 161:33–72
DOI 10.1007/s10107-016-1002-x

FULL LENGTH PAPER

## Continuous location under the effect of '*refraction*'

**Víctor Blanco**[1] · **Justo Puerto**[2,3] · **Diego Ponce**[2,3]

Received: 11 April 2014 / Accepted: 8 March 2016 / Published online: 30 April 2016
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2016

**Abstract** In this paper we address the problem of locating a new facility on a $d$-dimensional space when the distance measure ($\ell_p$- or polyhedral-norms) is different at each one of the sides of a given hyperplane $\mathcal{H}$. We relate this problem with the physical phenomenon of refraction, and extend it to any finite dimensional space and different distances at each one of the sides of any hyperplane. An application to this problem is the location of a facility within or outside an urban area where different distance measures must be used. We provide a new second order cone programming formulation, based on the $\ell_p$-norm representation given in Blanco et al. (Comput Optim Appl 58(3):563–595, 2014) that allows to solve the problem in any finite dimensional space with second order cone or semidefinite programming tools. We also extend the problem to the case where the hyperplane is considered as a rapid transit media (a different third norm is also considered over $\mathcal{H}$) that allows the demand to travel, whenever it is convenient, through $\mathcal{H}$ to reach the new facility. Extensive computational experiments run in Gurobi are reported in order to show the effectiveness of the approach. Some extensions of these models are also presented.

✉ Víctor Blanco
vblanco@ugr.es

Justo Puerto
puerto@us.es

Diego Ponce
dponce@us.es

[1] Department of Quantitative Methods for Economics and Business, Universidad de Granada, 18011 Granada, Spain

[2] Institute of Mathematics University of Seville (IMUS), Seville, Spain

[3] Department of Statistics and OR, Universidad de Sevilla, 41012 Seville, Spain



European Journal of Operational Research 262 (2017) 509–520

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

## The probabilistic $p$-center problem: Planning service for potential customers

Luisa I. Martínez-Merino[a], Maria Albareda-Sambola[b], Antonio M. Rodríguez-Chía[a,*]

[a] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Avda. República Saharaui, 11510 Puerto Real (Cádiz). Spain
[b] Departamento de Estadística e Investigación Operativa, Universitat Politècnica de Catalunya-BarcelonaTech, C. Colom, 11, 08222 Terrassa (Barcelona), Spain

ARTICLE INFO

*Article history:*
Received 7 April 2016
Accepted 15 March 2017
Available online 22 March 2017

*Keywords:*
Discrete location
p-center
Mixed integer linear formulations
Demand uncertainty

ABSTRACT

This work deals with the probabilistic p-center problem, which aims at minimizing the expected maximum distance between any site with demand and its center, considering that each site has demand with a specific probability. The problem is of interest when emergencies may occur at predefined sites with known probabilities. For this problem we propose and analyze different formulations as well as a Variable Neighborhood Search heuristic. Computational tests are reported, showing the potentials and limits of each formulation, the impact of their enhancements, and the effectiveness of the heuristic.

© 2017 Elsevier B.V. All rights reserved.

### 1. Introduction

Many discrete location models have been inspired by a variety of applications in logistics, telecommunications, emergency services, etc. The goal is to locate a number of facilities within a set of candidate sites and assign customers to them optimizing some effectiveness measure, usually depending on the assignment distances (see Laporte, Nickel, & Saldanha da Gama, 2015; Daskin, 1995; Drezner & Hamacher, 2002).

Among them, the $p$-center problem (*pCP*) aims at locating $p$ centers out of $n$ sites and assigning the remaining sites to the centers, so that the maximum distance between a site and its assigned center is minimized (see Calik et al., 2015). Although the *pCP* is NP-hard (Kariv & Hakimi, 1979), it can be solved efficiently via bisection search (see Daskin, 1995; 2000). Nonetheless, extensive literature exists proposing exact and heuristic algorithms for (Calik & Tansel, 2013; Irawan, Salhi, & Drezner, 2015). The main applications of the *pCP* are the location of emergency services like ambulances, hospitals or fire stations, since, in this context, the whole population should be timely reachable from some center. However, as already observed in the past, locating services according to the *pCP* may increase the effective service distances (see Ogryczak, 1997). This motivated alternative models, such as the cent-dian (Halpern, 1978).

This work presents a stochastic *pCP* variant (see, for instance, Snyder & Daskin, 2005; Berman, Krass & Menezes, 2007; O'Hanley, Scaparra, and García, 2013; Snyder, 2006). This variant aims at smoothing this loss of spatial efficiency, trying to keep the centers close to where they are needed. Namely, the probabilistic $p$-center problem (*PpCP*) aims at finding $p$ centers, out of $n$ sites, that minimize the expected maximum distance between a site with demand and its allocated center, assuming that demands can occur at each site independently, and with a known probability.

As stated above, considering the expected maximum service cost instead of the maximum assignment distance, prevents situations where a remote site with a low demand probability forces to place centers further from the remaining sites than it is desirable. In applications like firefighting, for instance, one pretends to provide service to a whole region but it wouldn't make sense to use a worst-case approach if the region contains areas with high risk of fire, and others where a fire is very unlikely to take place. Another real situation that fits to our model could be the case of locating several security offices to attend as fast as possible burglary alarms in different neighborhoods of a city. Obviously, by beforehand studies and previous experience, it would be possible to know that some neighborhoods are safer than others, and then an estimation of the probability that a theft takes place in a neighborhood can be computed. In this case, the probabilities that a theft occurs in two different neighborhoods are independent. In such a situation, the *PpCP* would be much more convenient than the classical *pCP*.

From the modeling point of view, the *PpCP* falls into the stochastic programming paradigm, where uncertain values are described through probability distributions (see, for instance, Albareda-Sambola, Fernández, & Saldanha da Gama, 2011; Huang, Kim, & Menezes, 2010) as opposite to the robust optimization approach, which attempts to optimize the worst-case system

* Corresponding author.
*E-mail address:* antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

http://dx.doi.org/10.1016/j.ejor.2017.03.043
0377-2217/© 2017 Elsevier B.V. All rights reserved.

European Journal of Operational Research 268 (2018) 432–449

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

## Multi-period stochastic covering location problems: Modeling framework and solution approach

Alfredo Marín[a], Luisa I. Martínez-Merino[b,*], Antonio M. Rodríguez-Chía[b], Francisco Saldanha-da-Gama[c,d]

[a] Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Murcia, Murcia, Spain
[b] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Cádiz, Spain
[c] Departamento de Estatística e Investigação Operacional, Faculdade de Ciências da Universidade de Lisboa, Portugal
[d] Centro de Matemática, Aplicações Fundamentais e Investigação Operacional, Faculdade de Ciências da Universidade de Lisboa, Portugal

ARTICLE INFO

*Article history:*
Received 6 October 2017
Accepted 18 January 2018
Available online 31 January 2018

*Keywords:*
Location
Covering
Discrete optimization
Multi-period facility location
Lagrangian relaxation

ABSTRACT

This paper introduces a very general discrete covering location model that accounts for uncertainty and time-dependent aspects. A MILP formulation is proposed for the problem. Afterwards, it is observed that most of the models existing in the literature related with covering location can be considered as particular cases of this formulation. In order to tackle large instances of this problem a Lagrangian relaxation based heuristic is developed. A computational study is addressed to check the potentials and limits of the formulation and some variants proposed for the problem, as well as to evaluate the heuristic. Finally, different measures to report the relevance of considering a multi-period stochastic setting are studied.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

A covering location problem consists of deciding where to locate a set of facilities/equipments in such a way that a set of demand points can be reached or served within some previously defined time or distance; in that case, a demand point is referred to as covered. When the set of potential locations for the facilities is discrete, we are facing a discrete covering location problem. This is the case we focus in this work. Some reviews of covering location problems can be found in Schilling, Jayaraman, and Barkhi (1993), Berman, Drezner, and Krass (2010), Snyder (2011), Farahani, Asgari, Heidari, Hosseininia, and Goh (2012), and García and Marín (2015).

Like for many other problems, the use of optimization techniques in the context of covering facility location calls for an *a priori* development of mathematical models. In this respect, two main modeling frameworks can be found in the literature depending on the objective of the problem: the set covering location problem (SCP) and the maximal covering location problem (MCLP). The former was initially introduced by Toregas, Swain, ReVelle, and Bergman (1971) and aims at minimizing the total cost associated with the installed facilities that guarantees the coverage of all demand points. The latter was initially proposed by Church and ReVelle (1974) and consists of locating a certain number of facilities maximizing the resulting coverage.

One interesting and relevant feature of covering location problems is that they can be applied to different areas. Such applications often regard the location of emergency or surveillance facilities such as ambulances, fire fighting bases, or surveillance points. In such a case, by covering the "demand points" with a set of facilities we are ensuring that a certain response level is guaranteed upon the occurrence of an event. The SCP has been applied to deployment of emergency services (Toregas et al., 1971) and to the analysis of markets (Storbeck, 1988). On the other hand, the MCLP has been applied, for instance, to cluster and discriminant analysis (Chung, 1986). Overall, the flexibility provided by covering location problems explains why they have been deeply studied. For further details, the reader can refer to the reviews by Schilling et al. (1993) and Caprara, Toth, and Fischetti (2000) and to the references therein.

In this paper we aim at bringing further the comprehensiveness of the optimization models in use when it comes to discrete covering location problems. With this purpose, we extend the general modeling framework proposed by García and Marín (2015) in order to capture two features (alone or combined) of practical relevance: time-dependency of aspects such as demand and costs in addition to the uncertainty that often underlies the parameters of the problem. It is worth noticing that the model proposed by

* Corresponding author.
*E-mail addresses:* amarin@um.es (A. Marín), luisa.martinez@uca.es (L.I. Martínez-Merino), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía), fsgama@ciencias.ulisboa.pt (F. Saldanha-da-Gama).

https://doi.org/10.1016/j.ejor.2018.01.040
0377-2217/© 2018 Elsevier B.V. All rights reserved.

TOP (2018) 26:187–226
https://doi.org/10.1007/s11750-018-0476-5

INVITED PAPER

# Extensive facility location problems on networks: an updated review

**Justo Puerto**[1] · **Federica Ricca**[2] · **Andrea Scozzari**[3]

Published online: 9 April 2018
© Sociedad de Estadística e Investigación Operativa 2018

**Abstract** Location problems with extensive facilities represent a challenging field of research. According to the specialized literature, a facility is called extensive if, for purposes of location, it is too large in relation to its environment to be considered a point. There are many examples of this type of structures that appear in real-world applications both in the continuous space (straight lines, circles, strips) and in networks (paths, cycles, trees). There exists a recent literature review on the location of dimensional facilities on continuous space (Díaz-Báñez et al. in TOP 154:22–44, 2004; Schöbel in Location of dimensional facilities in a continuous space, 2015) that does not cover similar problems on networks. The goal of this paper is to review the location of dimensional facilities in networks. We mainly concentrate on the location of paths and trees considering the most common objective functions in the location literature, namely median and center. However, we also consider some other alterna-

This invited paper is discussed in the comments available at https://doi.org/10.1007/s11750-018-0473-8, https://doi.org/10.1007/s11750-018-0474-7, and https://doi.org/10.1007/s11750-018-0475-6.

This research has been partially supported by Spanish Ministry of Economía and Competitividad/FEDER grant number MTM2016-74983-C02-01.

✉ Justo Puerto
puerto@us.es

Federica Ricca
federica.ricca@uniroma1.it

Andrea Scozzari
andrea.scozzari@unicusano.it

1 IMUS and Dept. Estadística e Investigación Operativa, Universidad de Sevilla, Seville, Spain

2 Università di Roma, La Sapienza, Rome, Italy

3 Università degli Studi Niccolò Cusano, Rome, Italy



TOP (2018) 26:236–238
https://doi.org/10.1007/s11750-018-0477-4

DISCUSSION

# Rejoinder on: Extensive facility location problems on networks: an updated review

**Justo Puerto**[1] · **Federica Ricca**[2] · **Andrea Scozzari**[3]

Published online: 9 April 2018
© Sociedad de Estadística e Investigación Operativa 2018

We are grateful to the discussants of this survey for providing thoughtful comments and interesting suggestions. We are pleased to note that all of them found the survey complete, well structured, and useful to provide a clear overview on the topic of location of extensive facilities on graphs. They also recognized the potential of this article to initiate and encourage further research in this area.

First of all, we thank the discussants for their careful reading that led to fix some technicalities and missing details that improved the quality of the original manuscript. On top of that, we would like to comment on some specific remarks and suggestions raised in the discussion.

Wang pointed out some improved complexity results related to the $p$-core problem, $p \geq 2$ Novik (1996), the $k$-tree center problem and the minmax regret path location problem that have been included in the revised version of the paper.

This research has been partially supported by Spanish Ministry of Economía and Competitividad/FEDER Grant Number MTM2016-74983-C02-01.
This rejoinder refers to the comments available at https://doi.org/10.1007/s11750-018-0473-8, https://doi.org/10.1007/s11750-018-0474-7, and https://doi.org/10.1007/s11750-018-0475-6.

✉ Justo Puerto
puerto@us.es

Federica Ricca
federica.ricca@uniroma1.it

Andrea Scozzari
andrea.scozzari@unicusano.it

1 IMUS and Departamento Estadística e Investigación Operativa, Universidad de Sevilla, Seville, Spain

2 Università di Roma, La Sapienza, Rome, Italy

3 Università degli Studi Niccolò Cusano, Rome, Italy



Computers and Operations Research 107 (2019) 95–106

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# New models for the location of *controversial* facilities: A bilevel programming approach

Martine Labbé[a], Marina Leal[b,*], Justo Puerto[b]

[a] Université Libre de Bruxelles, Belgium
[b] IMUS, Universidad de Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 15 November 2018
Revised 1 March 2019
Accepted 11 March 2019
Available online 13 March 2019

*Keywords:*
Bilevel optimization
Locational analysis
Combinatorial optimization

ABSTRACT

Motivated by recent real-life applications in Location Theory in which the location decisions generate *controversy*, we propose a novel bilevel location model in which, on the one hand, there is a leader that chooses among a number of fixed potential locations which ones to establish. Next, on the second hand, there is one or several followers that, once the leader location facilities have been set, chooses his location points in a continuous framework. The leader's goal is to maximize some proxy to the weighted distance to the follower's location points, while the follower(s) aim is to locate his location points as close as possible to the leader ones. We develop the bilevel location model for one follower and for any polyhedral distance, and we extend it for several followers and any $\ell_p$-norm, $p \in \mathbb{Q}$, $p \geq 1$. We prove the NP-hardness of the problem and propose different mixed integer linear programming formulations. Moreover, we develop alternative Benders decomposition algorithms for the problem. Finally, we report some computational results comparing the formulations and the Benders decompositions on a set of instances.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Location is a research area devoted to the optimal placement of facilities (Albareda-Sambola et al., 2015; Boonmee et al., 2017; Hammad et al., 2018; Kalcsics et al., 2014; Labbé et al., 1995; Nickel and Puerto, 2006; Owen and Daskin, 1998; Puerto and Rodríguez-Chía, 2015), including among many others emergency systems (Bélanger et al., 2019; Boonmee et al., 2017; Schmid and Doerner, 2010), service providers (Albareda-Sambola et al., 2009; Berman et al., 2010), infrastructures, etc., and it is a basic building block of most transportation, communication or logistic problems. Moreover, these models also fit perfectly to positioning problems, as for instance of firms or products, in spaces of attributes or features where the dimension is much higher than in standard location problems on the plane or the tri-dimensional space. An optimal location can be chosen according to different criteria depending on the rationale behind the considered model. The most popular ones are the minimization of the total or maximum transportation cost (Albareda-Sambola et al., 2015; Boonmee et al., 2017), the maximization of some coverage goal (Albareda-Sambola et al., 2009; Bélanger et al., 2019; Berman et al., 2010; Boonmee et al., 2017), or the minimization of the undesirable effects induced by the facilities (Brimberg and Juel, 1998; Church and Scaparra, 2007; Hammad et al., 2018).

Location Theory includes a number of real-life applications in which the location decisions generate *controversy*. This controversy must be understood as a disagreement among users with different, non-aligned or opposite interests. Examples of this controversial location can be found in the literature, for example, in the areas of semi-obnoxious facility location or in problems that involve the location and protection of critical infrastructures, the positioning of products in attribute spaces or facilities sensitive to intentional attacks.

In the last decades, the consciousness-raising in environmental issues has grown substantially, specially in those aspects that affect human health or have adverse effects on people. As a consequence of this awareness-raising, the location of (semi-) obnoxious facilities has been extensively studied. Obnoxious facilities are those that generate a disservice to the people nearby while producing an intended product or service (Erkut and Neuman, 1989). However, if only these undesirable effects are taken into account when locating them, these facilities would never be opened or would be located too far from the population centers making use of the produced services, thus generating huge costs. For that reason, in the last years, there has been an increasing focus in analyzing the problem of locating semi-obnoxious facilities (Brimberg and Juel, 1998; Hammad et al., 2018; Melachrinoudis and Xanthopulos, 2003). Semi-obnoxious facilities has been defined as

* Corresponding author.

https://doi.org/10.1016/j.cor.2019.03.003
0305-0548/© 2019 Elsevier Ltd. All rights reserved.

Computers and Operations Research 108 (2019) 213–225

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# The stratified *p*-center problem

Maria Albareda-Sambola[a], Luisa I. Martínez-Merino[b,*], Antonio M. Rodríguez-Chía[b]

[a] Departamento de Estadística e Investigación Operativa, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain
[b] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain

ARTICLE INFO

*Article history:*
Received 27 July 2018
Revised 23 January 2019
Accepted 10 April 2019
Available online 13 April 2019

*Keywords:*
Location
*p*-Center
Discrete optimization
Sample average approximation

ABSTRACT

This work presents an extension of the discrete *p*-center problem. In this new model, called Stratified *p*-Center Problem (*Sp*CP), the demand is concentrated in a set of sites and the population of these sites is divided into different strata depending on the kind of service that they require. The aim is to locate *p* centers to cover the different types of services demanded minimizing the weighted average of the largest distances associated with each of the different strata. In addition, it is considered that more than one stratum can be present at each site. Different formulations, valid inequalities and preprocessings are developed and compared for this problem. An application of this model is presented in order to implement a heuristic approach based on the Sample Average Approximation method (SAA) for solving the probabilistic *p*-center problem in an efficient way.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Discrete location problems have been widely studied since the seminal paper Balinski (1965), where the first MILP formulation for such a problem was proposed. Among the fundamental problems in this area, the *p*-Center Problem (*p*CP) aims at selecting, from *n* given sites, the locations of *p* service centers that minimize the maximum distance between any of the sites and its closest service center. This model, in contraposition to the *p*-median problem, was motivated by the need not to discriminate spatially dispersed clients when locating essential or emergency centers (see Calik et al., 2015; Garfinkel et al., 1977, for more details).

Both, continuous and discrete versions have already been addressed by several authors. Examples of works concerning the continuous version are Callaghan et al. (2017) and Elshaikh et al. (2016) where the planar version is analyzed. This paper focuses on an extension of the discrete *p*CP. The discrete *p*CP, also known as vertex *p*CP, has been proven to be NP-hard (Kariv and Hakimi, 1979). However, many efficient exact and heuristic algorithms have been introduced for this problem. See for instance, Contardo et al. (2019), Calik and Tansel (2013) or Irawan et al. (2016).

In the last decades, several extensions of the discrete *p*CP have been introduced in the literature. These include variants considering capacities (Albareda-Sambola et al., 2010; Özsoy and Pınar, 2006; Quevedo-Orozco and Rios-Mercado, 2015) or pre-existing centers, as in the conditional *p*-center problem (Drezner, 1989). Other extensions, such as the $\alpha$-neighbor *p*-center problem, consider the largest distance of a demand point to its $\alpha$ closest centers, see Chen and Chen (2013).

In addition, the *p*CP with uncertain parameters has been addressed both, from the perspective of robust optimization (Averbakh and Berman, 1997; Lu and Sheu, 2013) and of stochastic programming (Espejo et al., 2015; Martínez-Merino et al., 2017; Revelle and Hogan, 1989). In particular, in Martínez-Merino et al. (2017) the probabilistic *p*-center problem (*Pp*CP) is introduced. In this problem, the goal is to minimize the expected largest distance between any demand point and its corresponding center. The *Pp*CP considers that the demand can occur independently at each demand site with a certain known probability. Observe, that the *p*CP is a particular case of this problem where all sites have demand with probability one. In this paper, we will introduce a heuristic method for the *Pp*CP making use of the formulations of the *p*CP extension that will be proposed.

A common characteristic of most of the considered problem variants is that customers are assumed to be homogeneous in the sense that they are all considered in the same way in the objective function. The only exception would be the weighted *p*CP, where the distances between each site and its closest center are affected by site-dependent weights. See, for instance, Jeger and Kariv (1985) where the particular case of this problem defined on trees is addressed.

In this paper, we consider situations where, for instance, the population of a region is divided into different strata, and people of different strata can live together in the same cities. The goal of the

* Corresponding author.
*E-mail address:* luisa.martinez@uca.es (L.I. Martínez-Merino).

https://doi.org/10.1016/j.cor.2019.04.013
0305-0548/© 2019 Elsevier Ltd. All rights reserved.

European Journal of Operational Research 289 (2021) 44–58

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# Covering problems with polyellipsoids: A location analysis perspective☆

Víctor Blanco[a,*], Justo Puerto[b]

[a] IEMath-GR, Universidad de Granada, Spain
[b] IMUS, Universidad de Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 3 April 2020
Accepted 30 June 2020
Available online 3 July 2020

*Keywords:*
Polyellipsoids
Covering location
Minimum enclosing disk
Second order cone programming

ABSTRACT

In this paper we analyze the extension of the classical smallest enclosing disk problem to the case of the location of a polyellipsoid to fully cover a set of demand points in $\mathbb{R}^d$. We prove that the problem is polynomially solvable in fixed dimension and analyze mathematical programming formulations for it. We also consider some geometric approaches for the problem in case the foci of the polyellipsoids are known. Extensions of the classical algorithm by Elzinga-Hearn are also derived for this new problem. Moreover, we address two extensions of the problem, as the case where the foci of the enclosing polyellipsoid are not given and have to be determined among a potential set of points or the induced covering problems when instead of polyellipsoids, one uses ordered median polyellipsoids. For these problems we also present Mixed Integer (Non) Linear Programming strategies that lead to efficient ways to solve it. Extensive computational experiments on different datasets show the usefulness of our solution methods.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Given a set of demand points in a finite dimensional normed space $\mathbb{R}^d$. Continuous Facility Location Problems (CFLP) deal with the determination of the optimal placement of some new points in order to minimize certain measure of the distances from the given to these new points. The most popular CFLP is the Weber Problem (Weber, 1929) in which a single facility is to be located minimizing the overall (weighted) sum of the Euclidean distances from the demand points to the facility. A common approach to solve the Weber Problem is via the Weiszfeld algorithm (Weiszfeld, 1937). The extension to multiple facilities may either consider that the overall sum of the distances from the demand points to all facilities (multiple-allocation case) or to their closest facilities (single-allocation case) has to be minimized. Different *aggregation* measures, apart from the overall sum, have been also proposed to find optimal facilities for CFLP (see e.g., Blanco, ElHaj-Ben Ali, & Puerto, 2013).

The level curves of the sum of Euclidean distances to given points on the plane (the Euclidean planar Weber level curve), in case the ($n$) demand points (also known as foci within this framework) are not collinear, are called in the literature *polyellipses*, *n-ellipses* or *multifocal ellipses*. The regions bounded by those curves are known to be convex bodies and can also be defined in higher dimensional spaces and with different norm measures, inducing the notion of *polyellipsoid*. Several authors have been interested in the analysis of these convex bodies, from Maxwell (1851) to Erdös and Vincze (1982), mainly from an algebraic or geometric point of view (see Nie, Parrilo, & Sturmfels, 2008). In particular, the problem whether all the convex plane curves can be arbitrarily approximated by polyellipses under a sufficiently large number of points (also known as *foci*) was posed by Weiszfeld. In this sense, the Erdös-Vincze's theorem states that regular triangles cannot be presented as the Hausdorff limit of polyellipses even if the number of the foci can be arbitrary large. Recently, in Vincze, Kovács, and Csorvássy (2017), the authors extend the analysis for regular polygons in the plane. However, these sets (polyellipsoids) have only been partially analyzed under optimization lens beyond their connections with the Weber problem (see, e.g. Melzak & Forsyth (1977)).

On the other hand, Covering Location Problems consist of locating facilities to satisfy the demand of a set of customers whenever they are only allowed to be served within certain coverage areas. In the continuous case there are two main families of problems in this field: Full Covering Location problems (FCP) and Maximal Covering Location problems (MCP). In the former, all the demand points have to be covered and the goal is to minimize the set up

☆ This research has been partially supported by Spanish Ministry of Education and Science/FEDER grant number MTM2016-74983-C02-(01–02), and projects FEDER-US-1256951, CEI-3-FQM331 and *NetmeetData*: Ayudas Fundación BBVA a equipos de investigación científica 2019.
* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.ejor.2020.06.048
0377-2217/© 2020 Elsevier B.V. All rights reserved.

Computers and Operations Research 102 (2019) 75–90

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# Upgrading nodes in tree-shaped hub location

Víctor Blanco[a,b,*], Alfredo Marín[c]

[a] IEMath-GR, Universidad de Granada, Spain
[b] Department of Quantitative Methods for Economics & Business, Universidad de Granada, Spain
[c] Department of Statistics & Operational Research, Universidad de Murcia, Spain

ARTICLE INFO

*Article history:*
Received 15 February 2018
Revised 4 October 2018
Accepted 6 October 2018
Available online 9 October 2018

*MSC:*
90B80
90C11
05C05

*Keywords:*
Hub location
Spanning tree
Upgrading

ABSTRACT

Upgrading a node in a graph means to improve the quality of the edges that have it as an extreme. The first problem where upgrading was considered is the Minimum Cost Spanning Tree Problem. There, upgrading one or two of the extremes of an edge reduces progressively its cost and makes the edge more likely to be in the optimal tree. In this paper we analyze upgradings of nodes in a discrete location problem, the so-called Tree of Hubs Location Problem. Hubs are transshipment points through which a product is sent from origins to destinations. In addition to locate the hubs, to determine a tree that connects the hubs and to allocate non-hub nodes to hubs, i.e., the decisions to be made in the Tree of Hubs Location Problem, a decision has to be made about which of the hubs will be upgraded to reduce the overall cost of the distribution network. We provide two different Mixed Integer Linear Programming formulations for the problem, tighten the formulations and generate several families of valid inequalities for them. A computational study is presented showing the improvements attained with the strengthening of the formulations and comparing them.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

The main goal of Discrete Facility Location is to decide which set of facilities, among a finite set of potential sites, must be opened in order to optimize a given objective function which represents traveling costs of a set of given customers and/or set-up costs. Reaching such a goal is closely related with Mixed Integer Linear Programming (MILP) models. According to Laporte et al. (2015), the first MILP formulation for a discrete location problem, the *Uncapacitated Facility Location Problem*, was proposed by Balinski (1965). In the following few years, MILP formulations appeared also for the discrete *p*-Median Problem (ReVelle and Swain, 1970) and for a covering-location problem (Toregas et al., 1971). Since then, Discrete Facility Location has gained importance over the years as a fundamental part of both Integer Programming and Location Science.

Among the wide family of Discrete Facility Location problems, hub location problems have taken on significance in the last decades. Hub Location Problems (HLP) emerge as an important design tool in transportation systems, where several sites must send and receive some goods through some transshipment points, *the hubs*. At each of the hubs, the product is collected from different sources and distributed either by sending it to its final destination or to another hub. Using this system, clear cost savings are obtained compared to the direct *origin-to-destination* distribution network, because of economies of scale. Note that the amount of possible routes is considerably reduced, each of them traversed by a large amount of product. In this sense, a common characteristic of HLP is that the unit transportation cost is reduced if the product is routed between two hubs. The interest of hub location is still increasing, and one can read a number of surveys on the matter published over the last years (see Alumur and Kara, 2008; Campbell et al., 2002; Campbell and O'Kelly, 2012; Contreras, 2015; Farahani et al., 2013, among many others).

Hub location problems are usually classified according to the way of allocating non-hub customers (also called *spokes*) to hub nodes. On the one hand, *multiple–allocation* hub location problems assume that the flow with origin a single node may be distributed through any of the available hub nodes. Thus, depending on the destination, the flow may be initially routed by a different hub node. The first formulation for this family of problems was presented in Campbell (1994). On the other hand, in *single–allocation* models, the customers are only allowed to send and receive the service from a unique hub, i.e., each origin/destination is allocated to exactly one hub. Our study fits within this category, to which the pioneering works (Aykin, 1995; Ernst and Krishnamoorthy, 1996; 1998; O'Kelly, 1987; O'Kelly et al., 1996) and (Skorin-Kapov et al., 1996) belong to.

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), amarin@um.es (A. Marín).

https://doi.org/10.1016/j.cor.2018.10.007
0305-0548/© 2018 Elsevier Ltd. All rights reserved.

Computers & Operations Research 132 (2021) 105310

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# Continuous maximal covering location problems with interconnected facilities

Víctor Blanco *, Ricardo Gázquez

*Institute of Mathematics, Universidad de Granada, Spain*
*Dpt. Quantitative Methods for Economics & Business, Universidad de Granada, Spain*

ARTICLE INFO

*MSC:*
90B85
90C10
90C90
90C26

*Keywords:*
Maximal covering location
Continuous location
Mixed Integer Non Linear Programming
Integer linear programming
Branch-&-cut approaches

ABSTRACT

In this paper we analyze a continuous version of the maximal covering location problem, in which the facilities are required to be linked by means of a given graph structure (provided that two facilities are allowed to be linked if a given distance is not exceed). We propose a mathematical programming framework for the problem and different resolution strategies. First, we provide a Mixed Integer Non Linear Programming formulation for the problem and derive some geometrical properties that allow us to reformulate it as an equivalent pure integer linear programming problem. We propose two branch-&-cut approaches by relaxing some sets of constraints of the former formulation. We also develop a math-heuristic algorithm for the problem capable to solve instances of larger sizes. We report the results of an extensive battery of computational experiments comparing the performance of the different approaches.

## 1. Introduction

The goal of Covering Location problems is to find the position of a set of facilities to give service to a finite set of users provided that the demand of each user is served just in case a facility is located within its coverage area. The most popular covering location problems are the *p*-center (Hakimi, 1964) and the maximal covering location (Church and ReVelle, 1974) problems. In particular, the Maximal Covering Location Problem (MCLP) consists of locating a fixed number of facilities in order to maximize the covered demand of a set of users. This problem has proven to be a seminal contribution in logistics and network design, in terms of both technical merit and practical interest. The MCLP is known to be useful in the geographic location of warehouses, health-care centers, fire stations, etc. (Church and ReVelle, 1974) and also in other non-logistic problems, as in data abstraction, statistical classification, selection of stocks or grouping problems (Chung, 1986). The interested readers are referred to the recent book by Church and Murray (2018) or the chapter by García-Quiles and Marín (2020) for further details on covering location problems.

Covering Location problems have been analyzed in both the discrete and the continuous frameworks. In the discrete case, a finite set of potential facilities is provided, and the goal is to choose, among them, the optimal ones. These problems are useful in the location of physical services (ATMs, stores, hospitals, etc.) in which a finite set of potential locations can be previously determined by the decision maker. However, in telecommunication networks, the positions of routers, alarms or sensors, can be more flexibly positioned if the facilities are allowed to be located in the whole space, being the continuous framework more adequate. This framework is also useful to determine the set of potential facilities that serves as input for a discrete version of the problem. The main difference between these two families of problems is that in the discrete case the costs/distances between the facilities and the users are given as input data (or can be preprocessed before solving the problem) while in the continuous case, the costs/distances are part of the decision and they have to be incorporated to the decision problem.

In many practical situations in Location Analysis, the facilities to be located are required to be interconnected. This is the case of the optimal design of forest fire-fighters centers that must be communicated to a central server at a give radius (Demaine et al., 2009) or in the location of sensors that have to be connected to each others (Romich et al., 2015). Some facility location problems have been analyzed with some kind of interconnection between facilities (see, e.g., Blanco et al., 2016), although interconnection between services has been mostly studied in the context of hub location, in which the routing costs induce such a connectivity between the hub nodes (see Contreras and O'Kelly, 2019 and the references therein). Another possibility is to consider that an interconnection graph for the centers has to be constructed

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), rgazquez@ugr.es (R. Gázquez).

https://doi.org/10.1016/j.cor.2021.105310
Received 6 May 2020; Received in revised form 23 March 2021; Accepted 24 March 2021
Available online 31 March 2021
0305-0548/© 2021 Elsevier Ltd. All rights reserved.

Computers and Operations Research 130 (2021) 105181

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/caor

# Minmax regret maximal covering location problems with edge demands

Marta Baldomero-Naranjo [a,*], Jörg Kalcsics [b], Antonio M. Rodríguez-Chía [a]

[a] *Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Cádiz, Spain*
[b] *School of Mathematics, The University of Edinburgh, Edinburgh, United Kingdom*

ARTICLE INFO

*Article history:*
Received 30 October 2019
Revised 7 November 2020
Accepted 8 December 2020
Available online 11 December 2020

*Keywords:*
Location
Covering
Minmax regret
Networks
Robust optimization

ABSTRACT

This paper addresses a version of the single-facility Maximal Covering Location Problem on a network where the demand is: (i) distributed along the edges and (ii) uncertain with only a known interval estimation. To deal with this problem, we propose a minmax regret model where the service facility can be located anywhere along the network. This problem is called Minmax Regret Maximal Covering Location Problem with demand distributed along the edges (MMR-EMCLP). Furthermore, we present two polynomial algorithms for finding the location that minimises the maximal regret assuming that the demand realisation is an unknown constant or linear function on each edge. We also include two illustrative examples as well as a computational study for the unknown constant demand case to illustrate the potential and limits of the proposed methodology.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

In this work, we focus our research on the Maximum Covering Location Problem (MCLP), first introduced by Church and ReVelle (1974). For a given set of demand points (clients or customers), the goal of this problem is to locate one or more service facilities in such a way that the demand of covered clients is maximized. A client is considered to be covered if the distance to an open facility is smaller than or equal to a fixed coverage radius, *R*. We can find many different variants of this problem in the literature, some of them in continuous space and others on networks or in discrete spaces (see Plastria, 2002; García and Marín, 2015; Berman and Krass, 2002 for a detailed survey of each case). In the past, almost all of the models discussed in the literature assume that the demand is represented by a finite set of points. However, recently, an increasing body of literature in facility location addresses problems where the demand is continuously distributed over a region (polygon, straight lines, edges, etc.), (see Murray et al., 2007 Murray et al., 2010).

In the case of network problems (the subject of study of this paper), there are several applications where the assumption that the demand only occurs at the nodes of the network is not realistic; e.g. for the location of emergency facilities (see Li et al., 2011), for the planning of garbage collection, meter reading, or mail delivery, or situations where the coverage areas are extremely distance-dependent such as the location of Automated External Defibrillators (AED), bus stops, Automated Teller Machines (ATM), etc. Similarly, assuming the situation that the service facilities can only be located at the nodes of the network, the so-called node-restricted version, also leads to unrealistic models for some situations (see ReVelle et al., 1976; Church and Meadows, 1979). For example, in the case of locating bus stops, this assumption would mean that they can only be placed at the corner of streets. Thus, assuming that the demand is concentrated at nodes and/or service facilities can only be located at nodes may yield models that are far from the real situation under study, providing solutions that are of little use, see Current et al. (1990) for further details.

In the following, we review the main models in the literature related to our problem. The MCLP where the demand is assumed to exist only at nodes is solved in Church and ReVelle (1974). They proposed an integer programming formulation to solve the multi-facility node-restricted problem. García and Marín (2015) and Marín et al. (2018) provided an overview of models for the discrete covering location problem. Recently, Cordeau et al. (2019) proposed an effective decomposition approach for the MCLP based on a branch-and-Benders-cut formulation to solve realistic cases where the number of customers is much larger than the number of potential facility locations. To solve the continuous version of the problem, i.e., facilities can be located not only at the nodes but also along the edges, Church and Meadows (1979) proposed a Finite Dominating Set (FDS). An FDS for a location problem is a finite set of points that is guaranteed to contain an optimal solution for the problem. If such a set can be found, then the continuous optimization problem can equivalently be formulated as a

* Corresponding author.
*E-mail addresses:* marta.baldomero@uca.es (M. Baldomero-Naranjo), joerg.kalcsics@ed.ac.uk (J. Kalcsics), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.cor.2020.105181
0305-0548/© 2020 Elsevier Ltd. All rights reserved.

# DOMP: 8 publicaciones

European Journal of Operational Research 260 (2017) 886–903

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

## Ordered Weighted Average optimization in Multiobjective Spanning Tree Problem

Elena Fernández[a], Miguel A. Pozo[b], Justo Puerto[b,*], Andrea Scozzari[c]

[a] Department of Statistics and Operational Research, Universitat Politècnica de Catalunya, Barcelona Tech. Jordi Girona 1-3, Barcelona 08034, Spain
[b] IMUS and Faculty of Mathematics, Universidad de Sevilla, Seville 41012, Spain
[c] Facoltà di Economia, Università degli Studi Niccolò Cusano, Roma - 00166 Roma, Italy

ARTICLE INFO

*Article history:*
Received 2 July 2015
Accepted 12 October 2016
Available online 18 October 2016

*Keywords:*
Combinatorial optimization
Multiobjective optimization
Ordered median
Ordered Weighted Average
Spanning trees

ABSTRACT

Multiobjective Spanning Tree Problems are studied in this paper. The ordered median objective function is used as an averaging operator to aggregate the vector of objective values of feasible solutions. This leads to the Ordered Weighted Average Spanning Tree Problem, a nonlinear combinatorial optimization problem. Different mixed integer linear programs are proposed, based on the most relevant minimum cost spanning tree models in the literature. These formulations are analyzed and several enhancements presented. Their empirical performance is tested over a set of randomly generated benchmark instances. The results of the computational experiments show that the choice of an appropriate formulation allows to solve larger instances with more objectives than those previously solved in the literature.

© 2016 Elsevier B.V. All rights reserved.

### 1. Introduction

Optimization problems related to spanning trees, or simply Spanning Tree Problems are among the core problems in combinatorial optimization. On the one hand, the combinatorial object that represents spanning trees has important structural properties. On the other hand, from a practitioner point of view, spanning trees are found in a wide range of applications in many fields (e.g. computer networks design, telecommunications networks, transportation, etc.). Furthermore, they often appear as subproblems of other more complex optimization problems.

The most relevant property of trees is their matroid structure. This implies that the basic problem of finding a minimum cost spanning tree, can be solved efficiently (Kruskal, 1956; Prim, 1957). This also implies that formulations with the integrality property can be obtained, which allow to solve the minimum cost Spanning Tree Problem (STP) with linear programing tools. However, these good features can be lost for several reasons. For instance, when the objective function does not preserve Gale optimality, i.e., it is not monotonic on the edges costs (Fernández, Pozo, & Puerto, 2014; Lawler, 1966), as it happens in the Optimum Communication Spanning Tree Problem (Hu, 1974). The reader may refer to Landete and Marín (2014) for a description of alternative objective functions for STP. An optimization STP also becomes a hard problem when several objectives are considered simultaneously (Ehrgott, 2005). In such cases, no efficient combinatorial algorithm is known so the choice of an appropriate mathematical programing representation of the combinatorial object may become crucial. In this sense, formulations for STP with good properties can be outperformed by other formulations in the new environment.

From a different point of view, in the multiobjective case, it is widely accepted that the use of order and aggregation functions may yield compromise solutions for the different criteria. The literature includes many works on this area related to combinatorial optimization. Some examples, among many others, include minimax problems (Hansen, 1979; Schrijver, 1983), combining minisum and minimax (Averbakh & Berman, 1995; Hansen & Labbé, 1988; Hansen, Labbé, & Thisse, 1991; Minoux, 1989; Punnen, Nair, & Aneja, 1995; Tamir, Puerto, & Perez-Brito, 2002), $k$-centrum optimization (Garfinkel, Fernández, & Lowe, 2006; Kalcsics, Nickel, Puerto, & Tamir, 2002; Punnen, 1992; Slater, 1978a; 1978b; Tamir, 2000), lexicographic optimization (Calvete & Mateo, 1998; Croce, Paschos, & Tsoukias, 1999), $k$th best solutions (Lawler, 1972; Martello, Pulleyblank, Toth, & de Werra, 1984; Pascoal, Captivo, & Clímaco, 2003; Yen, 1971), most uniform solutions (Galil & Schieber, 1998; López de los Mozos, Mesa, & Puerto, 2008), minimum-envy solutions (Espejo, Marín, Puerto, & Rodríguez-Chía, 2009), solutions with minimum deviation (Gupta, Gupta, & Bector, 1990), regret solutions (Averbakh, 2001; Conde, 2004; Puerto & Rodríguez-Chía, 2003), equity measures (Gupta & Punnen, 1988; Mesa, Puerto, & Tamir, 2003; López de los Mozos et al., 2008; then it performs the scalar product of this so obtained sorted cost vector and the given vector of weights. This objective function has been widely applied in the field of location analysis and distribution models [10,9,17,18]. In addition, it has the potential to yield new models for order statistics embedded within mathematical programming formulations; thus enlarging the applications of optimization tools to the resolution of statistical problems in data analysis.

DOMP is known to be $\mathcal{NP}$ – complete, see Nickel and Puerto [14].

The first formulation of DOMP, proposed by Nickel [13], consists in an integer nonlinear problem. Then, Boland et al. [3] propose several linearizations of Nickel's model.

Instances with up to 30 clients could be solved to optimality by Boland et al. [3]. Further, if clients and facility locations coincide and if the allocation cost of a client to itself is equal to zero (the so-called free self-service), then instances with up to 100 clients could be solved by Marín et al. [10,11]. We observe that in all previously considered formulations the gaps with respect to the linear programming relaxations of those models are rather large, as mentioned in all those papers.

In this paper, we propose new formulations for DOMP and we develop a theoretical comparison of the lower bounds obtained from their LP-relaxations and show that our new formulations are rather tight. Our theoretical results also attempt to shed some light on the polyhedral structure of the new formulation based on scheduling constraints. We conclude with extensive computational

* Corresponding author. Fax: +34 954622800.
*E-mail addresses:* e.fernandez@upc.edu (E. Fernández), miguelpozo@us.es (M.A. Pozo), puerto@us.es, justo.puerto@gmail.com (J. Puerto), andrea.scozzari@unicusano.it (A. Scozzari).

http://dx.doi.org/10.1016/j.ejor.2016.10.016
0377-2217/© 2016 Elsevier B.V. All rights reserved.

Computers & Operations Research 78 (2017) 230–242

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

## A comparative study of formulations and solution methods for the discrete ordered $p$-median problem

Martine Labbé[a], Diego Ponce[b,c], Justo Puerto[b,c,*]

[a] Computer Science Department, Université Libre de Bruxelles, Belgium
[b] IMUS, Instituto de Matemáticas, Universidad de Sevilla, Spain
[c] Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 28 July 2015
Received in revised form
30 May 2016
Accepted 6 June 2016
Available online 8 June 2016

*Keywords:*
Discrete multifacility location
MIP formulations
Ordered median problem

ABSTRACT

This paper presents several new formulations for the Discrete Ordered Median Problem (DOMP) based on its similarity with some scheduling problems. Some of the new formulations present a considerably smaller number of constraints to define the problem with respect to some previously known formulations. Furthermore, the lower bounds provided by their linear relaxations improve the ones obtained with previous formulations in the literature even when strengthening is not applied. We also present a polyhedral study of the assignment polytope of our tightest formulation showing its proximity to the convex hull of the integer solutions of the problem. Several resolution approaches, among which we mention a branch and cut algorithm, are compared. Extensive computational results on two families of instances, namely randomly generated and from Beasley's OR-library, show the power of our methods for solving DOMP.

© 2016 Elsevier Ltd. All rights reserved.

### 1. Introduction

Recognizing the need for more flexible logistic models, Nickel [13] proposed the Discrete Ordered Median Location Problem (DOMP) which could be used to model different locations problems, as the $p$-median or the $p$-center. It is a flexible formulation based on applying an ordered weighted averaging operator to the costs as they appear in the solution and taking them into account with a suitable $n$-vector $\lambda$.

Given a vector of weights, the ordered weighted average of $n$ real numbers is obtained by first ranking those numbers by non-decreasing order and then computing the scalar product of the ranked allocation cost vector and the weight vector, see e.g. Nickel and Puerto [14].

Consider a set of clients and a set of candidate locations where some facility can be established. Further we are given the costs for allocating clients to facilities. DOMP consists in choosing $p$ facility locations and assigning each client to a facility with smallest allocation cost in order to minimize a special objective function, the so-called ordered median function. Given a vector of weights, this function consists in an ordered weighted average of the allocation costs, namely it sorts these costs in non-decreasing sequence and

* Corresponding author at: IMUS, Instituto de Matemáticas, Universidad de Sevilla, Spain.
*E-mail addresses:* mlabbe@ulb.ac.be (M. Labbé), dponce@us.es (D. Ponce), puerto@us.es (J. Puerto).

http://dx.doi.org/10.1016/j.cor.2016.06.004
0305-0548/© 2016 Elsevier Ltd. All rights reserved.

TOP (2021) 29:78–105
https://doi.org/10.1007/s11750-020-00572-z

ORIGINAL PAPER

## The ordered median tree of hubs location problem

Miguel A. Pozo[1] · Justo Puerto[1] · Antonio M. Rodríguez Chía[2]

Received: 3 February 2020 / Accepted: 26 May 2020 / Published online: 22 June 2020
© Sociedad de Estadística e Investigación Operativa 2020

**Abstract**
In this paper, we propose the Ordered Median Tree of Hub Location Problem (OMTHL). The OMTHL is a single-allocation hub location problem where $p$ hubs must be placed on a network and connected by a non-directed tree. Each non-hub node is assigned to a single hub and all the flow between origin–destination pairs must cross at least one hub. The objective is to minimize the sum of the ordered weighted averaged collection and distribution costs plus the sum of the interhub flow costs. We present different MILP formulations for the OMTHL based on the properties of the Minimum Spanning Tree Problem, the ordered median optimization and on the different ways of modeling flow within the network. Given that ordered median hub location problems are rather difficult to solve, we have improved the OMTHL solution performance by introducing covering variables in two valid reformulations. In addition, we have developed two pre-processing phases to reduce the size of these formulations. We establish an empirical comparison between these new formulations and we also provide enhancements that together with a proper formulation allow to solve medium-size instances on general random graphs.

**Keywords** Minimum spanning tree · Hub-and-spoke models · Ordered median

**Mathematics Subject Classification** 90B80 · 90C11

✉ Antonio M. Rodríguez Chía
antonio.rodriguezchia@uca.es

Miguel A. Pozo
miguelpozo@us.es

Justo Puerto
puerto@us.es

[1] Department of Statistics and Operational Research, University of Seville, Seville, Spain

[2] Department of Statistics and Operational Research, University of Cádiz, Puerto Real (Cádiz), Spain

Computational Optimization and Applications (2019) 73:603–645
https://doi.org/10.1007/s10589-019-00077-x

# Ordered *p*-median problems with neighbourhoods

Víctor Blanco[1,2]

Received: 23 October 2017 / Published online: 18 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

**Abstract**
In this paper, we introduce a new variant of the *p*-median facility location problem in which it is assumed that the exact location of the potential facilities is unknown. Instead, each of the facilities must be located in a region around their initially assigned location (the neighborhood). In this problem, two main decisions have to be made simultaneously: the determination of the potential facilities that must be open to serve the customers' demand and the location of the open facilities in their neighborhoods, at global minimum cost. We present several mixed integer non-linear programming formulations for a wide family of objective functions which are common in Location Analysis: ordered median functions. We also develop two math-heuristic approaches for solving the problem. We report the results of extensive computational experiments.

**Keywords** Facility location · Mixed integer second order cone programming · Neighborhoods · Ordered median

**Mathematics Subject Classification** 90B85 · 90C22 · 90C30 · 47A30

## 1 Introduction

Facility location concerns with the optimal placement of one or several new facilities/plants to satisfy the customers' demands. In discrete facility location problems, the positions of both the customers and the potential new facilities are part of the input as well as the travel costs between them. On the other hand, in continuous facility location problems, although the (geographical) coordinates (in a given *d*-dimensional space) of the customers are provided, the information about the potential location of the facilities is unknown, in the sense that the facilities can be located at any place of

✉ Víctor Blanco
vblanco@ugr.es

1 IEMath-GR, Universidad de Granada, Granada, Spain

2 Department of Quantitative Methods for Economics and Business, Universidad de Granada, Granada, Spain



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Computers and Operations Research 102 (2019) 75–90

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# Upgrading nodes in tree-shaped hub location

Víctor Blanco[a,b,*], Alfredo Marín[c]

[a] IEMath-GR, Universidad de Granada, Spain
[b] Department of Quantitative Methods for Economics & Business, Universidad de Granada, Spain
[c] Department of Statistics & Operational Research, Universidad de Murcia, Spain

ARTICLE INFO

*Article history:*
Received 15 February 2018
Revised 4 October 2018
Accepted 6 October 2018
Available online 9 October 2018

*MSC:*
90B80
90C11
05C05

*Keywords:*
Hub location
Spanning tree
Upgrading

ABSTRACT

Upgrading a node in a graph means to improve the quality of the edges that have it as an extreme. The first problem where upgrading was considered is the Minimum Cost Spanning Tree Problem. There, upgrading one or two of the extremes of an edge reduces progressively its cost and makes the edge more likely to be in the optimal tree. In this paper we analyze upgradings of nodes in a discrete location problem, the so-called Tree of Hubs Location Problem. Hubs are transshipment points through which a product is sent from origins to destinations. In addition to locate the hubs, to determine a tree that connects the hubs and to allocate non-hub nodes to hubs, i.e., the decisions to be made in the Tree of Hubs Location Problem, a decision has to be made about which of the hubs will be upgraded to reduce the overall cost of the distribution network. We provide two different Mixed Integer Linear Programming formulations for the problem, tighten the formulations and generate several families of valid inequalities for them. A computational study is presented showing the improvements attained with the strengthening of the formulations and comparing them.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

The main goal of Discrete Facility Location is to decide which set of facilities, among a finite set of potential sites, must be opened in order to optimize a given objective function which represents traveling costs of a set of given customers and/or set-up costs. Reaching such a goal is closely related with Mixed Integer Linear Programming (MILP) models. According to Laporte et al. (2015), the first MILP formulation for a discrete location problem, the *Uncapacitated Facility Location Problem*, was proposed by Balinski (1965). In the following few years, MILP formulations appeared also for the discrete *p*-Median Problem (ReVelle and Swain, 1970) and for a covering-location problem (Toregas et al., 1971). Since then, Discrete Facility Location has gained importance over the years as a fundamental part of both Integer Programming and Location Science.

Among the wide family of Discrete Facility Location problems, hub location problems have taken on significance in the last decades. Hub Location Problems (HLP) emerge as an important design tool in transportation systems, where several sites must send and receive some goods through some transshipment points, *the hubs*. At each of the hubs, the product is collected from different sources and distributed either by sending it to its final destination or to another hub. Using this system, clear cost savings are obtained compared to the direct *origin-to-destination* distribution network, because of economies of scale. Note that the amount of possible routes is considerably reduced, each of them traversed by a large amount of product. In this sense, a common characteristic of HLP is that the unit transportation cost is reduced if the product is routed between two hubs. The interest of hub location is still increasing, and one can read a number of surveys on the matter published over the last years (see Alumur and Kara, 2008; Campbell et al., 2002; Campbell and O'Kelly, 2012; Contreras, 2015; Farahani et al., 2013, among many others).

Hub location problems are usually classified according to the way of allocating non-hub customers (also called *spokes*) to hub nodes. On the one hand, *multiple–allocation* hub location problems assume that the flow with origin a single node may be distributed through any of the available hub nodes. Thus, depending on the destination, the flow may be initially routed by a different hub node. The first formulation for this family of problems was presented in Campbell (1994). On the other hand, in *single–allocation* models, the customers are only allowed to send and receive the service from a unique hub, i.e., each origin/destination is allocated to exactly one hub. Our study fits within this category, to which the pioneering works (Aykin, 1995; Ernst and Krishnamoorthy, 1996; 1998; O'Kelly, 1987; O'Kelly et al., 1996) and (Skorin-Kapov et al., 1996) belong to.

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), amarin@um.es (A. Marín).

https://doi.org/10.1016/j.cor.2018.10.007
0305-0548/© 2018 Elsevier Ltd. All rights reserved.

informs
http://pubsonline.informs.org/journal/ijoc

INFORMS JOURNAL ON COMPUTING
Vol. 32, No. 3, Summer 2020, pp. 582–599
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

# A Branch-Price-and-Cut Procedure for the Discrete Ordered Median Problem

Samuel Deleplanque,[a] Martine Labbé,[b] Diego Ponce,[c,d,e] Justo Puerto[c]

[a] Ifsttar, COSYS, ESTAS, Université Lille Nord de France, 59000 Lille, France; [b] Département d'Informatique, Faculté des Sciences, Université Libre de Bruxelles, 1050 Bruxelles, Belgium; [c] Instituto de Matemáticas de la Universidad de Sevilla (IMUS), 41012 Sevilla, Spain; [d] Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada; [e] Centre interuniversitaire de recherche sur les réseaux d'enterprise, la logistique et le transport (CIRRELT), Montreal, Quebec H3T 1J4, Canada

**Contact:** samuel.deleplanque@ifsttar.fr, http://orcid.org/0000-0003-4119-6006 (SD); mlabbe@ulb.ac.be, http://orcid.org/0000-0001-7471-2308 (ML); dponce@us.es, http://orcid.org/0000-0003-3380-6601 (DP); puerto@us.es, http://orcid.org/0000-0003-4079-8419 (JP)

**Received:** February 9, 2018
**Revised:** December 7, 2018; March 11, 2019
**Accepted:** June 2, 2019
**Published Online in Articles in Advance:** July 2, 2020

https://doi.org/10.1287/ijoc.2019.0915

**Copyright:** © 2020 INFORMS

**Abstract.** The discrete ordered median problem (DOMP) is formulated as a set-partitioning problem using an exponential number of variables. Each variable corresponds to a set of demand points allocated to the same facility with the information of the sorting position of their corresponding costs. We develop a column generation approach to solve the continuous relaxation of this model. Then we apply a branch-price-and-cut algorithm to solve small- to large-sized instances of DOMP in competitive computational time.

**Funding:** This research was supported by Spanish/FEDER [Grant MTM2016-74983-C02-01-R]. The research of the second and third authors was supported by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.
**History:** Accepted by Andrea Lodi, Area Editor for Design & Analysis of Algorithms—Discrete.

**Keywords:** discrete optimization • location theory • branch and price • ordered median problems

## 1. Introduction

Broadly speaking, a facility-location problem consists of locating one or several facilities to minimize an objective function of the assignment costs of clients to facilities. Median and center location problems constitute the most popular ones. The ordered median location problem is a flexible model that provides a common framework to cast most popular location problems. The objective function to be minimized is a weighted function of the allocation costs in which the weights are assigned to the ordered values of the costs rather than to specific costs. Ordered median location problems were first introduced in networks and continuous spaces by Nickel and Puerto (1999) and Puerto and Fernández (2000), respectively. Later, they were extended to the discrete setting by Nickel (2001) and Boland et al. (2006).

Given a set of clients and a set of candidate locations and assuming that the allocation costs of clients to facilities are known, the discrete ordered median problem (DOMP) consists of choosing $p$ facility locations and assigning each client to a chosen facility with the smallest allocation cost to minimize the ordered weighted average of these costs. The ordered weighted average sorts the allocation costs in a nondecreasing sequence and then it performs the scalar product of this so-obtained sorted cost vector with a given vector of weights.

DOMP has been widely studied since the 1990s, and there are a number of different formulations, solution approaches, and applications available in the literature. To cite a few, DOMP has been applied to discrete facility location in Boland et al. (2006), Marín et al. (2009, 2010), Nickel (2001), Puerto (2008), and Puerto et al. (2009); to location on networks in Nickel and Puerto (1999); to hub network design problems in Puerto et al. (2011, 2013, 2016); to determine values in cooperative game theory in Perea and Puerto (2013); to combinatorial optimization problems with ordering in Fernández et al. (2013, 2014, 2017); and to voting problems in Ponce et al. (2018), etc. The reader is referred to the monographs by Nickel and Puerto (2005) and Puerto and Rodríguez-Chía (2015) for some other applications.

There exist several valid formulations for DOMP that exploit specific features of the problem, for instance, free self-service, ties in the matrix of costs, or null elements in the vector of weights (see, e.g., Boland et al. (2006), Marín et al. (2009, 2010), Puerto et al. (2013), Labbé et al. (2017), and the references therein). In Labbé et al. (2017), a new formulation for DOMP has been proposed based on a set-packing approach that is valid for general cost coefficients. This formulation gives rise to rather tight integrality gaps and was shown to be reasonably efficient to solve medium-sized instances when embedded in a branch-and-cut (B&C) scheme. For general cost coefficients (with no ties), all these formulations have a very large (cubic) number of

European Journal of Operational Research 286 (2020) 839–848

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# A fresh view on the Discrete Ordered Median Problem based on partial monotonicity

Alfredo Marín[a], Diego Ponce[b,c,d], Justo Puerto[b,*]

[a] Department of Statistics and Operations Research, Faculty of Mathematics, Universidad de Murcia, Spain
[b] Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Spain
[c] Centre interuniversitaire de recherche sur les réseaux d'enterprise, la logistique et le transport (CIRRELT), Montreal, Canada
[d] Department of Mechanical, Industrial and Aerospace Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, Canada

ARTICLE INFO

*Article history:*
Received 7 January 2020
Accepted 14 April 2020
Available online 28 April 2020

*Keywords:*
Location
Combinatorial optimization
Logistics
Ordered median problem

ABSTRACT

This paper presents new results for the Discrete Ordered Median Problem (DOMP). It exploits properties of $k$-sum optimization to derive specific formulations for the monotone DOMP (MDOMP), that arises when the $\lambda$ weights are non-decreasing monotone, and new formulations for the general non-monotone DOMP. The main idea in our approach is to express ordered weighted averages as telescopic sums whose terms are $k$-sums, with positive and negative coefficients. Formulations of $k$-sums with positive coefficients derive from the linear programming representations obtained by Ogryczack and Tamir (2003) and Blanco, Ali, and Puerto (2014). Valid formulations for $k$-sums with negative coefficients are more elaborated and we present 4 different approaches, all of them based on mixed integer programming formulations. An extensive computational experience based on a collection of well-known instances shows the usefulness of the new formulations to solve difficult problems such as trimmed and anti-trimmed mean.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

In a remarkable paper by Ogryczack and Tamir (2003) these authors introduce a novel linear time algorithm to compute the sum of the $k$-largest entries ($k \leq n$) of an arbitrary vector of $n$ components. This approach gives rise to a linear programming representation of such an evaluation. More recently, a different linear programming representation of the sum of the $k$ largest components of a vector has also been obtained Blanco, Ali, and Puerto (2014). The earlier idea in Ogryczack and Tamir (2003) was later exploited by Kalcsics, Nickel, Puerto, and Tamir (2002) to develop an efficient representation for the $k$-centrum location problem and more generally extended to deal with some classes of Discrete Ordered Median Problems (DOMP) (the reader is referred to Kalcsics, Nickel, Puerto, & Rodríguez-Chía, 2010; Nickel & Puerto, 2005; Puerto, Rodríguez-Chía, & Tamir, 2009 for further details on this problem). More recently, the same approach has been used in Puerto, Rodríguez-Chía, and Tamir (2017) to provide new algorithms to the general $k$-sum optimization problem. The interested reader is referred to Aouad and Segev (2019), Blanco (2019), Deleplanque, Labbé, Ponce, and Puerto (2020), Labbé, Ponce, and Puerto (2017), Olender and Ogryczak (2019) or Puerto (2020) for the most recent results on DOMP appeared in the literature.

$k$-sum optimization problems are well-understood and solved in a number of important cases (see Puerto et al., 2017). In particular, it has been observed that DOMP, under the hypothesis that the $\lambda$ modeling weights are monotone, can be easily reformulated using as building blocks $k$-sum terms.

The main goal of this paper is to show how to exploit the powerfulness of $k$-sum optimization within the framework of the monotone and general non-monotone DOMP. Our contribution is to present a range of formulations for the monotone and general version of DOMP using different forms of representing $k$-sums. In this journey, we resort to express ordered weighted averages as telescopic sums whose terms are $k$-sums, with positive and negative coefficients. Next, to get valid mathematical programming formulations we distinguish between reformulations for $k$-sums with positive and negative coefficients in the objective function. Formulations of $k$-sums with positive coefficients derive from the linear programming representations obtained by Ogryczack and Tamir (2003) and Blanco et al. (2014). Valid formulations for $k$-sums with negative coefficients are more cumbersome and we present 4 different approaches, all of them mixed integer programming formulations. The first two are based on radius formulations in the

* Corresponding author.
*E-mail addresses:* amarin@um.es (A. Marín), dponce@us.es (D. Ponce), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.ejor.2020.04.023
0377-2217/© 2020 Elsevier B.V. All rights reserved.

Journal of Global Optimization (2020) 77:341–359
https://doi.org/10.1007/s10898-019-00863-1

# An exact completely positive programming formulation for the discrete ordered median problem: an extended version

J. Puerto[1]

Received: 8 September 2018 / Accepted: 2 December 2019 / Published online: 9 December 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

**Abstract**
This paper presents a first continuous, linear, conic formulation for the discrete ordered median problem (DOMP). Starting from a binary, quadratic formulation in the original space of location and allocation variables that are common in location analysis (L.A.), we prove that there exists a transformation of that formulation, using the same space of variables, that allows us to cast DOMP as a continuous, linear programming problem in the space of completely positive matrices. This is the first positive result that states equivalence between the family of continuous, convex problems and some hard combinatorial problems in L.A. The result is of theoretical interest because it allows us to share the tools from continuous optimization to shed new light into the difficult combinatorial structure of the class of ordered median problems that combines elements of the $p$-median, quadratic assignment and permutation polytopes.

**Keywords** Discrete ordered median problem · Completely positive reformulation · Conic linear programming

## 1 Introduction

The discrete ordered median problem (DOMP) represents a generalization of several well-known discrete location problems, such as p-median, $p$-center or $(k_1 + k_2)-$trimmed mean, among many others. DOMP provides a very flexible tool in accommodating actual aspects, as subsidized or compensation costs, in models of Logistics and Location, see Puerto and Fernández [22] and Nickel and Puerto [19]. This way different objective functions are cast within the same framework and similar tools can be used to solve problems with apparently different structure. The ordered median objective function computes ordered weighted averages of vectors. In the case when that objective function is applied to location problems those vectors correspond to distances or allocation costs from clients to service facilities. The problem was introduced in Nickel [18] and later studied by Kalcsics et al. [13], Nickel and Puerto [19], Boland et al. [3], Stanimirovic et al. [25], Marín et al. [17], Perea and Puerto

✉ J. Puerto
puerto@us.es

[1] IMUS, Universidad de Sevilla, Seville, Spain



Computers and Operations Research 125 (2021) 105067

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/caor

# A comparative study of different formulations for the capacitated discrete ordered median problem

I. Espejo[a,*], J. Puerto[b,c], A.M. Rodríguez-Chía[a]

[a] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain
[b] IMUS Instituto de Matemáticas, Universidad de Sevilla, Spain
[c] Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 25 June 2019
Revised 25 July 2020
Accepted 25 July 2020
Available online 5 August 2020

*Keywords:*
Discrete location
Capacitated location problem
Ordered median location problem

ABSTRACT

This paper deals with the capacitated version of discrete ordered median problems. We present different formulations considering three-index variables or covering variables to address the order requirements in this problem. Some preprocessing phases for fixing variables and some valid inequalities are developed to enhance the initial formulations. Finally, an extensive computational analysis is addressed with data taken from the OR-library and AP-library showing the efficiency of the formulations and the improvements presented in the paper.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

A very important aspect of most optimization problems is the correct choice of their objective functions. If the considered problem models an actual application, the objective function plays a crucial role because it has to describe the utility (disutility) to be optimized in order to achieve a reasonable outcome once a solution is found. Therefore, modelling flexible families of objective functions that are applicable in contemporary Operations Research, and more specifically in Location Analysis, is a fruitful area of research. Apart from the choice of the optimization criterion, another crucial aspect in the literature on facility location, is the assumption of capacity constraints. This assumption implies more realistic models at the price of increasing its complexity with respect to their uncapacitated counterparts. In many cases new formulations are needed and a more specialized analysis is often required to solve even smaller sizes than those previously addressed for the uncapacitated versions of the same problems. This paper tries to contribute to this area of research providing new insights, models and algorithms to solve the so called capacitated discrete ordered median problem (CDOMP), see Nickel and Puerto (2005).

The Discrete Ordered Median Problem (DOMP) is a flexible model that provides a common framework to cast most popular problems in discrete location analysis. These models share in common a similar multiparametric objective function called *ordered median function* that was introduced for the first time in Nickel and Puerto (1999) and Puerto and Fernández (2000). An ordered median function is an ordered weighted average of the arguments of that function. Therefore, when it is applied to a location problem it results in the ordered weighted average of the distances or allocation costs from clients to service facilities. Given a set of clients, a set of candidate locations and assuming that the allocation costs of clients to facilities are known, DOMP consists in choosing $p$ facility locations and assigning each client to a chosen facility with the smallest allocation cost in order to minimize the ordered weighted average of these costs. The ordered weighted average sorts the allocation costs in a non-decreasing sequence and then it performs the scalar product of this so-obtained sorted cost vector with a given vector of weights, $\lambda$-vector.

This idea of sorting elements and exploiting radii is in the basis of current state of the art formulation in location theory. Specifically, state of the art models for the p-median are those by Elloumi (2010) and García et al. (2011) that solve the largest instances in the literature. These models are specifications of the general idea of radius formulations that is the basis of ordered median formulations. Actually, these formulations can be seen as particular instances of the latter whenever all lambdas are equal to one. The same happens for the p-center problem where the best formulations, namely those by Elloumi et al. (2004) and Calik and Tansel (2013) are also based on radius formulations that appear from ordered median ones with particular choices of lambdas. Therefore, the use of the rationale behind ordered median models

* Corresponding author. Tel.: +34 956012715.
*E-mail address:* inmaculada.espejo@uca.es (I. Espejo).

https://doi.org/10.1016/j.cor.2020.105067
0305-0548/© 2020 Elsevier Ltd. All rights reserved.

# Datos: 13 publicaciones

European Journal of Operational Research 261 (2017) 43–53

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

## Clustering data that are graph connected

Stefano Benati[a], Justo Puerto[b], Antonio M. Rodríguez-Chía[c,*]

[a] Dipartimento di Sociologia e Ricerca Sociale, Università di Trento, Via Verdi 26, 38122 Trento. Italy
[b] IMUS. Universidad de Sevilla, Avda. Reina Mercedes, s/n, 41012 Sevilla. Spain
[c] Faculty of Sciences, Universidad de Cádiz, Avda. República Saharaui, 11510 Puerto Real (Cádiz). Spain

ARTICLE INFO

*Article history:*
Received 18 June 2016
Accepted 7 February 2017
Available online 12 February 2017

*Keywords:*
Combinatorial optimization
Clustering
Clique partitioning
Integer programming

ABSTRACT

A new combinatorial model for clustering is proposed for all applications in which individual and relational data are available. Individual data refer to the intrinsic features of units, they are stored in a matrix $D$, and are the typical input of all clustering algorithms proposed so far. Relational data refer to the observed links between units, representing social ties such as friendship, joint participation to social events, and so on. Relational data are stored in the graph $G = (V, E)$, and the data available for clustering are the triplet $G = (V, E, D)$, called attributed graph. Known clustering algorithms can take advantage of the relational structure of $G$ to redefine and refine the units membership. For example, uncertain membership of units to groups can be resolved using the sociological principle that ties are more likely to form between similar units. The model proposed here shows how to take into account the graph information, combining the clique partitioning objective function (a known clustering methodology) with connectivity as the structural constraint of the resulting clusters. The model can be formulated and solved using Integer Linear Programming and a new family of cutting planes. Moderate size problems are solved, and heuristic procedures are developed for instances in which the optimal solution can only be approximated. Finally, tests conducted on simulated data show that the clusters quality is greatly improved through this methodology.

© 2017 Elsevier B.V. All rights reserved.

### 1. Introduction

The problem of clustering consists in discovering or detecting how a population is partitioned into two or more subgroups, each subgroup specified by distinct features. The typical outcome of a clustering algorithm is the assignment of observations to groups and, most of the times, some estimation of the characteristics of every group. There are many techniques proposed for clustering. Some of them are constructive, in the sense that observations are merged together one at a time until the required partition is found, hierarchical trees being an example. Other techniques have combinatorial structure, for example the $k$-means, in which clusters are the solutions of an optimization problem, and different partitions are evaluated through the objective function. Finally, some techniques assume that data can be described through probability distributions, so that the optimal clustering is calculated maximizing the likelihood function, for example with the EM algorithm. Each technique has its pros and cons, consequently it can be the most appropriate for some application. Clustering is an easy task if groups are well separated. Conversely, if the borders between groups are uncertain and populated by many observations of different groups with similar features, then the precision of the classification decreases to the point that algorithms just output random clusters.

In this contribution, it is shown that the capability of detecting the true clusters is enhanced if, together with the standard individual data (that is, data that are pertinent to the single population unit), the researcher can observe relational data too, that is, data describing the connections among units. To make an example in social network analysis, suppose that a researcher wants to determine groups with homogeneous cultural orientations within a population of individuals. Data come from a survey in which people answered to questions about their attitudes to religion, politics, family and so on. It is often the case that individuals classified as *"secular"* are not neatly separated from individuals classified as *"religious"*, as they can share the same views on social matters different from religion, see (Inglehart & Baker, 2000). Therefore the clustering algorithm does not correctly classify individuals and the subsequent analysis is flawed, for example the estimation of the groups parameters could result wrong. We propose that a possible remedy to the border uncertainty is to consider an additional data: the relationships, such as friendship, kinship, sympathy, and so on,

* Corresponding author.
*E-mail address:* antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

http://dx.doi.org/10.1016/j.ejor.2017.02.009
0377-2217/© 2017 Elsevier B.V. All rights reserved.

Computers and Operations Research 87 (2017) 137–145

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## Lagrangian relaxation for SVM feature selection

M. Gaudioso[a,*], E. Gorgone[d,e], M. Labbé[b], A.M. Rodríguez-Chía[c]

[a] Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica, Università della Calabria, Rende (CS) 87036, Italia
[b] Département d'Informatique, Université Libre de Bruxelles, CP 212 Boulevard du Triomphe, Brussels B-1050, Belgium
[c] Faculty of Sciences, Universidad de Cádiz Avenida República Saharaui s/n 11510. Puerto Real Cádiz, Spain
[d] Indian Institute of Management Bangalore (IIMB), Bannerghatta Road, Bangalore 560076, India
[e] Dipartimento di Matematica e Informatica, Università di Cagliari, Cagliari 09124, Italy

ARTICLE INFO

*Article history:*
Received 20 July 2016
Revised 1 June 2017
Accepted 1 June 2017
Available online 15 June 2017

*Keywords:*
SVM classification
Feature selection
Lagrangian relaxation
Nonsmooth optimization

ABSTRACT

We discuss a Lagrangian-relaxation-based heuristics for dealing with feature selection in the Support Vector Machine (SVM) framework for binary classification. In particular we embed into our objective function a weighted combination of the $L_1$ and $L_0$ norm of the normal to the separating hyperplane. We come out with a Mixed Binary Linear Programming problem which is suitable for a Lagrangian relaxation approach.

Based on a property of the optimal multiplier setting, we apply a consolidated nonsmooth optimization ascent algorithm to solve the resulting Lagrangian dual. In the proposed approach we get, at every ascent step, both a lower bound on the optimal solution as well as a feasible solution at low computational cost.

We present the results of our numerical experiments on some benchmark datasets.

© 2017 Elsevier Ltd. All rights reserved.

### 1. Introduction

The focus of pattern classification is to recognize similarities in the data, categorizing them in different subsets (Cristianini and Shawe-Taylor, 2000; Carrizosa and Morales, 2013; Vapnik, 1995). In many fields, such as the financial and the medical ones (Guyon et al., 2002; Lee and Tsung-Lin, 2009), classification of data (samples in Machine Learning language) is useful for analysis or diagnosis purposes.

Quite often datasets are formed by a small number of samples, which in turn are characterized by a huge number of attributes (features). The handling of the entire feature set would be computationally very expensive and its outcome would lack from insight. For this reason, it is convenient to reduce the set of features which is expected to be easier to interpret and also easy to evaluate. However, it is not always easy to predict which of those are relevant for classification purposes.

Hence it is necessary to screen off the relevant features from those which are irrelevant (Bi et al., 2003).

The process that selects the features entering the subset of the relevant ones is known as *Feature Selection* (FS) and the related literature is extremely rich (see e.g. Guyon and Elisseeff (2003); Guyon et al. (2006); Kittler (1986); Meyer et al. (2008)). As far as mathematical programming-based approaches are concerned, we cite here Weston et al. (2000) and, more specifically, Bradley et al. (1998), where (FS) is pursued by formulating a mathematical program with a parametric objective function embedding a concave approximation of the zero-norm of the feature vector. Other approaches based on zero-norm minimization are in Weston et al. (2003) and Rinaldi and Sciandrone (2010). In a recent paper (Aytug, 2015) a model based on penalization of the number of features entering into the classification process is treated by means of generalized Benders decomposition.

The objective of this paper is to treat explicitly the (FS) problem as a Mixed Binary Programming (MBP) one (Bertolazzi et al., 2016; Maldonado et al., 2014), in the framework of the SVM (Support Vector Machine) (Chen and Lin, 2006; Cristianini and Shawe-Taylor, 2000; Nguyen and de la Torre, 2010; Vapnik, 1995) approach. Consequently the objective of our model is threefold: to minimize the classification error, to maximize the separation margin and to minimize the number features playing a role in the classification process. To this aim we embed into the objective function, in addition to a measure of the classification error, the weighted combination of the $L_1$ and $L_0$ norms of the feature vector. We note that use of such norms (in particular the $L_1$ is also known as LASSO penalty) has been adopted in several papers in the SVM framework.

* Corresponding author.
*E-mail addresses:* gaudioso@dimes.unical.it (M. Gaudioso), enrico.gorgone@iimb.ernet.in, egorgone@unica.it (E. Gorgone), mlabbe@ulb.ac.be (M. Labbé), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

http://dx.doi.org/10.1016/j.cor.2017.06.001
0305-0548/© 2017 Elsevier Ltd. All rights reserved.

Computers and Operations Research 95 (2018) 172–193

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## Locating hyperplanes to fitting set of points: A general framework

Víctor Blanco[a,*], Justo Puerto[b], Román Salmerón[c]

[a] Dpt. Quant. Methods for Economics & Business and IEMath-GR, Universidad de Granada, Spain
[b] Dpt. Statistics & OR and IMUS, Universidad de Sevilla, Spain
[c] Dpt. Quant. Methods for Economics & Business, Universidad de Granada, Spain

ARTICLE INFO

*Article history:*
Received 17 November 2016
Revised 21 March 2018
Accepted 22 March 2018
Available online 23 March 2018

*2010 MSC:*
90B85
90C26
52C35
65D10

*Keywords:*
Fitting hyperplanes
Mathematical programming
Location of structures
Robust fitting
Linear regression

ABSTRACT

This paper presents a family of methods for locating/fitting hyperplanes with respect to a given set of points. We introduce a general framework for a family of aggregation criteria, based on ordered weighted operators, of different distance-based errors. The most popular methods found in the specialized literature, namely least sum of squares, least absolute deviation, least quantile of squares or least trimmed sum of squares among many others, can be cast within this family as particular choices of the errors and the aggregation criteria. Unified mathematical programming formulations for these methods are provided and some interesting cases are analyzed. The most general setting give rise to mixed integer nonlinear programming problems. For those situations we present inner and outer linear approximations to assess tractable solution procedures. It is also proposed a new goodness of fitting index which extends the classical coefficient of determination and allows one to compare different fitting hyperplanes. A series of illustrative examples and extensive computational experiments implemented in R are provided to show the applicability of the proposed methods.

© 2018 Elsevier Ltd. All rights reserved.

### 1. Introduction

The problem of locating hyperplanes with respect to a given set of point is well-known in Location Analysis (LA) Schöbel (1999). This problem is closely related to another common question in Data Analysis (DA): to study the behavior of a given set of data with respect to a fitting body expressed with an equation of the form $f(x) = 0$, with $x = (X_1, \ldots, X_d) \in \mathbb{R}^d$. This last problem reduces to the estimation of the 'best' function $f$ that expresses the relationship between the data or, in the jargon of LA, to the location of the surface $f(x) = 0$ that minimizes some aggregation function of the distances to these points (see Amaldi et al., 2016; Díaz-Báñez et al., 2004; Drezner et al., 2002). In many cases the family of functions where $f$ belongs to is fixed and then, the parameters defining such an *optimal* function must be determined. The family of linear functions is the most widely used. This implies that the above equation is of the form $f(x) = \beta_0 + \sum_{k=1}^{d} \beta_k X_k = 0$ for $\beta_0, \beta_1, \ldots, \beta_d \in \mathbb{R}$.

To perform such a fitting, we are given a set of points $\{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, and the goal is to find the vector $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_d)$ that minimizes some measure of the deviation of the data with respect to the hyperplane it induces, $\mathcal{H}(\hat{\boldsymbol{\beta}}) = \{z \in \mathbb{R}^d : \hat{\beta}_0 + \sum_{k=1}^{d} \hat{\beta}_k z_k = 0\}$. For a given point $x \in \mathbb{R}^d$, we define the *residual* with respect to a generic $x$ as a mapping $\varepsilon_x : \mathbb{R}^{d+1} \to \mathbb{R}_+$, that maps any set of coefficients $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_d) \in \mathbb{R}^{d+1}$, into a measure $\varepsilon_x(\boldsymbol{\beta})$ that represents the deviation of the given point $x$ from the hyperplane with those parameters. The problem of locating a hyperplane for a given set of points $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$ consists of finding the coefficients minimizing an aggregation function, $\Phi : \mathbb{R}^n \to \mathbb{R}$, of the residuals of all the points. Different choices for the residuals and the aggregation criteria will give, in general, different optimal values for the parameters and thus different properties for the resulting hyperplanes. This problem is not new and some of these criteria, as the minisum, minimax and some other alternatives, have been widely analyzed from a LA perspective (see Carrizosa and Plastria, 1995; Megiddo and Tamir, 1983; Schöbel, 1996; Schöbel, 1997; Schöbel, 1998; Schöbel, 1999, among other).

A first approach to locate a hyperplane is to consider that residuals, with respect to given points, are individual measures of error and thus, each residual should be minimized independently of the remaining (Carrizosa et al., 1995; Narula and Wellington, 2007). It is clear that this simultaneous minimization will not be possible in most of the cases and then several strategies can be followed: one can try to find the set of Pareto fitting curves (Carrizosa et al.,

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), puerto@us.es (J. Puerto), romansg@ugr.es (R. Salmerón).

https://doi.org/10.1016/j.cor.2018.03.009
0305-0548/© 2018 Elsevier Ltd. All rights reserved.

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY, 2018
VOL. 69, NO. 9, 1379–1395
https://doi.org/10.1080/01605682.2017.1398206

# Mixed integer linear programming and heuristic methods for feature selection in clustering

Stefano Benati[a], Sergio García[b] and Justo Puerto[c]

[a]School of International Studies, University of Trento, Trento, Italy; [b]School of Mathematics, The University of Edinburgh, Edinburgh, UK; [c]Institute of Mathematics of the University of Seville (IMUS), Universidad de Sevilla, Sevilla, Spain

**ABSTRACT**

This paper studies the problem of selecting relevant features in clustering problems, out of a data-set in which many features are useless, or masking. The data-set comprises a set $U$ of units, a set $V$ of features, a set $R$ of (tentative) cluster centres and distances $d_{ijk}$ for every $i \in U$, $k \in R$, $j \in V$. The feature selection problem consists of finding a subset of features $Q \subseteq V$ such that the total sum of the distances from the units to the closest centre is minimised. This is a combinatorial optimisation problem that we show to be NP-complete, and we propose two mixed integer linear programming formulations to calculate the solution. Some computational experiments show that if clusters are well separated and the relevant features are easy to detect, then both formulations can solve problems with many integer variables. Conversely, if clusters overlap and relevant features are ambiguous, then even small problems are unsolved. To overcome this difficulty, we propose two heuristic methods to find that, most of the time, one of them, called $q$-vars, calculates the optimal solution quickly. Then, the $q$-vars heuristic is combined with the $k$-means algorithm to cluster some simulated data. We conclude that this approach outperforms other methods for clustering with variable selection that were proposed in the literature.

**ARTICLE HISTORY**
Received 4 July 2016
Accepted 20 October 2017

**KEYWORDS**
Integer linear programming; heuristics; $q$-vars; cluster analysis; $p$-median problem

## 1. Introduction

Clustering is a useful and important unsupervised learning technique widely studied in the literature. The goal of clustering is to group similar units (or objects) into one cluster, while partitioning dissimilar units into different clusters. Clustering becomes difficult if data contain features (also referred to as variables) with no relevant information. When those features are not detected, the calculation of the dissimilarity between units is biased by their presence, resulting in inconsistent clusters (Fowlkes, Gnanadesikan, & Kettering, 1988). The problem becomes more and more relevant as the number of features in a database increases, as frequently occurs nowadays with data containing hundreds and even thousands of covariates (see Guyon & Elisseef, 2003). Therefore, researchers from different disciplines need tools to discard the noising or masking features that are useless to recognise the true clusters. Previous literature has focused on three methods to select or reject variables. The first is the most simple and consists of calculating peculiar indices, one for each variable, to distinguish those features that contain recognisable patterns. The second approach is the most elaborate and consists of optimising the maximum likelihood function, assuming that data are generated by multivariate distributions. The third approach, whose difficulty is intermediate between the previous ones, is to solve an optimisation model whose structure is easier than the maximum likelihood.

The first approach, developing "clusterability" indices, has been suggested in Andrews and McNicholas (2014), Carmone, Kara, and Maxwell (1999), Morlini and Zani (2013), and Steinley and Brusco (2008b). In Carmone et al. (1999), an index called Total Pairwise Rand Index (TOPRI) has been used to select those features that form the best clusters. Features are selected using a constructive procedure that selects features one at a time until a stop criterion is met. In Steinley and Brusco (2008b) an index called clusterability index is defined and used within a multi-step procedure. This procedure is composed of variable preprocessing, data standardisation and variable selection through optimisation. In Andrews and McNicholas (2014), the role of the index is played by an inequality that balances correlation with variability reduction, used for preliminary data screening. In Morlini and Zani (2013), another index is used to compare the outcome of hierarchical clusterings applied to data subsets. The advantage of all these methods is that the calculation of an index is usually a fast task. However, the "greedy" constructive way in which features are selected is clearly sub-optimal.

The second approach, optimising the data likelihood function, assumes that data are described by multivariate distribution functions. This distribution function

**CONTACT** Stefano Benati stefano.benati@unitn.it

© Operational Research Society 2017

European Journal of Operational Research 273 (2019) 441–451

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# Exact algorithms for handling outliers in center location problems on networks using $k$-max functions

Teresa Schnepper[a,*], Kathrin Klamroth[a], Michael Stiglmayr[a], Justo Puerto[b]

[a] School of Mathematics and Natural Sciences, University of Wuppertal, Gaußstraße 20, Wuppertal 42103, Germany
[b] Mathematical Research Institute (IMUS), University of Seville, C/ Tarfia 21, Seville 41012, Spain

ARTICLE INFO

*Article history:*
Received 30 January 2018
Accepted 18 August 2018
Available online 23 August 2018

*Keywords:*
Location theory
Center problem
Outlier
$k$-max function
Finite dominating set

ABSTRACT

In this paper, the identification and exclusion of very distant facilities in center location problems is modeled by $k$-max functions: One or several new facilities are to be located such that not the largest, but the $k$th largest weighted distance to the customers is minimized, with $k \geq 1$. It is shown that $k$-max location problems on networks can be solved efficiently by enumerating candidate solutions from a finite dominating set (FDS) that is independent from the particular value of $k$. As a consequence, $k$-max locations can be found for all reasonable values of $k$ at little extra cost as compared to a single solver call, for one fixed value of $k$. The location of several new facilities is naturally more complex. An FDS based recursive algorithm is developed for this case that allows the exact solution of medium size instances.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Location models typically use the distances between a set of new facilities (i.e., facilities to locate) and a set of already existing customers (also called clients) as a quality criterion for the service provided by the new facilities. Common ways to aggregate these distances in one objective function are the center function, which relies on the largest distance between any customer and its closest new facility and the median function, which relies on the sum of all these distances (see, e.g., Hakimi, 1964).

When the customers are distributed such that there is a (small) number of "far-away" customers, i.e., clients that are located very distant from the majority of the other clients, these classical objective functions may be arguable from an economic point of view. Particularly in center location problems the distance to a majority of customers may be unnecessarily large due to the attracting effect of only one or two very distant customers.

In situations where mostly economic aspects are relevant, this motivates an adaptive and problem specific exclusion of so called *outliers*. Excluding outliers makes location problems more robust with respect to the influence of far-away customers. For example, when locating delivery stations supplying food or maintenance services, the possibility to exclude very distant customers from the placement decision of a new delivery station can lead to a more reasonable and economic solution from the decision maker point of view. However, it should be emphasized that excluding outliers from location models is usually not acceptable in situations where emergency facilities such as fire stations, hospitals or police departments have to be located. In these cases it is important that *all* houses are reachable within a certain time limit.

To limit the influence of far-away customers on the optimal center location, the $k$th largest distance between any customer and its corresponding closest new facility is used as a decision criterion in this paper. With the parameter $k$, the number of acceptable outliers is specified as $k-1$. Thus, $k$ can be varied from 1 up to the number of customers. The resulting objective function is referred to as a $k$-max function in the following.

There are various applications for $k$-max location problems in practice. As far-away customers may have a negative effect on the location of new facilities, efficient approaches to handle outliers are particularly useful for center problems with non-uniformly distributed customers. To mention a concrete example, the US-based department store chain Kmart uses outlier handling methods to cover 88% of the US population within a radius of six miles, which is economically more useful to them than to cover the whole population (see Charikar, Khuller, Mount, & Narasimhan, 2001). Another real life example appeared in the design of the high-speed railway network in Spain. According to the specifications of PEIT-2005 (Spanish Infrastructure and Transportation Plan) the final network should cover 90% of the total Spanish population (thus considering outliers the remaining 10%) within a radius of 50 kilometers. Further details on modelling and solution approaches to

* Corresponding author.
*E-mail addresses:* schnepper@math.uni-wuppertal.de, teresa.schnepper@gmx.de (T. Schnepper), klamroth@math.uni-wuppertal.de (K. Klamroth), stiglmayr@math.uni-wuppertal.de (M. Stiglmayr), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.ejor.2018.08.030
0377-2217/© 2018 Elsevier B.V. All rights reserved.

Discrete Applied Mathematics 261 (2019) 276–304

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

# Mixed integer linear programming for feature selection in support vector machine

Martine Labbé [a,b], Luisa I. Martínez-Merino [c], Antonio M. Rodríguez-Chía [c,*]

[a] *Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium*
[b] *INOCS, INRIA Lille Nord-Europe, Lille, France*
[c] *Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain*

ARTICLE INFO

*Article history:*
Received 20 March 2017
Received in revised form 2 October 2018
Accepted 19 October 2018
Available online 22 November 2018

*Keywords:*
Mathematical programming
Kernel search algorithm
Supervised classification
Support vector machine
Feature selection

ABSTRACT

This work focuses on support vector machine (SVM) with feature selection. A MILP formulation is proposed for the problem. The choice of suitable features to construct the separating hyperplanes has been modeled in this formulation by including a budget constraint that sets in advance a limit on the number of features to be used in the classification process. We propose both an exact and a heuristic procedure to solve this formulation in an efficient way. Finally, the validation of the model is done by checking it with some well-known data sets and comparing it with classical classification methods.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

In supervised classification, we are given a set of objects $\Omega$ partitioned into classes and the goal is to build a procedure for classifying new objects into these classes. This type of problem is faced in many fields, including insurance companies (to determine whether an applicant is a high insurance risk or not), banks (to decide whether an applicant is a credit risk or not), medicine (to determine whether a tumor is benign or malignant), etc. This wide field of application has attracted the attention of a number of researchers from different areas. Currently, these problems are analyzed from different perspectives: artificial intelligence, machine learning, optimization or statistical pattern recognition among others. In this paper we analyze these problems from the point of view of optimization and more specifically, from the perspective of Mathematical Programming [17,18].

In the partitioning process, the objects of $\Omega$ are considered as points in an $n$-dimensional feature space. However, the number of features is often much larger than the number of objects in the population. Handling such a high number of features could therefore be a difficult task and in addition, the interpretation of the results could also be impossible. In this sense, feature selection consists of eliminating as many features as possible in a given problem, keeping an acceptable accuracy where spurious random variables are removed. Actually, having a minimal number of features often leads to better generalization and simpler models that can be interpreted more easily.

The support vector machine (SVM) is a type of mathematical programming approach developed by Vapnik [21], and Cortes and Vapnik [10]. It has been widely studied and has become popular in many fields of application in recent years; see the

* Corresponding author.
*E-mail address:* antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.dam.2018.10.025
0166-218X/© 2018 Elsevier B.V. All rights reserved.

Journal of Machine Learning Research 21 (2020) 1-29 Submitted 9/18; Revised 6/19; Published 02/20

# On $\ell_p$-Support Vector Machines and Multidimensional Kernels

**Víctor Blanco** VBLANCO@UGR.ES
*IEMath-GR, Universidad de Granada, SPAIN*

**Justo Puerto** PUERTO@UGR.ES
*IMUS, Universidad de Sevilla, SPAIN*

**Antonio M. Rodríguez-Chía** ANTONIO.RODRIGUEZCHIA@UCA.ES
*Dpt. Statistics & OR, Universidad de Cádiz, SPAIN*

**Editor:** Mark Schmidt

## Abstract

In this paper, we extend the methodology developed for Support Vector Machines (SVM) using the $\ell_2$-norm ($\ell_2$-SVM) to the more general case of $\ell_p$-norms with $p > 1$ ($\ell_p$-SVM). We derive second order cone formulations for the resulting dual and primal problems. The concept of kernel function, widely applied in $\ell_2$-SVM, is extended to the more general case of $\ell_p$-norms with $p > 1$ by defining a new operator called multidimensional kernel. This object gives rise to reformulations of dual problems, in a transformed space of the original data, where the dependence on the original data always appear as homogeneous polynomials. We adapt known solution algorithms to efficiently solve the primal and dual resulting problems and some computational experiments on real-world datasets are presented showing rather good behavior in terms of the accuracy of $\ell_p$-SVM with $p > 1$.

**Keywords:** Support Vector Machines, Kernel functions, $\ell_p$-norms, Mathematical Optimization.

## 1. Introduction

In supervised classification, given a finite set of objects partitioned into classes, the goal is to build a mechanism, based on current available information, for classifying new objects into these classes. Examples of such techniques are Support Vector Machines (SVM), Classification Trees and $k$ Nearest Neighbours, among many others. In the last decades, SVM has become a popular methodology for supervised classification (Burges, 1998), due to their successful applications, as for instance in writing recognition (Bahlmann et al., 2002), evaluating insurance risks (Kascelan et al., 2016) or credit risk for lending (Harris, 2013) or detecting maligne/benigne tumors (Majid et al., 2014; Radhimeenakshi, 2016). SVM is a mathematical programming tool, that as other optimization-based approaches has helped to the successful development of supervised classification (see Bertsimas and Shioda, 2007, among others).

Support vector machine was originally developed by Vapnik (Vapnik, 1995, 1998) and Cortes and Vapnik (Cortes and Vapnik, 1995), and it consists in finding a hyperplane to separate a set of data into two classes, so that the distance from the hyperplane to the nearest point of each class is maximized. In order to do that, the standard SVM solves an

©2020 Víctor Blanco, Justo Puerto, and Antonio M. Rodríguez-Chía.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v21/18-601.html.

Advances in Data Analysis and Classification (2020) 14:175–199
https://doi.org/10.1007/s11634-019-00367-6

REGULAR ARTICLE

# Optimal arrangements of hyperplanes for SVM-based multiclass classification

**Víctor Blanco**[1] · **Alberto Japón**[2] · **Justo Puerto**[2]

Received: 22 October 2018 / Revised: 18 July 2019 / Accepted: 19 July 2019 / Published online: 26 July 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

In this paper, we present a novel SVM-based approach to construct multiclass classifiers by means of arrangements of hyperplanes. We propose different mixed integer (linear and non linear) programming formulations for the problem using extensions of widely used measures for misclassifying observations where the *kernel trick* can be adapted to be applicable. Some dimensionality reductions and variable fixing strategies are also developed for these models. An extensive battery of experiments has been run which reveal the powerfulness of our proposal as compared with other previously proposed methodologies.

**Keywords** Multiclass support vector machines · Mixed integer non linear programming · Classification, hyperplanes

**Mathematics Subject Classification** 62H30 · 90C11 · 68T05 · 32S22

## 1 Introduction

Support vector machine (SVM) is a widely-used methodology in supervised binary classification, firstly proposed by Cortes and Vapnik (1995). Given a number of observations with their corresponding labels, the SVM technique consists of finding a strip in the feature space so that each class is included in a different semispace maximizing the separation between classes (in a training sample) and minimizing some measure

✉ Víctor Blanco
vblanco@ugr.es

Alberto Japón
ajapon1@us.es

Justo Puerto
puerto@us.es

[1] IEMath-GR, Universidad de Granada, 18001 Granada, Spain

[2] IMUS, Universidad de Sevilla, 41012 Sevilla, Spain



European Journal of Operational Research 286 (2020) 84–100

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# Tightening big Ms in integer programming formulations for support vector machines with ramp loss

Marta Baldomero-Naranjo*, Luisa I. Martínez-Merino, Antonio M. Rodríguez-Chía

*Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Cádiz, Spain*

ARTICLE INFO

*Article history:*
Received 28 June 2019
Accepted 4 March 2020
Available online 19 March 2020

*Keywords:*
Location
Support vector machine
Ramp loss model
Mixed integer programming
Indicator constraints

ABSTRACT

This paper considers various models of support vector machines with ramp loss, these being an efficient and robust tool in supervised classification for the detection of outliers. The exact solution approaches for the resulting optimization problem are of high demand for large datasets. Hence, the goal of this paper is to develop algorithms that provide efficient methodologies to exactly solve these optimization problems. These approaches are based on three strategies for obtaining tightened values of the big M parameters included in the formulation of the problem. Two of them require solving a sequence of continuous problems, while the third uses the Lagrangian relaxation to tighten the bounds. The proposed resolution methods are valid for the $\ell_1$-norm and $\ell_2$-norm ramp loss formulations. They were tested and compared with existing solution methods in simulated and real-life datasets, showing the efficiency of the developed methodology.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

The goal of supervised classification is to provide a decision rule to classify the new individuals of a population in classes, based on the learning over a set of individuals already classified in those classes. The field of mathematical programming has resulted in a very useful tool for supervised classification, introduced by Mangasarian (1965, 1968). In particular, one of its most useful approaches are the support vector machines (SVM). Since their introduction by Cortes and Vapnik (1995) and Vapnik (1998), SVM have been applied in many fields, including: biology (Zhang, Ahn, Lin, & Park, 2006), medicine (Furey et al., 2000; Ren, 2012; Wang, Zheng, Yoon, & Ko, 2018), bioinformatics (Byvatov & Schneider, 2003; Chen, Yang, Feng, Ding, & Lin, 2017), machine vision (Barkana, Saricicek, & Yildirim, 2017; Decoste & Schölkopf, 2002), text classification (Zhang, Yoshida, & Tang, 2008), and finance (Gavrishchaka & Banerjee, 2006; Maldonado, Pérez, & Bravo, 2017; Min & Lee, 2005), among others.

In order to define a supervised classification problem more precisely, we consider the following notation. Given a set $N$ of individuals partitioned into two classes $\mathcal{Y} = \{-1, 1\}$, each individual $i \in N$ is associated with a pair $(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$, where $d$ is the number of features analyzed in each individual of $N$, $x_i$ contains the feature values, and $y_i$ provides the class membership (1 or -1). The aim of the SVM is to build a separating hyperplane $w \cdot x + b = 0$ to classify new individuals. The search for this hyperplane is based on a compromise between maximizing the distance (margin) between the two parallel hyperplanes supporting individuals of each class and minimizing the error caused by misclassified data. As proven in Carrizosa and Romero-Morales (2013), the maximization of the margin between the parallel hyperplanes associated with each class can equivalently be expressed as the minimization of $\|w\|^\circ$, (where $\|\cdot\|^\circ$ is the dual norm of the norm used to measure the distance to the hyperplanes in $\mathbb{R}^d$). In Blanco, Puerto, and Rodríguez-Chía (2020), the methodology applied to classical SVM (using $\ell_2$-norm) is extended to the general $\ell_p$-norm with $p \geq 1$. Although we will concentrate on cases $p = 1$ and $p = 2$, we provide the following common formulation for both cases that it is also valid for any value $p \geq 1$,

$$\text{(SVM-}\ell_p\text{)} \quad \min \ \frac{1}{p}(\|w\|_p)^p + C\sum_{i=1}^{n}\xi_i,$$
$$\text{s.t.} \ \ y_i\left(\sum_{k=1}^{d} w_k x_{ik} + b\right) \geq 1 - \xi_i, \quad i \in N, \tag{1}$$
$$\xi_i \geq 0, \quad i \in N, \tag{2}$$

where the $w$- and $b$-variables are the separation hyperplane coefficients and the $\xi$-variables represent classification errors. $C$ is a non-negative constant, $N$ is the set $\{1, \ldots, n\}$, and $\|w\|_p$ represents

* Corresponding author.
*E-mail addresses:* marta.baldomero@uca.es (M. Baldomero-Naranjo), luisa.martinez@uca.es (L.I. Martínez-Merino), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.ejor.2020.03.023
0377-2217/© 2020 Elsevier B.V. All rights reserved.

Discrete Applied Mathematics 296 (2021) 217–234

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

# A local analysis to determine all optimal solutions of *p*-*k*-max location problems on networks

Teresa Schnepper [a], Kathrin Klamroth [a], Justo Puerto [b], Michael Stiglmayr [a,*]

[a] School of Mathematics and Natural Sciences, University of Wuppertal, Gaußstraße 20, 42103 Wuppertal, Germany
[b] Mathematical Research Institute (IMUS), University of Seville, C/ Tarfia 21, 41012 Seville, Spain

ARTICLE INFO

*Article history:*
Received 28 May 2019
Received in revised form 9 April 2020
Accepted 12 May 2020
Available online 13 June 2020

*Keywords:*
Center problem on networks
Outlier
k-max function
Local analysis
Complete optimal set
Secondary objective

ABSTRACT

As a generalization of *p*-center location problems, *p*-*k*-max problems minimize the *k*th largest weighted distance to the customers. In this way, outlier facilities can be detected automatically and excluded from consideration when locating new facilities. Similar to *p*-center problems, *p*-*k*-max problems often have many alternative optimal solutions. Knowledge of the complete optimal set allows to select a most preferred solution using secondary criteria. In this paper, a general solution method is suggested that guarantees to find all alternative optimal solutions of *p*-*k*-max problems on networks. This is realized by performing a local analysis on the edges of the underlying graph and identifying edge segments on which the *p*-*k*-max function is linear. It is shown that the complete optimal set can be represented by an extended finite dominating set (FDS) which is of polynomial size for fixed values of *p*. Numerical tests indicate that computing the set of optimal solutions compared to the computation of a single optimal solution of a *p*-*k*-max problem requires on average less than 15% additional computing effort. This computational efficiency allows one to select the most preferred solution among them using secondary objectives, like backup coverage or the Weber function.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Bottleneck objective functions are in general very sensitive to outliers and data errors. In the context of location problems a small number of far-away customers, which we denote as *outliers*, can influence the solution of location problems such that the distance to the majority of customers is increased dramatically. In economically motivated situations this might lead to unfavorable solutions. Outliers in location problems are mainly not caused by data errors (in contrast to applications in data science) but by very heterogeneous distributions of costumers.

There has been extensive work on different techniques for handling outliers in location problems. As the influence of far-away facilities is most significant for center location problems, most approaches focus on this case. Continuous location problems with outliers are considered, e.g., in [1] for Euclidean distances, and in [2] for $l_\infty$-distances. In [4,5] so-called minquantile location problems are investigated, which can be considered a general case of *k*-max location problems. More general cases are considered in [30] (high dimensional problems) and in [6] (finite metric spaces). Discrete location models with outliers are discussed, e.g., in [7,28] and in [6].

* Corresponding author.
*E-mail addresses:* schnepper@math.uni-wuppertal.de (T. Schnepper), klamroth@math.uni-wuppertal.de (K. Klamroth), puerto@us.es (J. Puerto), stiglmayr@math.uni-wuppertal.de (M. Stiglmayr).

https://doi.org/10.1016/j.dam.2020.05.013
0166-218X/© 2020 Elsevier B.V. All rights reserved.

Computers and Operations Research 128 (2021) 105124

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/caor

# On the multisource hyperplanes location problem to fitting set of points

Víctor Blanco [a], Alberto Japón [b], Diego Ponce [b,c], Justo Puerto [b,*]

[a] IEMath-GR, Universidad de Granada, Spain
[b] IMUS, Universidad de Sevilla, Spain
[c] Dep. Métodos Estadísticos, Universidad de Zaragoza, Spain

ARTICLE INFO

*Article history:*
Received 4 February 2020
Revised 1 October 2020
Accepted 7 October 2020
Available online 21 October 2020

*Msc:*
52C35
90B85
90C11
90C30

*Keywords:*
Hyperplanes location
Mixed Integer Non Linear programming
Column generation

ABSTRACT

In this paper we study the problem of locating a given number of hyperplanes minimizing an objective function of the closest distances from a set of points. We propose a general framework for the problem in which norm-based distances between points and hyperplanes are aggregated by means of ordered median functions. A compact Mixed Integer Linear (or Non Linear) programming formulation is presented for the problem and also an extended set partitioning formulation with a huge number of variables is derived. We develop a column generation procedure embedded within a branch-and-price algorithm for solving the problem by adequately performing its preprocessing, pricing and branching. We also analyze geometrically the optimal solutions of the problem, deriving properties which are exploited to generate initial solutions for the proposed algorithms. Finally, the results of an extensive computational experience are reported. The issue of scalability is also addressed showing theoretical upper bounds on the errors assumed by replacing the original datasets by aggregated versions.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

Location Analysis deals with the determination of the *optimal* positions of facilities to satisfy the demand of a set of customers. The problems analyzed in the field are diverse but can be usually classified as: Discrete Location problems (DLP) and Continuous Location problems (CLP). In the first family, a set of potential facilities is previously given and the goal is to select, among them, the optimal ones under one or more criteria. The main tools for solving these problems come from Discrete Optimization, or more precisely, from Integer Linear Programming. In the second family of problems, the facilities have to be located in a continuous space and then, convex analysis and global optimization tools are needed to solve the problems. The most popular problem in the latter family is the Weber problem (Weber, 1909) in which a single point-facility has to be positioned on the plane so as to minimize the overall sum of the (Euclidean) distances to a set of (planar) demand points. The applications of both types of location problems are vast. DLP are more common in the location of *physical* facilities (as ATMs, supermarkets, stations, etc), while CLP are more useful when locating facilities in telecommunication networks (as wifi routers, servers, etc) or even to provide the sets of potential facilities for a DLP.

In this paper we study a problem that falls into the family of CLP. More specifically, we focus on the determination of *optimal* hyperplanes fitting a given finite set of demand points. The location of a single hyperplane is a classical problem that has been addressed in different fields. On the one hand, this problem clearly extends the classical Weber problem, but where instead of locating zero-dimensional facilities one looks for locating higher dimensional structures. On the other hand, in Statistics and Data Analysis, the determination of a hyperplane minimizing the sum of squares of vertical residuals is key for estimating a multivariate linear regression model using the Least Sum of Squares (LSS) method, credited to Gauss (1809). One can also find recent useful applications, both in Location Science and Data Analysis, for the problem of finding *optimal* hyperplanes fitting a set of points. For instance, Espejo and Rodríguez-Chía (2011) deals with the location of a rapid transit line on the plane to be used as an alternative transportation mean. Analogously, the widely used Support Vector Machine (SVM) methodology due to Cortes and Vapnik (1995), is also based on constructing a hyperplane minimizing certain loss functions of the distances to a given set of points.

Scanning the literature one can find that most of the attention has been devoted to finding hyperplanes with any of the following

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), ajapon1@us.es (A. Japón), dponce@us.es (D. Ponce), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.cor.2020.105124
0305-0548/© 2020 Elsevier Ltd. All rights reserved.

Expert Systems With Applications 178 (2021) 115017

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# A robust SVM-based approach with feature selection and outliers detection for classification problems

Marta Baldomero-Naranjo [a,*], Luisa I. Martínez-Merino [a,b], Antonio M. Rodríguez-Chía [a]

[a] Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Cádiz, Campus Universitario, Puerto Real, Cádiz, Spain
[b] Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Spain

ARTICLE INFO

*Keywords:*
Data science
Classification
Support vector machine
Outliers detection
Feature selection
Mixed integer programming

ABSTRACT

This paper proposes a robust classification model, based on support vector machine (SVM), which simultaneously deals with outliers detection and feature selection. The classifier is built considering the ramp loss margin error and it includes a budget constraint to limit the number of selected features. The search of this classifier is modeled using a mixed-integer formulation with big M parameters. Two different approaches (exact and heuristic) are proposed to solve the model. The heuristic approach is validated by comparing the quality of the solutions provided by this approach with the exact approach. In addition, the classifiers obtained with the heuristic method are tested and compared with existing SVM-based models to demonstrate their efficiency.

## 1. Introduction

Support vector machine (SVM) is a frequently used mathematical programming approach in supervised classification. It was introduced by Cortes et al. (1995) and Vapnik (1998). It has been applied to different fields such as finance (Gavrishchaka et al., 2006; Maldonado, Pérez, & Bravo, 2017; Min et al., 2005; Ghoddusi, Creamer, & Rafizadeh, 2019); machine vision (Decoste & Schölkopf, 2002; Rehman, Mahmud, Chang, Jin, & Shin, 2019); biology (Zhang, Ahn, Lin, & Park, 2006; Guerrero, Pajares, Montalvo, Romeo, & Guijarro, 2012; Joloudari, Saadatfar, Dehzangi, & Shamshirband, 2019); or sustainability (Mrówczynska, Sztubecka, Skiba, Bazan-Krzywoszanska, & Bejga, 2019; Kim, Costello, & Lee, 2019). More applications can be found in Cervantes, Garcia-Lamont, Rodríguez-Mazahua, and Lopez (2020).

Given a set $N := \{1, \ldots, n\}$ of individuals partitioned into two classes $\mathcal{Y} = \{-1, 1\}$, each individual $i \in N$ is associated with a pair $(x_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$, where $d$ is the number of features analyzed over each individual of $N$, $x_i$ contains the feature values, and $y_i$ provides the class membership, 1 or $-1$. The aim of SVM is to find a separating hyperplane $w \cdot x + b = 0$ to classify new individuals by finding the balance between maximizing the distance (margin) between the two parallel hyperplanes supporting individuals of each class and minimizing the error of misclassified data. See Bradley, Mangasarian, and Street (1998) and Blanco, Puerto, and Rodríguez-Chía (2020) for the use of SVM model considering different norms to measure the margin.

Although classical SVM models have high predictive power in comparison with other state-of-the-art classifying methods, some drawbacks also arise from their use. As stated in Hastie, Tibshirani, and Friedman (2009), classical SVMs are not robust against outliers and they do not perform efficiently when the analyzed dataset contains a large number of irrelevant features. In this paper we introduce a model which deals with both aspects: outliers detection and feature selection.

In this context, an outlier is an individual whose feature values are not similar from the corresponding ones of most individuals that belong to the same class. The presence of outliers in real-life data is common and is caused by several factors which include but are not limited to: transcription errors, measurement errors, mislabelling, data processing errors, intentional instances (e.g. spam). There are many papers in the literature that use an SVM based model to address these issues, see for instance, Xu, Cao, Hu, and Principe (2017), Zhang, Meratnia, and Havinga (2009), Lukashevich, Nowak, and Dunker (2009) and Brooks (2011) and Blanco, Japón, and Puerto (2020a).

Besides, Karami (2018) also deals with outliers detection. Specifically, the mentioned paper considers a novel anomaly-based intrusion detection system in presence of long-range independence data called benign outliers and also provides visualized information and insights to end users. This type of methodology is particularly suitable in the abnormal network traffic analysis which is an important research topic

* Corresponding author.
*E-mail addresses:* marta.baldomero@uca.es (M. Baldomero-Naranjo), luisa.martinez@uca.es (L.I. Martínez-Merino), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.eswa.2021.115017
Received 29 May 2020; Received in revised form 10 October 2020; Accepted 6 April 2021
Available online 20 April 2021
0957-4174/© 2021 Elsevier Ltd. All rights reserved.

ARTICLE IN PRESS
JID: EOR [m5G;June 19, 2021;21:47]
European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# A branch-and-price procedure for clustering data that are graph connected

Stefano Benati [a], Diego Ponce [b], Justo Puerto [b,*], Antonio M. Rodríguez-Chía [c]

[a] Dipartimento di Sociologia e Ricerca Sociale, Università di Trento, Italy
[b] IMUS and Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Spain
[c] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain

ARTICLE INFO

*Article history:*
Received 28 May 2020
Accepted 25 May 2021
Available online xxx

*Keywords:*
Combinatorial optimization
Clustering
Mixed integer programming
Branch-and-price

ABSTRACT

This paper studies the Graph-Connected Clique-Partitioning Problem (GCCP), a clustering optimization model in which units are characterized by both individual and relational data. This problem, introduced by Benati, Puerto, and Rodríguez-Chía (2017) under the name of Connected Partitioning Problem, shows that the combination of the two data types improves the clustering quality in comparison with other methodologies. Nevertheless, the resulting optimization problem is difficult to solve; only small-sized instances can be solved exactly, large-sized instances require the application of heuristic algorithms. In this paper we improve the exact and the heuristic algorithms previously proposed. Here, we provide a new Integer Linear Programming (ILP) formulation, that solves larger instances, but at the cost of using an exponential number of variables. In order to limit the number of variables necessary to calculate the optimum, the new ILP formulation is solved implementing a branch-and-price (B&P) algorithm. The resulting pricing problem is itself a new combinatorial model: the Maximum-weighted Graph-Connected Single-Clique problem (MGCSC), that we solve testing various Mixed Integer Linear Programming (MILP) formulations and proposing a new fast "random shrink" heuristic. In this way, we are able to improve the previous algorithms: The B&P method outperforms the computational times of the previous MILP algorithms and the new random shrink heuristic, when applied to GCCP, is both faster and more accurate than the previous heuristic methods. Moreover, the combination of column generation and random shrink is itself a new MILP-relaxed matheuristic that can be applied to large instances too. Its main advantage is that all heuristic local optima are combined together in a restricted MILP, consisting in the application of the exact B&P method but solving heuristically the pricing problem.

© 2021 Published by Elsevier B.V.

## 1. Introduction

We consider a clustering problem in which units are characterized by both individual and relational data. Individual data take the form of a matrix F of $n$ rows, representing units, and $m$ columns, representing features that are measured for individuals. Individual data are then complemented by relational data, for example representing friendship, communication, co-participation and so on. Relational data are described as an undirected graph $G = (V, E)$ in which $V$ are the units, $|V| = n$, and there is an edge $e_{ij} \in E$ if and only if there is a relation between $i, j \in V$. The data structure that combines the graph $G = (V, E)$ with the data matrix $F$ forms the triplet $G = (V, E, F)$, called *attributed graph*, see Bothorel, Cruz, Magnani, and Micenkova (2015).

The simplest method of clustering attributed graphs is projecting the relational data into the individual data, or vice versa, the individual into the relational. In the former case, a dissimilarity measure between units $i$ and $j$ is calculated using *both* individual measures of F *and* the existence/non-existence of an arc $i, j$. Combe, Largeron, Egyed-Zsigmond, and Géry (2012), Cheng, Zhou, Huang, and Yu (2012). In the latter case, the matrix F is used to calculate a distance $d_{ij}$ attached to an existing arc $e_{ij}$, and to convert the *unweighted* graph into a *weighted* one, Neville, Adler, and Jensen (2003). In both cases, the problem is reduced to standard clustering or graph partitioning problems respectively, and any solution methods for those problems can be applied. The interested reader is referred to Gambella, Ghaddar, and Naoum-Sawaya (2021) for a recent survey. Alternatively, the two data structures are kept separate and then one can formulate an optimization model to determine the best classification. The optimization model

* Corresponding author.
*E-mail addresses:* stefano.benati@unitn.it (S. Benati), dponce@us.es (D. Ponce), puerto@us.es (J. Puerto), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.ejor.2021.05.043
0377-2217/© 2021 Published by Elsevier B.V.

Please cite this article as: S. Benati, D. Ponce, J. Puerto et al., A branch-and-price procedure for clustering data that are graph connected, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2021.05.043

# Optimización combinatoria: 5 public.

European Journal of Operational Research 262 (2017) 863–878

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

## Minimum Spanning Trees with neighborhoods: Mathematical programming formulations and solution methods

Víctor Blanco[a,*], Elena Fernández[b], Justo Puerto[c]

[a] Department of Quantitative Methods for Economics and Business, Universidad de Granada, 18011 Granada, Spain
[b] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
[c] Department of Statistics and Operations Research, Universidad de Sevilla, 41012 Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 10 November 2016
Accepted 8 April 2017
Available online 15 April 2017

*Keywords:*
Combinatorial Optimization
Minimum Spanning Trees
Neighborhoods
Mixed Integer Non Linear Programming
Second order cone programming,

ABSTRACT

This paper studies Minimum Spanning Trees under incomplete information assuming that it is only known that vertices belong to some neighborhoods that are second order cone representable and distances are measured with a $\ell_q$-norm. Two Mixed Integer Non Linear mathematical programming formulations are presented, based on alternative representations of subtour elimination constraints. A solution scheme is also proposed, resulting from a reformulation suitable for a Benders-like decomposition, which is embedded within an exact branch-and-cut framework. Furthermore, a mathheuristic is developed, which alternates in solving convex subproblems in different solution spaces, and is able to solve larger instances. The results of extensive computational experiments are reported and analyzed.

© 2017 Elsevier B.V. All rights reserved.

### 1. Introduction

Nowadays Combinatorial Optimization (CO) lies in the heart of multiple applications in the field of Operations Research. Many such applications can be formulated as optimization problems defined on graphs where some particular structure is sought satisfying some optimality property. Traditionally this type of problems assumed implicitly the exact knowledge of all input elements, and, in particular, of the precise position of vertices and edges. Nevertheless, this assumption does not always hold, as uncertainty, lack of information, or some other factors may affect the relative position of the elements of the input graph. Hence, new tools are required to give adequate answers to these challenges, which have been often ignored by standard CO tools.

A matter that, in this context, has attracted the interest of researchers over the last years is the solution of certain CO problems when the exact position of the vertices of the underlying graph is not known with certainty. If probabilistic information is available, then stochastic programming tools can be used, and optimization over expected values carried out. Moreover, even under the assumption of incomplete information one could use a uniform distribution and still apply such an approach. However, the use of probabilistic information and allowing to consider all possible locations for the vertices is not always suitable. For instance, when a unique representative associated with each point of the input graph must be determined. Scanning the related literature one can find papers applying both methodologies. Examples of stochastic approaches are for instance Bertsimas and Howell (1993) or Frank (1969). Examples of the second type of approach arise in variants of the traveling salesman problem (TSP), Minimum Spanning Tree (MST), or facility location problems that deal with *demand regions* instead of *demand points* (see Arkin and Hassin, 1994; Brimberg and Wesolowsky, 2002; Cooper, 1978; Dror, Efrat, Lubiw, and Mitchell, 2003; Juel, 1981; Nickel, Puerto, and Rodríguez-Chía, 2003; Yang, Lin, Xu, and Xie, 2007, to mention just a few).

A relevant common question raised by the latter class of problems is how to model and solve optimization problems on graphs when vertices are not points but regions in a given domain. The above mentioned case of the TSP, first introduced by Arkin and Hassin (1994, 2000), has been addressed recently by a number of authors. It generalizes the Euclidean TSP and the group Steiner tree problem, and has applications in VLSI-design and other routing problems, in which there exist constraints on the position of the vertices. Several inapproximability results and approximation algorithms have been developed for particular cases. The case of the spanning tree problem with neighborhoods (MSTN) was first addressed by Yang et al. (2007), who proved that the general case of the problem in the plane is NP-hard (result also reproved by Löffler & van Kreveld, 2010), and gave several approximation algorithms and a PTAS for the particular case of disjoint unit disks in the plane. Some extensions considering the maximization of the

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), e.fernandez@upc.edu (E. Fernández), puerto@us.es (J. Puerto).

http://dx.doi.org/10.1016/j.ejor.2017.04.023
0377-2217/© 2017 Elsevier B.V. All rights reserved.

Optim Lett (2017) 11:667–678
DOI 10.1007/s11590-016-1047-z

ORIGINAL PAPER

## Rank aggregation in cyclic sequences

Eva M. García-Nové[1] · Javier Alcaraz[1] · Mercedes Landete[1] · Juan F. Monge[1] · Justo Puerto[2]

Received: 8 October 2015 / Accepted: 21 May 2016 / Published online: 10 June 2016
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** In this paper we propose the problem of finding the cyclic sequence which best represents a set of cyclic sequences. Given a set of elements and a precedence cost matrix we look for the cyclic sequence of the elements which is at minimum distance from all the ranks when the permutation metric distance is the Kendall Tau distance. In other words, the problem consists of finding a robust cyclic rank with respect to a set of elements. This problem originates from the Rank Aggregation Problem for combining different linear ranks of elements. Next, we also introduce the problem of finding the cyclic sequence with minimum expected cost with respect to a probability measure based on dissimilarity between cyclic sequences on the Kendall Tau metric. Finally, we establish certain relationships among some classical problems and the new problems that we have proposed.

**Keywords** Linear ordering problem · Rank aggregation problem

### 1 Introduction

There is a vast number of combinatorial problems which look for the best permutation of a set of elements assuming a certain criterion. If the optimal permutation is the one which minimizes the cost of the path induced by the permutation we have the Linear Ordering Problem (LOP) [17–19] and when the optimal permutation is the closest to a given set of permutations and the distances among permutations are measured with the

✉ Mercedes Landete
landete@umh.es

[1] Department of Statistics, Mathematics and Computer Science, Center of Operations Research, Miguel Hernández University of Elche, Elche, Spain

[2] IMUS, University of Seville, Seville, Spain



European Journal of Operational Research 270 (2018) 968–981

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

## A minmax regret version of the time-dependent shortest path problem☆

Eduardo Conde*, Marina Leal, Justo Puerto

*Department of Statistics and Operations Research, Faculty of Mathematics. Campus Reina Mercedes, University of Seville, Seville 41011, Spain*

ARTICLE INFO

*Article history:*
Received 7 April 2017
Accepted 13 April 2018
Available online 21 April 2018

*Keywords:*
Integer Programming
Minmax-regret models
Benders decomposition

ABSTRACT

We consider a shortest path problem where the arc costs depend on their relative position on the given path and there exist uncertain cost parameters. We study a minmax regret version of the problem under different types of uncertainty of the involved parameters. First, we provide a Mixed Integer Linear Programming (MILP) formulation by using strong duality in the uncertainty interval case. Second, we develop three algorithms based on Benders decomposition for general polyhedral sets of uncertainty. In order to make the algorithms faster we use constant factor approximations to initialize them. Finally, we report some computational experiments for different uncertainty sets in order to compare the behavior of the proposed algorithms.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

Certain real applications require deciding the *best* schedule plans in which the cost of processing a task depends on its starting time. In many cases, we are given a set of planning tasks, usually represented as the arcs of a network that may model precedence relationships between them, while the goal is to determine the instant in which each task (arc) should start its processing (is traversed) in order to optimize some objective function. Natural examples are timetabling and scheduling problems (Borndörfer & Schlechte, 2007; Cacchiani, Caprara, & Toth, 2010; Cacchiani & Toth, 2012; Fischer & Helmberg, 2014; Gawiejnowicz, 2008), Shortest Path Problems (SPP) in dynamic networks (Ahuja, Orlin, Pallotino, & Scutella, 2003; Aronson, 1989; Cooke & Halsey, 1966; Fischer & Helmberg, 2014; Hashemi, Mokarami, & Nasrabadi, 2010) or Time-Dependent Traveling Salesman Problems (TDTSP) (Bigras, Gamache, & Savard, 2008; Furini, Persiani, & Toth, 2016; Pessoa, Uchoa, & Freitas, 2010; Picard & Queyranne, 1978; Tas, Gendreau, Jabali, & Laporte, 2016; Wander, Wiel, & Sahinidis, 1996), among others.

An interesting type of time-dependency that we would like to emphasize here and that appears in some of the above TDTSP models, (Bigras et al., 2008; Picard & Queyranne, 1978; Wander et al., 1996), is that of defining the travel cost between the two nodes of any arc as a function of the order in which the traveler moves from one of them to the other into the route. More specifically, the travel cost from node $i$ to node $j$ depends on the position that the node $i$ occupies in the ordering sequence of visited nodes. This problem with discrete travel times can be viewed as a single machine scheduling problem with sequence-dependent setup times (Bigras et al., 2008; Picard & Queyranne, 1978).

The time-dependent SPP (TDSPP) has been considered under different optimality criteria and although there are models that can be solved in polynomial time (Ahuja et al., 2003), their complexity level is usually very high. The size of an associated network called sometimes *time-expanded network* (Fischer & Helmberg, 2014), *multipartite network* (Bigras et al., 2008) or *layered graph* (Abeledo, Fukasawa, Pessoa, & Uchoa, 2013) plays an important role in this level of complexity. This network is the basis of one of the most useful modeling approaches for discrete planning. Depending on the number of arcs and discretized time periods, the full time-expanded network easily grows to a huge number of arcs in actual applications. In order to handle this behavior, Fischer and Helmberg (2014), developed a general dynamic graph generation framework to control the size of the network, what is crucial in order to obtain efficient algorithms to solve exactly the corresponding problem.

In this paper, we use a time-expanded network as the basis to obtain a Linear Programming (LP) formulation of time-dependent SPP models where the cost structure takes into account the position that each arc occupies in the path, specifically, the total cost associated to each arc is the sum of an amount proportional to the position of the arc in the path and another quantity associated to

☆ This research has been funded by the Spanish Ministry of Science and Technology projects MTM2013-46962-C2-1-P, MTM2016-74983-C2-1-R and the Regional Government of Andalusia project FQM331.
* Corresponding author.
*E-mail address:* educon@us.es (E. Conde).

https://doi.org/10.1016/j.ejor.2018.04.030
0377-2217/© 2018 Elsevier B.V. All rights reserved.

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

WILEY

INTERNATIONAL TRANSACTIONS IN OPERATIONAL RESEARCH

Intl. Trans. in Op. Res. 28 (2021) 48–69
DOI: 10.1111/itor.12680

# Computational comparisons of different formulations for the Stackelberg minimum spanning tree game

Martine Labbé[a,b], Miguel A. Pozo[c,*] and Justo Puerto[c]

[a] *Faculty of Sciences, Computer Science Department, Université Libre de Bruxelles, Brussels, Belgium*
[b] *INOCS Team, INRIA, Lille, France*
[c] *Facultad de Matemáticas, Department of Statistics and Operational Research, University of Seville, Sevilla, Spain*
*E-mail: mlabbe@ulb.ac.be [Labbé]; miguelpozo@us.es [Pozo]; puerto@us.es [Puerto]*

Received 26 November 2018; received in revised form 14 May 2019; accepted 14 May 2019

**Abstract**

Let $G = (V, E)$ be a given graph whose edge set is partitioned into a set $R$ of red edges and a set $B$ of blue edges, and assume that red edges are weighted and contain a spanning tree of $G$. Then, the Stackelberg minimum spanning tree game (StackMST) is that of pricing (i.e., weighting) the blue edges in such a way that the total weight of the blue edges selected in a minimum spanning tree of the resulting graph is maximized. In this paper, we present different new mathematical programming formulations for the StackMST based on the properties of the minimum spanning tree problem and the bilevel optimization. We establish a theoretical and empirical comparison between these new formulations that are able to solve random instances of 20–70 nodes. We also test our models on instances in the literature, outperforming previous results.

*Keywords:* minimum spanning tree; bilevel optimization; Stackelberg game

## 1. Introduction

Let $G$ be a given graph whose edge set is partitioned into a set of red edges and a set of blue edges, and assume that red edges are weighted and contain a spanning tree of $G$. Then, the Stackelberg minimum spanning tree game (StackMST) consists in pricing (i.e., weighting) the blue edges in such a way that the total weight of the blue edges selected in a minimum spanning tree (MST) of the resulting graph is maximized.

The StackMST can be seen as a bilevel optimization problem where the second level is a minimum spanning tree problem (MSTP) and where objective functions are bilinear at both levels. Optimization problems related to spanning trees, or simply spanning tree problems, are among the core problems in combinatorial optimization. On the one hand, the combinatorial object that represents

*Corresponding author.

© 2019 The Authors.
International Transactions in Operational Research © 2019 International Federation of Operational Research Societies
Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA02148, USA.

Computers and Operations Research 131 (2021) 105208

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

# Solving binary-constrained mixed complementarity problems using continuous reformulations

Steven A. Gabriel[a,b,1], Marina Leal[c,1], Martin Schmidt[d,*,1]

[a] *University of Maryland, Department of Mechanical Engineering/Applied Mathematics & Statistics, and Scientific Computation Program, College Park, MD, USA*
[b] *Norwegian University of Science and Technology, Trondheim, Norway*
[c] *Miguel Hernández University, Department of Statistics and Operations Research, Elche, Spain*
[d] *Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany*

ARTICLE INFO

*MSC:*
90Bxx
90Cxx
90C33
91B50

*Keywords:*
Binary-constrained mixed complementarity problems
Mixed-integer optimization
Continuous reformulations
Spatial price equilibrium problems

ABSTRACT

Mixed complementarity problems are of great importance in practice since they appear in various fields of applications like energy markets, optimal stopping, or traffic equilibrium problems. However, they are also very challenging due to their inherent, nonconvex structure. In addition, recent applications require the incorporation of integrality constraints. Since complementarity problems often model some kind of equilibrium, these recent applications ask for equilibrium points that additionally satisfy certain integer conditions. Obviously, this makes the problem even harder to solve. The solution approach used most frequently in the literature is to recast the complementarity conditions as disjunctive constraints using additional binary variables and big-$M$ constraints. However, both latter aspects create issues regarding the tractability and correctness of the reformulation. In this paper, we follow the opposite route and restate the integrality conditions as complementarity constraints, leading to purely continuous reformulations that can be tackled by local solvers. We study these reformulations theoretically and provide a numerical study that shows that continuous reformulations are useful in practice both in terms of solution times and solution quality.

## 1. Introduction

Many real-world engineering-economic systems such as energy markets, transportation networks, or supply chains—to name just three examples—involve models with multiple, competing agents. If the principle of symmetry holds, i.e., the Jacobian of the mixed complementarity problem (MCP) function is symmetric, these systems can be modeled as a single optimization problem (Gabriel et al., 2012). In this case, the assumption is that what is best for the system is also best for each participant and vice versa. For instance, in micro-economics, this is the case under the assumption of perfectly competitive markets. However, in many cases this is not realistic and, thus, a separate optimization problem for each agent is required and additional equilibrating conditions are needed to couple the optimal, self-interested decisions of every player. This situation usually leads to variational inequalities or MCPs; see, e.g., Gabriel et al. (2012).

The main assumption for obtaining an MCP derived from the concatenation of separate optimization problems (and possibly additional equilibrating conditions) is that for each of the separate optimization problems, the Karush–Kuhn–Tucker (KKT) conditions (Bazaraa et al., 2013) are both necessary and sufficient for global optimality. This usually means that the considered optimization problems are convex and that a suitable constraint qualification is satisfied. However, many important engineering-economic systems involve nonconvex models due to the modeling of the inherently nonconvex functions as part of the underlying physics (e.g., power flows in electricity (Schweppe et al., 2013) or the Weymouth equation in natural gas networks (Fügenschuh et al., 2015)) or involve integer variables (e.g., the unit commitment problem in power system dispatch models (O'Neill et al., 2005)).

The resulting nonconvexity of the separate optimization problems of the players makes solving them simultaneously much harder. This difficulty is due to the fact that, most often in these cases, the KKT conditions may not be necessary or sufficient for global optimality. Thus, their concatenation does not define a valid equilibrium anymore. One heuristic approach that often is applied in situations in which the nonconvexity is caused by integrality conditions in an underlying optimization problem is the following. First, all integer variables are relaxed, i.e., the continuous relaxation of the mixed-integer problem is considered. The optimal solutions of these relaxations can then be characterized by their KKT conditions assuming a constraint qualification holds. Finally, the integrality constraints are re-added to the overall resulting MCP derived from the concatenation of all KKT conditions (as well as possibly other equilibrating conditions).

* Corresponding author.
*E-mail addresses:* sgabriel@umd.edu (S.A. Gabriel), m.leal@uv.es (M. Leal), martin.schmidt@uni-trier.de (M. Schmidt).
[1] All authors equally contributed to all parts of the paper.

https://doi.org/10.1016/j.cor.2020.105208
Received 16 February 2020; Received in revised form 25 November 2020; Accepted 29 December 2020
Available online 7 January 2021
0305-0548/© 2020 Elsevier Ltd. All rights reserved.

# Transporte: 8 publicaciones

Transportation Research Part B 98 (2017) 94–112

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

## Multi-objective integration of timetables, vehicle schedules and user routings in a transit network

Gilbert Laporte[a], Francisco A. Ortega[b], Miguel A. Pozo[c,*], Justo Puerto[d]

[a] HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montreal, H3T 2A7, Canada
[b] Department of Applied Mathematics I, University of Seville, Spain
[c] Department of Statistics and Operational Research, University of Seville, Spain
[d] Department of Statistics and Operational Research, University of Seville, Spain

ARTICLE INFO

*Article history:*
Received 13 November 2015
Revised 28 July 2016
Accepted 29 December 2016
Available online 2 January 2017

*Keywords:*
Timetabling
Vehicle scheduling
Schedule delay
Location-allocation
Multi-objective optimization

ABSTRACT

The Transit Network Timetabling and Scheduling Problem (TNTSP) aims at determining an optimal timetable for each line of a transit network by establishing departure and arrival times at each station and allocating a vehicle to each timetable. The current models for the planning of timetables and vehicle schedules use the *a priori* knowledge of users' routings. However, the actual route choice of a user depends on the timetable. This paper solves the TNTSP in a public transit network by integrating users' routings in the model. The proposed formulation guarantees that each user is allocated to the best possible timetable, while satisfying capacity constraints. In addition, we perform a trade-off analysis by means of a multi-objective formulation which jointly optimizes the operator's and the users' criteria.

© 2016 Elsevier Ltd. All rights reserved.

### 1. Introduction

The Transit Network Timetabling and Scheduling Problem (TNTSP) aims at determining an optimal timetable for each line of a transit network by establishing departure and arrival times at each station and allocating a vehicle to each timetable (see e.g. Guihaire and Hao, 2008a). The input data consist of a public transportation network (PTN) made up of a set of stations and links between them, a set of lines, a fleet of capacitated vehicles, a fixed budget for line runs and an origin/destination (O/D) matrix. The output data consist of a set of arrival and departure times at the stations for the vehicles. Several criteria can be taken into consideration in these problems, for example waiting times, short transfers, fleet size, travel time, load factor, users utilization and, in general terms, users' and operator's costs. Traditionally, the Transit Network Timetabling Problem (TNTP, see e.g. Cacchiani and Toth, 2012) has been studied as a preliminary step for the Vehicle Scheduling Problem (VSP, see e.g. Bunte and Kliewer, 2009), i.e. the TNTP output is an input for the VSP. Unfortunately, this approach leads to suboptimal solutions for the TNTSP. In this paper we solve the TNTSP by integrating the TNTP and VSP.

In this paper we assume that users are interested in minimizing their *scheduled delay* (see, e.g., Small, 1982; de Palma and Lindsey, 2001; Mesa et al., 2014) that can be defined as the deviation between desired departure and arrival times and actual ones. When the travel time is constant, the actions of boarding late or arriving early do not incur any inconvenience themselves, but for many users boarding early involves some wasted time and arriving late typically has even more severe

* Corresponding author.
*E-mail addresses:* gilbert.laporte@cirrelt.ca (G. Laporte), riejos@us.es (F.A. Ortega), miguelpozo@us.es (M.A. Pozo), puerto@us.es (J. Puerto).

http://dx.doi.org/10.1016/j.trb.2016.12.018
0191-2615/© 2016 Elsevier Ltd. All rights reserved.

Downloaded from informs.org by [128.227.24.141] on 25 June 2018, at 06:49 . For personal use only, all rights reserved.

informs
http://pubsonline.informs.org/journal/trsc/

TRANSPORTATION SCIENCE
*Articles in Advance*, pp. 1–16
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

## On-Line Timetable Rescheduling in a Transit Line

**Francisco A. Ortega,[a] Miguel A. Pozo,[b] Justo Puerto[c]**

[a] Department of Applied Mathematics I, University of Seville, Seville 41012, Spain; [b] Department of Applied Mathematics II, University of Seville, Seville 41092, Spain; [c] Department of Statistics and Operational Research, University of Seville, Seville 41092, Spain

**Contact:** riejos@us.es, http://orcid.org/0000-0002-3523-796X (FAO); miguelpozo@us.es, http://orcid.org/0000-0002-5798-6558 (MAP); puerto@us.es, http://orcid.org/0000-0003-4079-8419 (JP)

**Received:** May 4, 2016
**Revised:** January 5, 2017; August 3, 2017
**Accepted:** August 28, 2017
**Published Online in Articles in Advance:** June 21, 2018

**https://doi.org/10.1287/trsc.2017.0807**

**Copyright:** © 2018 INFORMS

**Abstract.** Public transportation systems in metropolitan areas carry a high density of daily traffic, heterogeneously distributed, and exposed to the negative consequences derived from service disruptions. Breakdowns, accidents, strikes, etc., require on-line operation adjustments to address these incidents and thus reduce their side effects, such as passenger extra-waiting times, complaints, potential operational dangers, etc. The Vehicle Rescheduling Problem consists of defining a new schedule for a set of previously scheduled trips, given that one/several trips cannot be carried out. This paper addresses the rescheduling problem in a transit line that has suffered a fleet size reduction (also denoted as Reduced Fleet Rescheduling Problem). We present different modeling possibilities depending on the assumptions that must be included in the modelization and we show that the problem can be rapidly solved using a reformulation that will be proven to have the integrality property. We test our results in a testbed of random instances outperforming previous results in the literature. We also include a real-world case of the commuter trains of Madrid, Spain to illustrate our solutions.

**Funding:** This research has been partially supported by the Spanish Ministry of Science and Education through [grants MTM2013-46962-C02-01 (MICINN, Spain) MTM2015-67706-P (MINECO/FEDER, UE) and MTM2016-74983-C2-1-R].

**Keywords:** urban transportation network • disturbance management • real time rescheduling

### 1. Introduction

Timetable design is a central problem in transportation planning including many interfaces with other classical problems, e.g., line planning, vehicle scheduling, and vehicle rescheduling. The Transit Network Timetabling and Scheduling Problem (TNTSP) is devoted to obtaining and optimizing departure and arrival times for each line run (i.e., trip from an origin to a final destination) to and from each station over a planning horizon imposing/optimizing different constraints and objectives (Ibarra-Rojas, Giesen, and Ríos-Solís 2014). The TNTSP is based on the following general input: An infrastructure of a transport system described by a node set (network stations) and an edge set (i.e., roads/tracks between adjacent stations), a trip demand matrix between pairs of nodes of the infrastructure, a set of transit lines with associated frequencies that have already been determined to satisfy such trip demand and, finally, a vehicle fleet with specific characteristics. The objective of the TNTSP is finding the arrival and departure times of each vehicle at each station such that the demand satisfaction, required fleet size, and vehicle capacities can be optimized.

Accidents, strikes, and other sources of vehicle delays or cancellations may force modification of an initial timetable when vehicles in some sections cannot run according to the initial planning. Here disturbances are relatively small perturbations of the transit network that can be handled by modifying the timetable, but without modifying the duties of vehicles and drivers. Complementarily, disruptions are relatively large incidents requiring modification of the timetable and vehicle duties. There are several examples of possible disruptions that require the rescheduling of vehicles. Examples include: (1) interruptions coming from severe weather conditions, accidents, and the blockage of road or track sections (Cacchiani et al. 2012) or (2) fleet size reductions coming from vehicle breakdowns, drivers' and crew strikes (van Exel and Rietveld 2001) or vehicle reallocations made to reinforce other sections of the transit network (Burdett and Kozan 2009). A timetable may also become infeasible simply due to a heavy passenger flow (Mesa, Ortega, and Pozo 2009b).

To address these incidents, on-line operation adjustments are required to reduce their side effects. The Vehicle Rescheduling Problem (VRP) consists of defining a new timetable for a set of previously scheduled trips, given that one/several trips cannot be carried out. While many objectives and constraints remain from the timetabling problem, new requirements and objectives arise in this context. In terms of passenger transportation, the main decisions concern minimization of the deviations from the initial timetable in operation. This is done by cancelling some services and/or providing new reference times for some



RIGHTSLINK



## RAILWAY TRAFFIC DISTURBANCE MANAGEMENT BY MEANS OF CONTROL STRATEGIES APPLIED TO OPERATIONS IN THE TRANSIT SYSTEM

FRANCISCO A. ORTEGA[1], JUAN A. MESA[2], MIGUEL A. POZO[2] & JUSTO PUERTO[3]
[1]Higher Technical School of Architecture, Universidad de Sevilla, Spain.
[2]Higher Technical School of Engineering, Universidad de Sevilla, Spain.
[3]Faculty of Mathematics, Universidad de Sevilla, Spain.

ABSTRACT

Railway systems in metropolitan areas support a high density of daily traffic that is exposed to different types of disturbances in the service. An interesting topic in the literature is to obtain action protocols in the presence of contingencies which can affect the system operation, avoiding the propagation of perturbation and minimizing its negative consequences.

Assume that, with a small margin of time (e.g. one day), the decision-maker of the transportation network is knowing that a part of the train fleet will become inoperative temporarily along a specific transit line and none additional vehicle will be able to restore the affected services. The decision to be taken in consequence will require to reschedule the existing services by possibly reducing the number of expeditions (line runs). This will affect travellers who regularly use the transit system to get around.

Consider that the decision-maker aims to lose the least number of passengers as a consequence of having introduced changes into the transit line. A strategy that could be applied in this context is to remove those line runs which are historically less used by travellers without affecting the remaining services. Another alternative strategy might be to reschedule the timetables of the available units, taking into account the pattern of arrivals of users to the boarding stations and the user behavior during waiting times (announced *in situ*).

The aim of this work consists of assessing the strategy of train rescheduling along the current transportation line when the supply must be reduced in order to reinforce the service of another line, exploited by the same public operator, which has suffered an incidence or emergency.

*Keywords: disruption management, railways, timetable rescheduling.*

### 1 INTRODUCTION

Planning of public transport systems includes decisions about the design of lines and service frequency that seek a balance between supply and demand in a given context and with a medium-long time horizon. The occasional failures in the operation of the system are usually not considered in this initial planning stage; therefore, a system, even properly designed, may be unavailable for unforeseen occurrences (such as accidents), or expected episodes of peak demand as a consequence of massive events.

In order to reduce the disruptive effect of these imbalances, the operator of a passenger transport system must be able to implement some control strategy to adjust schedules to traffic conditions in real time. The most commonly control strategies that are implemented for this purpose are express service (to skip stations with little flow of passengers), deadheading (to skip all stations during the returning trip), short-turning [1] (to select a portion of the fleet in order to serve, by means of short cycles, those route sections exhibiting high demand) and a combination of them [2–4].

On the other hand, if the incident happens in another transit line managed by the same operator, this decision-maker may additionally choose to reallocate its global fleet of vehicles, in

© 2018 WIT Press, www.witpress.com
ISSN: 2058-8305 (paper format), ISSN: 2058-8313 (online), http://www.witpress.com/journals
DOI: 10.2495/TDI-V2-N4-362-372

*Francisco A. Ortega et al., Int. J. Transp. Dev. Integr., Vol. 3, No. 4 (2019) 306–316*

# A MATHEURISTIC FOR OPTIMIZING SKIP–STOP OPERATION STRATEGIES IN RAIL TRANSIT LINES

FRANCISCO A. ORTEGA[1], JUAN A. MESA[2], RAMÓN PIEDRA-DE-LA-CUADRA[3], MIGUEL A. POZO[4]
[1]Higher Technical School of Architecture, Universidad de Sevilla, Spain
[2]Higher Technical School of Engineering, Universidad de Sevilla, Spain
[3]Department of Mathematics, Universidad de Cordoba, Spain
[4]Faculty of Mathematics, Universidad de Sevilla, Spain

ABSTRACT

In many transit systems, operators use skip–stop strategies to reduce travel time of particular train services by not stopping (skipping) at less densely populated stations. This decision of omitting some stops reduces the travel time for the users within the vehicle and increases the speed of operation, favouring the provision of new transit services where are more necessary. In this work, the best A/B stop–skip patterns for a set of transit services along a railway corridor are determined by means a three-phase methodology that includes the formulation of a nonlinear integer programming inspired in the multiple knapsack problem and the application of a heuristic algorithm based on mathematical properties (matheuristic).

*Keywords: matheuristics, optimization, railways, stop–skip strategy.*

## 1 INTRODUCTION

The skip–stop (or limited-stop) mechanism is a strategy that some transport companies follow to reduce travel times without the need to increase the fleet of vehicles. It consists of privileging a majority of passengers by offering shorter travel times, as a result of having previously selected a group of low-activity stations, where the trains will not stop to leave/pick up passengers.

Distinguishing between express and local stations appears first in the Northwestern Elevated of Chicago by July 1900. The skip–stop service was also first developed for the Chicago Metro system in 1947, and later implemented in Philadelphia and New York (see [1]). In 1947, system of express and local schedules provided by the Chicago Transit Authority (CTA) had become a nuisance for users, because the really available services for riders were hard to comprehend. In order to stop the fall in demand that was being happening, the CTA planned a clever way of running express service on its two-track lines. This proposal was known as the A/B skip–stop plan (see [2]). As illustrative examples, the skip–stop operation mode has been used (and/or is being used) in practice in Santiago de Chile since 2001 [3], in Bogota's Transmilenio system [4], in Los Angeles's Metro Rapid system [5], in Singapur's Transit Link [6] and in the bus line that connects the East zone of Seville area with its historic centre [7]. Figure 1 shows four train services that run along a railway corridor with seven stations. The services appear classified in types A and B and the stations in types A, B and A/B. The horizontal sections in the polygonal lines represent the stopping time of the trains at the stations.

© 2019 WIT Press, www.witpress.com
ISSN: 2058-8305 (paper format), ISSN: 2058-8313 (online), http://www.witpress.com/journals
DOI: 10.2495/TDI-V3-N4-306-316

European Journal of Operational Research 277 (2019) 153–165

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

# Reformulated acyclic partitioning for rail-rail containers transshipment

Maria Albareda-Sambola[a,*], Alfredo Marín[b], Antonio M. Rodríguez-Chía[c]

[a] *Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya. Barcelona Tech, Spain*
[b] *Departamento de Estadística e Investigación Operativa, Universidad de Murcia, Spain*
[c] *Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain*

ARTICLE INFO

*Article history:*
Received 20 February 2018
Accepted 8 February 2019
Available online 14 February 2019

*Keywords:*
Combinatorial optimization
Rail-rail transshipment
Integer programming
Graph partitioning

ABSTRACT

Many rail terminals have loading areas that are properly equipped to move containers between trains. With the growing throughput of these terminals all the trains involved in a sequence of such movements may not fit in the loading area simultaneously, and storage areas are needed to place containers waiting for their destination train, although this storage increases the cost of the transshipment. This increases the complexity of the planning decisions concerning these activities, since now trains need to be packed in groups that fit in the loading area, in such a way that the number of containers moved to the storage area is minimized. Additionally, each train is only allowed to enter the loading area once.

Similarly to previous authors, we model this situation as an *acyclic graph partitioning problem* for which we present a new formulation, and several valid inequalities based on its theoretical properties. Our computational experiments show that the new formulation outperforms the previously existing ones, providing results that improve even on the best exact algorithm designed so far for this problem.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Inland freight transportation is an increasing activity with a high economical and environmental impact. Among the alternative transport modes, road transport has traditionally had a very large share (for instance, according to Eurostat Statistics Explained, 2017 in 2014 it accounted for 74.9% of the total inland freight transport -t–Km- within the EU). In order to increase its competitiveness and to avoid the drawbacks of this activity such as congestion and pollution, other transportation modes are encouraged by the authorities (see, e.g., European Commission, 2011). In this regard, rail transport (accounting already in 2014 for almost 20% of the total inland freight transport in the European Union) is a relevant alternative, and presents now an increasing trend. This desirable increase in the rail freight transportation requires an adaptation of the infrastructures and involves increasingly complex decisions from the managerial point of view. In particular, the new challenges within the freight handling processes in railway yards have originated a stream of literature on suitable optimization approaches and decision support systems. A recent survey on these works can be found in Boysen, Fliedner, Jaehn, and Pesch (2013).

Among the surveyed problems, Boysen, Jaehn, and Pesch (2011) consider the rail-rail transshipment system described next. In a given railway yard, a series of containers must be transshipped among a number of trains. There is a *loading area* composed by a series of parallel train tracks equipped with gantry cranes, and a storage area. Each train has a given length and the tracks have a limited capacity, so that not all trains can be within the loading area at the same time, and a schedule must be designed where one group of trains enters the loading area after the other. Each time, containers are moved between trains within the loading area. If a container has to be moved from one of those trains to another one which is not there, it is placed in the storage area to wait until the train that must carry it enters the loading area. Feasible schedules are those associated with groups of trains that fit in the loading area, where no train needs to enter this area more than once. The goal is to find a feasible schedule that minimizes the number of containers that need to be moved to and from the storage area (see Fig. 1).

Already in Boysen et al. (2011) and later in Nossack and Pesch (2014a), this situation was addressed through a graph partitioning problem (GPP). In these problems, the nodes of a given graph must be assigned to clusters in order to optimize an objective function. Different GPP variants arise when alternative conditions are imposed on the number or the size of these clusters, or on the structure of the graph they induce. Some of them are surveyed in Alpert and Kahng (1995) and in the recent book in Parrochia and Neuville (2013). GPPs are rather difficult to solve through integer programming mainly due to the symmetries inherent to the

* Corresponding author.
*E-mail address:* maria.albareda@upc.edu (M. Albareda-Sambola).

https://doi.org/10.1016/j.ejor.2019.02.022
0377-2217/© 2019 Elsevier B.V. All rights reserved.

Omega 92 (2020) 102165

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

# An optimization model for line planning and timetabling in automated urban metro subway networks. A case study☆☆☆

Víctor Blanco[a,*], Eduardo Conde[b], Yolanda Hinojosa[c], Justo Puerto[b]

[a] *IEMath-GR, Universidad de Granada, Spain*
[b] *Department of Statistics & OR, Universidad de Sevilla, Spain*
[c] *Department of Applied Economics I, Universidad de Sevilla, Spain*

ARTICLE INFO

*Article history:*
Received 1 July 2019
Accepted 19 November 2019
Available online 24 November 2019

*Keywords:*
Line planning
Short-turns
Timetabling
Mixed integer linear programming
Matheuristic

ABSTRACT

In this paper we present a Mixed Integer Linear Programming model that we developed as part of a pilot study requested by the R&D company Metrolab® in order to design tools for finding solutions for line planning and timetable situations in automated urban metro subway networks. Our model incorporates important factors in public transportation systems from both, a cost-oriented and a passenger-oriented perspective, as time-dependent demands, interchange stations, short-turns and technical features of the trains in use. The incoming flows of passengers are modeled by means of piecewise linear demand functions which are parameterized in terms of arrival rates and bulk arrivals. Decisions about frequencies, train capacities, short-turning and timetables for a given planning horizon are jointly integrated to be optimized in our model. Finally, a novel matheuristic approach is proposed to solve the problem. The results of extensive computational experiments are reported to show its applicability and effectiveness to handle real-world subway networks.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

In this paper, we propose a model for line planning and timetabling on general urban subway transportation systems. This study was originated by a real-world problem proposed by Metrolab®, a French R&D company, dealing with the line planning and timetabling of trains of existing subway networks. It was a pilot experience to automatize the decision making process, at the tactical and operational level, of a small section of the Paris subway network.

The development of flexible tools to control, automatically, a transportation system, in order to improve its service quality, may have considerable impact in its efficiency and usefulness. The perception, from the customer's point of view, of the quality of a public transportation system is highly dependent of its reliability, comfortability and effectiveness, when comparing with alternative transportation means. If just poor-quality connections are offered or the quality-price relationship does not fulfil the passengers' expectations, they may decide to use a different transportation mode. Therefore, one of the key objectives when designing and managing a public transportation system, besides infrastructure constraints, operational limitations or budget considerations, must be its service quality [16].

The existing literature on line planning is very extensive (see e.g., [14,16,34] and the references therein), including different models which can be classified according to the decisions covered (determination of train routes, frequency setting, or both), infrastructure and operational aspects, objective functions and the way in which passengers' decisions are taken into account in the decision making process. For instance, in [16], line planning models are classified with respect to their objective functions into models with cost-oriented or with passenger-oriented objective functions. Our model will consider both points of view in an attempt to find an equilibrium between these conflicting objectives which, as mentioned in [18], is an important challenge in a public transportation system.

However, building an effective model is much more complex than selecting the appropriate nature of the objective function. The correct delimitations of the considered features in the context of the transportation system or in the set of customers served by this system is a very important phase in the actual modeling. Often, an initial line planning must be re-engineered motivated

☆ This manuscript was processed by Associate Editor Yagiura.
☆☆ Société Metrolab®, Service Contrôle de Gestion, registered on the Paris Trade and Companies Register under the number 532 684 685 RCS Paris, with its registered office at 117/119 Quai de Valmy - 75010 PARIS - FRANCE.

* Corresponding author.
*E-mail addresses:* vblanco@ugr.es (V. Blanco), educon@us.es (E. Conde), yhinojos@us.es (Y. Hinojosa), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.omega.2019.102165
0305-0483/© 2019 Elsevier Ltd. All rights reserved.

TOP (2020) 28:442–474
https://doi.org/10.1007/s11750-019-00537-x

ORIGINAL PAPER

# Transportation infrastructure network design in the presence of modal competition: computational complexity classification and a genetic algorithm

**Federico Perea**[1] · **Mozart B. C. Menezes**[2] · **Juan A. Mesa**[3] · **Fernando Rubio-Del-Rey**[1]

Received: 10 May 2019 / Accepted: 27 December 2019 / Published online: 13 January 2020
© Sociedad de Estadística e Investigación Operativa 2020

## Abstract

In this paper we analyze the computational complexity of transportation infrastructure network design problems, in the presence of a competing transportation mode. Some of these problems have previously been introduced in the literature. All problems studied have a common objective: the maximization of the number of travelers using the new network to be built. The differences between them are due to two factors. The first one is the constraints that the new network should satisfy: (1) budget constraint, (2) no-cycle constraint, (3) both constraints. The second factor is the topology of the network formed by the feasible links and stations: (1) a general network, (2) a forest. By combining these two factors, in total we analyze six problems, five of them are shown to be NP-hard, the sixth being trivial. Due to the NP-hardness of these problems, a genetic algorithm is proposed. Computational experiments show the applicability of this algorithm.

**Keywords** Networks/graphs · Transportation · Computational complexity · Genetic algorithms

✉ Federico Perea
perea@eio.upv.es; fedperea@gmail.com

Mozart B. C. Menezes
mozart.menezes@neoma-bs.fr

Juan A. Mesa
jmesa@us.es

Fernando Rubio-Del-Rey
fernando.rubio.rey@outlook.es

[1] Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Valencia, Spain

[2] Department of Information Systems, Supply Chain Management and Decision Sciences, NEOMA Business School, 1 Rue du Maréchal Juin, 76130 Mont-Saint-Aignan, France

[3] Departamento de Matemática Aplicada II, Universidad de Sevilla, Sevilla, Spain



JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2020.1854628

ORIGINAL ARTICLE

# Assessing the effectiveness of park-and-ride facilities on multimodal networks in smart cities

Juan A. Mesa[a], Francisco A. Ortega[b], Miguel A. Pozo[c] and Ramón Piedra-de-la-Cuadra[b]

[a]Departamento de Matemática Aplicada II, Universidad de Sevilla, Sevilla, Spain; [b]Departamento de Matemática Aplicada I, Universidad de Sevilla, Sevilla, Spain; [c]Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Sevilla, Spain

**ABSTRACT**

This paper presents an optimization procedure to choose a parking facility according to different criteria: total travel time including transfers, parking fee and a factor depending on the risk of not having an available spot in the parking facility at the arrival time. An integer programming formulation is proposed to determine an optimal strategy of minimum cost considering the available information, different scenarios, and each user profile. To evaluate the performance, a computational experience has been carried out on Seville (Spain), where a historical city center restricts the traffic of private vehicles and encourages the use of parking facilities.

**ARTICLE HISTORY**
Received 14 November 2019
Accepted 16 November 2020

**KEYWORDS**
Park-and-ride facilities; routing on multimodal networks; shortest paths with fixed nodes; networks with time-dependent arcs

## 1. Introduction

One of the key transport issues facing many countries in the world is the increase in congestion in urban areas and their accesses, due to the heavy dependence on the use of private cars. Since park-and-ride facilities allow citizens to use their cars for a part of the trip, while completing the rest by means of public transportation, such facilities have been identified as a tool that can help to reduce traffic congestion (Mesa & Ortega, 2001). During recent decades increased attention has been paid to park-and-ride facilities. Reports commissioned by urban or transportation agencies of cities, metropolitan areas or governments, and papers published in specialized journals have informed about inquires and research related to several aspects of park-and-ride facilities. In the report Steer Davis Glease (2017) typical objectives of park-and-ride facilities are listed.

Searching-for-parking traffic comprises a significant amount of the total traffic volume, with average values reported equivalent to 30–50% of the total traffic volume at certain hours and districts (Arnott & Williams, 2017; Chaniotakis & Pel, 2015; Shoup, 2006). Parking choice can be considered as a search process where drivers make a number of linked decisions based on updated knowledge gained by experience (Thompson & Richardson, 1998). According to studies carried out in several big cities, the daily parking-search time is eight minutes on average (Shoup, 2006). This fact was validated from surveys conducted in Europe, which show that 25–40% of the travel time to central urban areas is attributed to the search for parking (Bonsall & Palmer, 2004). Holguín-Veras et al. (2012) compare the generalized cost of the only drive trip with that of using park-and-ride. The generalized cost for the only drive mode is a combination of the in-vehicle cost, the value of time, and out-of-pocket expenses, whereas that of the combined mode consists of in-vehicle and transit costs, the values of time, and out-of-pocket expenses both for the auto and transit branches. This paper aims at locating park-and-ride facilities, both in the linear and two-dimensional city cases, by maximizing their markets. Hence, reducing the parking-search time could lead to significant improvements in terms of travel time, production, traffic flows, fuel consumption, pollution, and noise emission. This is one of the reasons why searching for parking process has been investigated in recent years (see, e.g., Benenson et al., 2008; Kaplan & Bekhor, 2012).

The study of drivers' behavior regarding uncertainties about search times for finding vacant spot has been carried out in Chaniotakis and Pel (2015), where stated preference experiments are applied to several discrete choice models. One of the findings of this research is that most of the drivers searching for a parking lot who make a trip for shopping purposes start the searching process when they

CONTACT Francisco A. Ortega riejos@us.es Departamento de Matematica Aplicada I, Escuela Tecnica Superior de Arquitectura, Avda. Reina Mercedes 2, 41012-Sevilla, Spain
© Operational Research Society 2021

# Programación matemática: 7 publ.

Math. Program., Ser. A (2017) 165:579–604
DOI 10.1007/s10107-016-1096-1

FULL LENGTH PAPER

## Revisiting $k$-sum optimization

**J. Puerto**[1] · **A. M. Rodríguez-Chía**[2] · **A. Tamir**[3]

Received: 3 February 2016 / Accepted: 22 November 2016 / Published online: 16 December 2016
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2016

**Abstract** In this paper, we address continuous, integer and combinatorial $k$-sum optimization problems. We analyze different formulations of this problem that allow to solve it through the minimization of a relatively small number of minisum optimization problems. This approach provides a general tool for solving a variety of $k$-sum optimization problems and at the same time, improves the complexity bounds of many ad-hoc algorithms previously reported in the literature for particular versions of this problem. Moreover, the results developed for $k$-sum optimization have been extended to the more general case of the convex ordered median problem, improving upon existing solution approaches.

**Mathematics Subject Classification** 90B80 · 90B85 · 90C10 · 90C27 · 90C35

### 1 Introduction

The concept of $k$-sum optimization is a natural extension of the standard minisum optimization that looks for a feasible solution of a problem with the property that it minimizes the sum of the $k$ largest components of an $n$-dimensional cost vector. The term $k$-sum optimization was coined by Gupta and Punnen in their paper from 1990. Throughout the years, this type of an objective function has attracted the attention of researchers, and one can find a number of results that consider $k$-sum optimization

✉ J. Puerto
puerto@us.es

1 Instituto de Investigacion Matematica de la Universidad de Sevilla (IMUS), Universidad de Sevilla, Seville, Spain

2 Facultad de Ciencias, Universidad de Cádiz, Cádiz, Spain

3 School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

Optim Lett (2018) 12:203–220
https://doi.org/10.1007/s11590-017-1156-3

ORIGINAL PAPER

## The geometry of optimal partitions in location problems

**Lina Mallozzi**[1] · **Justo Puerto**[2]

Received: 3 November 2015 / Accepted: 26 May 2017 / Published online: 1 June 2017
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Given the position of some facilities, we study the shape of optimal partitions of the customers' area in a general planar demand region minimizing total average cost that depends on a set up cost plus some function of the travelling distances. By taking into account different norms, according to the considered situation of the location problem, we characterize optimal consumers' partitions and describe their geometry. The case of dimensional facilities is also investigated.

**Keywords** Facility location · Optimal transport · Polyhedral- and $\ell_p$-norms

### 1 Introduction

Most research in location analysis focuses on locating facilities to better cover customers' demand. This traditional analysis can be seen as a primal approach. Alternatively, from a dual point of view, one can approach the location problem from a different perspective, namely the allocation of customers to given facilities. This is also an important aspect in the area of location and it is usually done by means of some allocation rules such as most preferred, cheapest or closest. Obviously, the choice of an allocation rule has an important impact in the final solution since the same set of facilities may give rise to different solutions.

The problem that we wish to consider in this paper falls within the latter area. We wish to design "optimal" districting of regions of demand originated from given set of

✉ Lina Mallozzi
lina.mallozzi@unina.it

Justo Puerto
puerto@us.es

1 University of Naples Federico II, Naples, Italy

2 Universidad de Sevilla, Sevilla, Spain



European Journal of Operational Research 277 (2019) 215–226

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

## New algorithmic framework for conditional value at risk: Application to stochastic fixed-charge transportation

Elena Fernández[a], Yolanda Hinojosa[b,*], Justo Puerto[c], Francisco Saldanha-da-Gama[d]

[a] Departamento de Estadística e Investigación Operativa, Universidad de Cádiz, Spain
[b] IMUS and Departamento de Economía Aplicada I, Universidad de Sevilla, Spain
[c] IMUS and Departamento Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla, Spain
[d] Departamento de Estatística e Investigação Operacional and Centro de Matemática, Aplicações Fundamentais e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Portugal

ARTICLE INFO

*Article history:*
Received 25 April 2018
Accepted 4 February 2019
Available online 7 February 2019

*Keywords:*
Transportation
Stochastic mixed-integer programming
CVaR

ABSTRACT

This paper introduces a new algorithmic scheme for two-stage stochastic mixed-integer programming assuming a risk averse decision maker. The focus is the minimization of the conditional value at risk for a hard combinatorial optimization problem. Some properties of a mixed-integer non-linear programming formulation for conditional value at risk are studied as well as their algorithmic implications. This yields to a procedure for obtaining lower and upper bounds on the optimal value of the problem that may lead to an optimal solution. The new developments are applied to a fixed-charge transportation problem with stochastic demand, and they are computationally tested. The corresponding results are thoroughly presented and discussed.

© 2019 Elsevier B.V. All rights reserved.

### 1. Introduction

In this work we propose an algorithmic scheme for two-stage stochastic mixed-integer programming assuming a risk averse decision maker.

As recently pointed out by Escudero, Garín, and Unzueta (2017), many risk measures can be found in the literature, among which the most popular ones are scenario immunization, semi-deviation, min-risk, stochastic dominance, value-at-risk, and the conditional value-at-risk (CVaR) (see e.g., Chelst & Canbolat, 2011, for further details on these measures). The two latter measures are possibly the most widely used in the literature. The risk measure adopted in this work is the CVaR (Uryasev & Rockafellar, 2001).

Stochastic Programming has become a very important tool to hedge against uncertainty when compared to approaches derived from simplified (deterministic) counterpart models (Birge & Louveaux 2011, Wallace & Ziemba 2005). Due to applications arising in many relevant fields (e.g., logistics), an increasing amount of work is being done on two-stage, and multi-stage, stochastic mixed-integer linear programming (e.g., Hinojosa, Puerto, & Saldanha-da-Gama, 2014; Nickel, Saldanha-da-Gama, & Ziegler, 2012; Santoso, Ahmed, Goetschalckx, & Shapiro, 2005).

Initially, stochastic programming literature focused predominantly on risk-neutral decision makers, i.e., the future assets are accounted for using a mean-type objective function. More recently this trend has changed and there exists an increasing interest on models capturing a risk-averse attitude. To the best of our knowledge, the first work considering a mean-risk objective function in two-stage stochastic programming is Ahmed (2006), where the complexity of mean-risk stochastic programming considering different risk measures was investigated. Fábián (2008) studied two-stage CVaR minimization problems (as well as CVaR-constrained problems) and devised decomposition and solution schemes based on the linear programming representation of CVaR proposed by Künzi-Bay and Mayer (2006). Also in the context of two-stage stochastic linear programming problems, Miller and Ruszczyński (2011) consider an objective function combining several conditional risk measures.

For the case of two-stage stochastic mixed-integer linear programming problems, Schultz and Tiedemann (2006) presented an explicit mixed-integer linear programming formulation for the CVaR, when the probability distribution is discrete and finite. For the same problem Carœ and Schultz (1999) devised a dual decomposition approach by relaxing the non-anticipativity constraints.

The use of risk measures in two-stage stochastic mixed-integer linear programming problems has been considered in several applications. Noyan (2012) proposed a CVaR mean-risk model in the context of disaster relief, which was solved with a

* Corresponding author.
*E-mail addresses:* elena.fernandez@uca.es (E. Fernández), yhinojos@us.es (Y. Hinojosa), puerto@us.es (J. Puerto), fsgama@ciencias.ulisboa.pt (F. Saldanha-da-Gama).

https://doi.org/10.1016/j.ejor.2019.02.010
0377-2217/© 2019 Elsevier B.V. All rights reserved.

Journal of Optimization Theory and Applications (2019) 182:730–767
https://doi.org/10.1007/s10957-018-01470-y

# On Location-Allocation Problems for Dimensional Facilities

**Lina Mallozzi**[1] · **Justo Puerto**[2] · **Moisés Rodríguez-Madrena**[2]

Received: 31 May 2018 / Accepted: 31 December 2018 / Published online: 16 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

**Abstract**
This paper deals with a bilevel approach of the location-allocation problem with dimensional facilities. We present a general model that allows us to consider very general shapes of domains for the dimensional facilities, and we prove the existence of optimal solutions under mild assumptions. To achieve these results, we borrow tools from optimal transport mass theory that allow us to give explicit solution structure of the considered lower level problem. We also provide a discretization approach that can approximate, up to any degree of accuracy, the optimal solution of the original problem. This discrete approximation can be optimally solved via a mixed-integer linear program. To address very large instance sizes, we also provide a GRASP heuristic that performs rather well according to our experimental results. The paper also reports some experiments run on test data.

**Keywords** Bilevel optimization · Dimensional facilities · Optimal transport mass · Mixed-integer programming · Heuristics

**Mathematics Subject Classification** 90B85 · 49M25 · 90B80 · 90C30

Communicated by Michel Théra.

✉ Justo Puerto
puerto@us.es

Lina Mallozzi
lina.mallozzi@unina.it

Moisés Rodríguez-Madrena
madrena@us.es

1 University of Naples Federico II, Naples, Italy

2 IMUS, Universidad de Sevilla, Seville, Spain



European Journal of Operational Research 287 (2020) 262–279

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

# On pricing-based equilibrium for network expansion planning. A multi-period bilevel approach under uncertainty☆

Laureano F. Escudero[a,*], Juan F. Monge[b], Antonio M. Rodríguez-Chía[c]

[a] Area de Estadística e Investigación Operativa, Universidad Rey Juan Carlos, Móstoles (Madrid), Spain
[b] Centro de Investigación Operativa, Universidad Miguel Hernández de Elche, Elche, Alicante, Spain
[c] Dpto. de Estadística e Investigación Operativa, Universidad de Cádiz, Cádiz, Spain

ARTICLE INFO

*Article history:*
Received 18 March 2019
Accepted 15 March 2020
Available online 22 March 2020

*Keywords:*
Stochastic optimization
Network expansion planning
Pricing-based equilibrium
Mixed binary bilevel optimization
Stochastic Nested Decomposition

ABSTRACT

This study focuses on the development of a mixed binary primal-dual bilinear model for multi-period bilevel network expansion planning under uncertainty, where pricing-based equilibrated strategic and operational decisions are to be made. The periodwise dependent parameters' uncertainty is represented by a finite set of scenarios. Pricing-based equilibrium is required in the models to be optimized at the nodes of a multi-period scenario tree. Given the size of the models, it is unrealistic to seek an optimal solution. Several versions of a Stochastic Nested Decomposition matheuristic algorithm are presented for problem solving. Additionally, an approach based on a stagewise-related Stochastic Lagrangean Decomposition is also considered together with a Frank-Wolfe Progressive Hedging-based algorithm. The state step variables device is key for the performance of both approaches. The solution's optimality gap is computed for three out of the four solution providers that are presented. An extension of the Toll Assignment Problem is considered as a pilot case. A broad computational experience is reported.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

It is well known that a mixed binary mathematical optimization (say, maximization) problem under uncertainty has a mixed binary Deterministic Equivalent Model (DEM). In that case, the uncertainty in the parameters is usually represented by a finite set of representative scenarios. Traditionally, special attention has been given to optimizing the DEM by maximizing the expected value of the objective function in the scenarios, subject to the satisfaction of the problem constraints in the defined scenarios. We will hereafter refer to it as Risk Neutral (RN) problem.

This study focuses on the development of a stochastic mixed binary bilinear modeling framework for the multi-period network expansion planning (NEP) problem under uncertainty, where pricing-based equilibrated strategic and operational decisions are to be made. Let us name it PE-NEP. The topic is relevant, particularly in the broader context of managing network interactions in which a manager (or social planner) wishes to optimize an objective that is subject to incentive compatibility of the lower level agents. These conditions arise in the design of network platforms such as those for ride hailing, worker-employer matching and hospitality, among others.

The deterministic static version of the problem has been widely discussed in the literature, ever since the problem was introduced in Brotcorne, Labbé, Marcotte, and Savard (2000), Brotcorne, Labbé, Marcotte, and Savard (2001), Brotcorne, Labbe, and Savard (2008) and Labbé, Marcotte, and Savard (1988), among others, as a bilevel problem, and the Toll Assignment Problem (TAP) was considered as a pilot case. In those works the parameters are known, all of the links of the nodes in the network are available and the tariff of the upper level to charge to the lower level for commodity transportation through the links is also known. Thus, all the (operational) decisions are both deterministic and static. So, the problem entails deciding what commodities are to be transported through the network links from origin to destination. The aim is to obtain a price-based equilibrium between profit maximization of the upper level and the cost minimization of the lower level by transporting commodities through network links and alternative means. See recent reviews in Burtscheidt, Claus, and Dempe (2019), DeNegre (2011), Dempe, Kalashnikov, Pérez-Valdés, and

☆ This research has been partially supported by the projects RTI2018-094269-B-I00 (L.F. Escudero), MTM2016-79765-P (J.F. Monge) and MTM2016-74983-C2-2-R (A.M. Rodríguez-Chía) from the Spanish Ministry of Economy, Industry and Competitiveness, the European Regional Development Fund (AEI/FEDER, UE), and projects "Netmeetdata Fundation BBVA convocatoria 2019" and P18-FR-1422, FEDER-UCA18-106895 by Junta de Andaluía/UCA/FEDER.
* Corresponding author. Tel.: +34 627 407 403.
*E-mail addresses:* laureano.escudero@urjc.es (L.F. Escudero), monge@umh.es (J.F. Monge), antonio.rodriguezchia@uca.es (A.M. Rodríguez-Chía).

https://doi.org/10.1016/j.ejor.2020.03.048
0377-2217/© 2020 Elsevier B.V. All rights reserved.

Optimization Letters (2021) 15:97–107
https://doi.org/10.1007/s11590-020-01589-w

ORIGINAL PAPER

# An improved algorithm to compute the $\omega$-primality

**Wissem Achour**[1] · **Djamal Chaabane**[1] · **Víctor Blanco**[2]

Received: 25 September 2019 / Accepted: 25 April 2020 / Published online: 21 May 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

In this paper, we present an improved methodology to compute the $\omega$-primality of a numerical semigroup. The approach is based on exploiting the structure of the problem on a resolution method for optimizing a linear function over the set of efficient solutions of a multiple objective integer linear programming problem. The numerical experiments show the efficiency of the proposed technique compared to the existing methods.

**Keywords** Global optimization · Numerical semigroups · Optimization over an efficient set · Multi-objective integer programming · Omega primality

## 1 Introduction

The problem of optimizing a function over the set of efficient solutions of a multi-objective optimization problem has attracted the attention of researchers in the last decades due to its wide range of applications. Needless to say that most (if not all) of the real-world problems, in different fields (manufacturing, logistic, finance, …), have to deal with several conflicting goals. Several techniques have been designed to solve the different types of problems under this framework (see e.g., [1–3]). However, in general, the solution set of these problems is large, which is not reasonable from the decision maker viewpoint. Hence, optimizing a function over the set of solutions of a multiobjective function becomes a very useful tool to select the "best" solution among that set (see [4] for a recent real-world application of this problem). However, optimiz-

✉ Víctor Blanco
vblanco@ugr.es

Wissem Achour
wissem_achour16@yahoo.com

Djamal Chaabane
dchaabane@usthb.dz

1 Department of Operations Research, USTHB, Algiers, Algeria

2 IEMath-GR, Universidad de Granada, Granada, Spain



European Journal of Operational Research 282 (2020) 453–463

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# An exact algebraic $\epsilon$-constraint method for bi-objective linear integer programming based on test sets

María Isabel Hartillo-Hermoso[a,1], Haydee Jiménez-Tafur[b], José María Ucha-Enríquez[a,b,2,*]

[a] Departamento de Matemática Aplicada I, E. T. S. de Ingeniería Informática, Universidad de Sevilla, Av. Reina Mercedes s/n, Sevilla 41012, Spain
[b] Instituto de Matemáticas, Universidad de Sevilla, Avda. Reina Mercedes, s/n, Sevilla 41012, Spain

ARTICLE INFO

*Article history:*
Received 21 March 2018
Accepted 18 September 2019
Available online 28 September 2019

*Keywords:*
Multiple objective programming
Non-dominated set
Pareto set
$\epsilon$-constraint method
Unbounded Knapsack Problem
Algebraic test sets

ABSTRACT

A new exact algorithm for bi-objective linear integer problems is presented, based on the classic $\epsilon$-constraint method and algebraic test sets for single-objective linear integer problems. Our method provides the complete Pareto frontier $\mathcal{N}$ of non-dominated points and, for this purpose, it considers exactly $|\mathcal{N}|$ single-objective problems by using reduction with test sets instead of solving with an optimizer. Although we use Gröbner bases for the computation of test sets, which may provoke a bottleneck in principle, the computational results are shown to be promising, especially for unbounded knapsack problems, for which any usual branch-and-cut strategy could be much more expensive. Nevertheless, this algorithm can be considered as a potentially faster alternative to IP-based methods when test sets are available.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

In real-life problems, the pursued objectives are multiple and decision makers usually need complete knowledge of the best decisions to be made from those different points of view. The concept of *non-dominated solution or Pareto optimal* gathers the solutions that cannot be improved in any objective without degrading at least one of the other objectives.

Multi-objective Integer Programming (MOIP) is the branch that deals with this kind of problems in the case of integer variables. Generation methods compute the whole space of Pareto optimal solutions. Among these type of methods, the *weighted-sum method* transforms a multi-objective problem into several single-objective problems that arise from the linear combination of the different costs. If the Pareto set of points is convex this method provides the whole set of non-dominated solutions, but only the supported Pareto optimal points are obtained in the general case. On the other hand, the *$\epsilon$-constraint technique* generates a grid in the objective space with ranges between the costs of ideal and nadir points and, for each point in the upper-bound set (cf. Ehrgott and Gandibleux (2007) and Bringmann, Friedrich, Igel, and Voß (2013)), a single-objective IP is solved, thereby avoiding incremental movements through the grid.

In Laumanns, Thiele, and Zitzler (2006), an adaptive variation of the constraints that depends on the Pareto-optimal solutions found so far has been proposed. They use a lexicographic optimization with respect to the costs in the constraints in order to avoid weakly efficient Pareto optimal solutions. In their process slackness of the constraints are forced. The time complexity of their algorithm is $\mathcal{O}(|\mathcal{N}|^{p-1})$, where $|\mathcal{N}|$ is the number of Pareto-optimal solutions and $p$ the number of objectives.

In Ehrgott and Ruzika (2008), the main weaknesses of the $\epsilon$-constraint technique are pointed out: the lack of easy-to-check conditions for properly efficient solutions and the inflexibility of the constraints. In order to address these weaknesses, slack and surplus variables are added.

In Mavrotas (2009), the so-called *augmented $\epsilon$-constraint method* is presented. By using the previous idea of slack variables, an effective implementation of the $\epsilon$-constraint technique is given. In seeking non-dominated solutions, the algorithm changes the cost function, by rewarding whenever the slack variables are positive. This method is improved in Mavrotas and Florios (2013), and a variation is proposed in Zhang and Reimann (2014).

In Kirlik and Sayın (2014) a new method is described, which is also based on the $\epsilon$-constraint technique, with an innovative way of partitioning the search space in terms of rectangles. This technique presents a two-stage formulation for each problem in contrast to the lexicographic option.

* Corresponding author.
*E-mail address:* ucha@us.es (J.M. Ucha-Enríquez).
[1] Partially supported by MTM2016-74983-C2-1-R and MTM2016-75024-P.
[2] Partially supported by MTM2016-74983-C2-1-R, MTM2016-75024-P and P12-FQM-2696.

https://doi.org/10.1016/j.ejor.2019.09.032
0377-2217/© 2019 Elsevier B.V. All rights reserved.

# Aplicaciones: 9 publicaciones

Group Decision and Negotiation (2018) 27:1029–1046
https://doi.org/10.1007/s10726-018-9583-9

## A Multicriteria Selection System Based on Player Performance: Case Study—The Spanish ACB Basketball League

Víctor Blanco[1,2] · Román Salmerón[2] · Samuel Gómez-Haro[3]

Published online: 22 June 2018
© Springer Nature B.V. 2018

**Abstract**
In this paper, we describe an approach to rank sport players based on their efficiency. Although is extremely useful to analyze the performance of team games there is no unanimity on the use of a single index to perform such a ranking. We propose a method to summarize the huge amount of information collected at different aspects of a sport team which is almost daily publicly available. The tool will allow agents involved in a player's negotiation to show the strengths (and weaknesses) of the player with respect to other players. The approach is based on applying a multicriteria outranking methodology using as alternatives the potential players and as criteria different efficiency indices. A novel automatic parameter tuning approach is detailed that will allow coaches and sports managers to design templates and sports strategies that improve the efficiency of their teams. We report the results performed over the available information on the ACB Basketball League, and we show how it can be easily implemented and interpreted in practice by decision-makers non familiar with the mathematical side of the methodology.

**Keywords** Multicriteria decision aids · Ranking strategies · Players efficiency · Agents negotiation

✉ Víctor Blanco
vblanco@ugr.es

Román Salmerón
romansg@ugr.es

Samuel Gómez-Haro
samugh@ugr.es

[1] IEMath-GR, Universidad de Granada, Granada, Spain

[2] Department of Quantitative Methods for Economics & Business, Universidad de Granada, Granada, Spain

[3] Department of Business Management II, Universidad de Granada, Granada, Spain

Springer

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

## ON THE AGGREGATION OF EXPERTS' INFORMATION IN BONUS–MALUS SYSTEMS

BY

VÍCTOR BLANCO AND JOSÉ M. PÉREZ–SÁNCHEZ

ABSTRACT

In this paper, we propose a new family of premium calculation principles based on the use of prior information from different sources. Under this framework and based on the use of Ordered Weighted Averaging operators, we provide alternative collective and Bayes premiums and describe some approaches to efficiently compute them. Several examples are detailed to illustrate the performance of the new methods.

KEYWORDS

Non-life insurance, premium calculation principles, Ordered Weighted Averaging, mathematical programming, Bayesian statistics.

### 1. INTRODUCTION

A Bonus–Malus System (BMS) is a rating method widely used by actuaries in non-life insurance, specially for automobile insurance (see Lemaire, 1979, 1988, 1998; Shengwang *et al.*, 1999; Gómez *et al.*, 2002; Gómez–Déniz *et al.*, 2006), to compute the premium to be paid by an insurer. This system is based on the assumption that the premium charged to a customer or a portfolio of customers must depend on the number of claims reported in the past. In practice, the BMS methodology implies high penalties in case of reporting claims and moderate premium rewards to claim-free policyholders. Due to the nature of the problem, it is usual to apply Bayesian techniques to compute those BMS premiums. In the Bayesian paradigm, the knowledge about the unknown (risk) parameter is quantified by means of a prior distribution. In case the actual data are available, the prior knowledge is updated by using the posterior distribution.

One of the most usual techniques to compute BMS premiums consists of finding the amount that minimizes the expected deviation of the estimated risk parameter with respect to its actual value (also known as the loss). To this end, the risk is plugged into the distribution function that models the number

*Astin Bulletin* 48(1), 311-337. doi: 10.1017/asb.2017.25 © 2017 by Astin Bulletin. All rights reserved.

Downloaded from https://www.cambridge.org/core. Universidad de Sevilla, on 31 May 2021 at 14:26:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/asb.2017.25

Computers and Operations Research 98 (2018) 127–136

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## Mathematical programming formulations for the efficient solution of the *k*-sum approval voting problem

Diego Ponce[a], Justo Puerto[a], Federica Ricca[b], Andrea Scozzari[c,*]

[a] IMUS and Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Spain
[b] Università di Roma, La Sapienza, Italy
[c] Università degli Studi Niccolò Cusano, Roma, Italy

ARTICLE INFO

*Article history:*
Received 24 July 2017
Revised 27 March 2018
Accepted 14 May 2018
Available online 24 May 2018

*Keywords:*
Approval voting
Ordered Weighted Averaging (OWA)
*k*-sum optimization problems

ABSTRACT

In this paper we address the problem of electing a committee among a set of *m* candidates on the basis of the preferences of a set of *n* voters. We consider the approval voting method in which each voter can approve as many candidates as he likes by expressing a preference profile (boolean *m*-vector). In order to elect a committee, a voting rule must be established to 'transform' the *n* voters' profiles into a winning committee. The problem is widely studied in voting theory; for a variety of voting rules the problem was shown to be computationally difficult and approximation algorithms and heuristic techniques were proposed in the literature. In this paper we follow an Ordered Weighted Averaging approach and study the *k*-sum approval voting (optimization) problem in the general case $1 \leq k < n$. For this problem, we provide different mathematical programming formulations that allow us to solve it in an exact solution framework. We provide computational results showing that our approach is efficient for medium size test problems (*n* up to 200, *m* up to 60), since in all tested cases it was able to find the exact optimal solution in very short computational times.

© 2018 Elsevier Ltd. All rights reserved.

### 1. Introduction

A typical problem in collective decision making is to select one (or more) winners among a set of *m* candidates on the basis of the preferences of a set of *n* voters. This situation arises in many different real-life contexts, as in sport competitions to select the set of winners, or in political elections, and, more in general, whenever a committee must be formed from a larger set of candidates (for example, in companies or universities). For $m > 1$ the problem is a *multi-winner* one and the ballot gives the possibility to each voter to express his preference for each single candidate by *approving* or not his nomination. This means that voters can approve as many candidates as they like by expressing their preference profile (*approval balloting*). Approval voting is a well-known method used for this kind of multi-winner elections. The method, introduced by Brams and Fishburn in 1978 Brams and Fishburn (1978), was widely studied in the literature on voting theory (see Brams and Fishburn, 2005; Brams and Fishburn, 2007; Laslier and Sanver, 2010; Menezes et al., 2016; Rapoport and Felsenthal, 1990; Taylor and Pacelli, 2008 and the references therein).

Consider a set $N$ of $n$ voters and a set $A$ of $m$ candidates. For each $i$, $P_i = (p_{i1}, \ldots, p_{ij}, \ldots, p_{im})$ denotes the preference profile of voter $i$ which corresponds to a boolean $m$-vector whose generic element $p_{ij}$ is equal to 1 if candidate $j$ is approved by $i$ and equal to 0 otherwise. Therefore, in the problem input we have $N$, $A$ and a set $P$ of preference profiles of the voters, so that a generic instance is denoted by $(N, A, P)$. The problem is to find in $A$ the "best" subset of candidates (*winning* or *elected* committee), according to a certain voting rule (criterion). A committee is represented by a boolean $m$-vector $x = (x_1, \ldots, x_j, \ldots, x_m)$, in which $x_j = 1$ means that candidate $j$ is in the committee, and $x_j = 0$ otherwise.

When we have a single-winner election, the typical voting rule is the one that elects the most approved candidate, i.e., the one who received the largest number of votes (with a tie-breaking rule if needed).

For multi-winner elections, several voting rules have been proposed for approval voting (Kilgour, 2010). For a majority of the voting rules computing a winning committee is a difficult problem (Aziz et al., 2015; Fishburn and Pekec, 2004). Among the many, there is a class of rules known as *centralization procedures* that was widely studied in the literature. Two rules in this class were mainly analyzed, namely, one based on the *minisum* criterion and one on the *minimax* criterion. According to the first criterion, the winning committee is the one that minimizes the sum

* Corresponding author.
*E-mail addresses:* dponce@us.es (D. Ponce), puerto@us.es (J. Puerto), federica.ricca@uniroma1.it (F. Ricca), andrea.scozzari@unicusano.it (A. Scozzari).

https://doi.org/10.1016/j.cor.2018.05.014
0305-0548/© 2018 Elsevier Ltd. All rights reserved.

Omega 97 (2020) 102105

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

# A market regulation bilevel problem: A case study of the Mexican petrochemical industry ☆

Héctor Maravillo [a], José-Fernando Camacho-Vallejo [a,*], Justo Puerto [b], Martine Labbé [c]

[a] Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Nuevo León, Monterrey, Mexico
[b] Department of Statistics and Operational Research & Instituto de Matemáticas Universidad de Sevilla, Universidad de Sevilla, Seville, Spain
[c] Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium and INRIA, Lille, France

ARTICLE INFO

*Article history:*
Received 2 November 2018
Accepted 19 August 2019
Available online 24 August 2019

*Keywords:*
Market regulation
Bilevel programming
Petrochemical industry

ABSTRACT

In this paper, a bilevel programming model is proposed to study a problem of market regulation through government intervention. One of the main characteristics of the problem herein analyzed is that the government monopolizes the raw material in one industry, and competes in another industry with private firms for the production of commodities. Under this scheme, the government controls a state-owned firm to balance the market; that is, to minimize the difference between the produced and demanded commodities. On the other hand, a regulatory organization that coordinates private firms aims to maximize the total profit by deciding the amount of raw material bought from the a state-owned firm. Two equivalent single-level reformulations are proposed to solve the problem. The first reformulation is based on the strong duality condition of the lower level and results in a continuous non-linear model. The second reformulation resorts to the complementarity slackness optimality constraints yielding a mixed-integer linear model. Additionally, three heuristic algorithms are designed to obtain good-quality solutions with low computational effort. In this problem, the feasible region of the dual problem associated to the follower is independent from the leader's decision. Therefore, the proposed heuristics exploit this particular characteristic of the bilevel model. Moreover, the third heuristic hybridizes the other two algorithms to enhance its performance. Extensive computational experimentation is carried out to measure the efficiency of the proposed solution methodologies. A case study based on the Mexican petrochemical industry is presented. Additional instances generated from the case study are considered to validate the robustness of the proposed heuristic algorithms. Numerical results indicate that the hybrid algorithm outperforms the other two heuristics. However, all of them demonstrate to be good alternatives for solving the problem. Additionally, optimal solutions of all the instances are obtained by using good quality solutions (given by the hybrid algorithm) as initial solutions when solving the second reformulation via a general purpose solver.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Market regulation through government intervention has appeared in different ways, such as fiscal policies (taxes and subsidies), anti-monopoly legislation, price control, quantity control, or nationalization of firms, among others [1]. The Public Interest Theory, developed initially by Pigou, explains the regulation as "response to the demand of the public for the correction of inefficient or inequitable market practices" [2]. A specific manner of government intervention is through the participation of state-owned firms, which may be a public monopoly or may compete with private firms. The goal to regulate the market by participation has been very noticeable in developing countries, mainly in raw materials industries [3,4]. For instance, a table of patterns of state ownership firms on the world in 2010 is presented in [5], and a list of countries in which market regulation by participation exists in the oil industry in [6–8].

The first paper that considers the idea of regulating the market by including a state-owned firm that competes with other firms is [9]. In that paper, a short-term analysis of the entrance of a state-owned firm into a three-firms oligopoly market is done. They conclude that the existence of a state-owned firm may improve the performance of the market. Following up with the existence of state-owned firms, [10] considers a situation in which the difference between the production and the production level is made up by the government. The novelty in that paper is to realize that

☆ This manuscript was processed by Associate Editor O. Prokopyev.
* Corresponding author.
*E-mail address:* jose.camachovl@uanl.edu.mx (J.-F. Camacho-Vallejo).

https://doi.org/10.1016/j.omega.2019.08.012
0305-0483/© 2019 Elsevier Ltd. All rights reserved.

*mathematics*

MDPI

*Article*

# A Risk-Aversion Approach for the Multiobjective Stochastic Programming Problem

**Javier León [1,*], Justo Puerto [2] and Begoña Vitoriano [1]**

[1] HUM-LOG Research Group, Instituto de Matemática Interdisciplinar (IMI), Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Plaza de las Ciencias 3, 28040 Madrid, Spain; bvitoriano@mat.ucm.es
[2] Mathematical Research Institute (IMUS), University of Seville, 41004 Sevilla, Spain; puerto@us.es
* Correspondence: javileon@ucm.es

Received: 10 October 2020; Accepted: 8 November 2020; Published: 13 November 2020

check for updates

**Abstract:** Multiobjective stochastic programming is a field that is well suited to tackling problems that arise in many fields: energy, financial, emergencies, among others; given that uncertainty and multiple objectives are usually present in such problems. A new concept of solution is proposed in this work, which is especially designed for risk-averse solutions. The proposed concept combines the notions of conditional value-at-risk and ordered weighted averaging operator to find solutions protected against risks due to uncertainty and under-achievement of criteria. A small example is presented in order to illustrate the concept in small discrete feasible spaces. A linear programming model is also introduced to obtain the solution in continuous spaces. Finally, computational experiments are performed by applying the obtained linear programming model to the multiobjective stochastic knapsack problem, gaining insight into the behaviour of the new solution concept.

**Keywords:** multiobjective stochastic programming; linear programming; risk-aversion

## 1. Introduction

Decision making is never easy, yet we often have to make decisions. Emergencies and disaster management are fields in which many difficulties often arise, such as high uncertainty and multiple conflicting objectives. Risk-averse decisions are usually sought to overcome such difficulties. Risk-aversion is the attitude for which we prefer to lower uncertainty rather than gambling extreme outcomes (positive or negative).

Risk-aversion, although typically studied in problems with uncertainty, can as well be considered when making decisions with multiple criteria. For instance, in the field of disaster management, solutions that are sufficiently good for all criteria are usually preferred to others that perform exceptionally good for some criteria, but inadequately for the others.

Multicriteria decision making (MCDM) is a field worth of consideration when studying real-world problems. This situation, in which multiple conflicting objectives have to be optimized, has led to the definition of different solution concepts and methodologies. A specific methodology should be applied depending on the problem and the type of solution considered. A key concept in MCDM is the notion of efficiency, which reflects the intuition that, for a solution to be acceptable, another cannot exist improving that one in every objective.

Uncertainty is another feature that is present in the studied problems, in which risk-averse decisions will be preferred. The most common ways for dealing with the uncertainty are stochastic programming and robust optimization, in which fuzzy optimization is also included [1]. Stochastic programming is the widest used technique when there are historical data or information to infer a probability distribution. Moreover, usually discrete distributions are used, calling scenarios the



# A METHOD FOR CONSTRUCTING AND INTERPRETING SOME WEIGHTED PREMIUM PRINCIPLES

BY

ANTONIA CASTAÑO-MARTÍNEZ, FERNANDO LÓPEZ-BLAZQUEZ, GEMA PIGUEIRAS AND MIGUEL Á. SORDO

ABSTRACT

We present a method for constructing and interpreting weighted premium principles. The method is based on modifying the underlying risk distribution in such a way that the risk-adjusted expected value (or premium) is greater than the expected value of some conveniently chosen function of claims, which defines the insurer's perception of the risk. Under some assumptions on the function of claims, the method produces distortion premium principles. We provide several examples under different assumptions on the claim arrival process and different functions of claims, including record claims and $k$th record claims.

KEYWORDS

Weighted premium principle, distortion premium principle, equilibrium distribution, order statistics, record value.

**JEL code:** G220.

## 1. INTRODUCTION

The actuarial literature provides a wide variety of approaches for pricing insurance risks (see Goovaerts and Laeven, 2008, for a review). Some theories are based on the idea that the price of a risk should be derived from a risk-adjusted distribution, which is obtained by modifying the underlying risk distribution to include risk aversion or the insurer's perception of the risk. Relevant examples of risk-adjusted distributions include distorted distributions (see Yaari, 1987, Denneberg, 1990, and Wang, 1996) and weighted distributions (see Furman and Zitikis, 2008a,b, 2009 and Zhu *et al.*, 2016).

*Astin Bulletin* 50(3), 1037-1064. doi:10.1017/asb.2020.15 © 2020 by Astin Bulletin. All rights reserved.

Downloaded from https://www.cambridge.org/core. Universidad de Sevilla, on 31 May 2021 at 15:12:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/asb.2020.15

Computational Statistics
https://doi.org/10.1007/s00180-021-01103-5

ORIGINAL PAPER

# Automatic differentiation and maximal correlation of order statistics from discrete parents

Fernando López-Blázquez[1] · Begoña Salamanca-Miño[1]

Received: 22 October 2020 / Accepted: 27 March 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

The maximal correlation is an attractive measure of dependence between the components of a random vector, however it presents the difficulty that its calculation is not easy. Here, we consider the case of bivariate vectors which components are order statistics from discrete distributions supported on $N \geq 2$ points. Except for the case $N = 2$, the maximal correlation does not have a closed form, so we propose the use of a gradient based optimization method. The gradient vector of the objective function, the correlation coefficient of pairs of order statistics, can be extraordinarily complicated and for that reason an automatic differentiation algorithm is proposed.

**Keywords** Automatic differentiation · Maximal correlation · Order statistics · Discrete distributions

## 1 Introduction

The Pearson's correlation coefficient, $\rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$, is a widely used measure of linear dependence between the components of a bivariate random vector $(X, Y)$ provided that $0 < \mathrm{Var}(X) < \infty$ and $0 < \mathrm{Var}(Y) < \infty$. The *maximal correlation coefficient* is defined as

$$R(X, Y) := \sup \rho(g(X), h(Y)), \tag{1}$$

where the supremum in (1) is taken over the set of measurable functions $g, h : \mathbb{R} \to \mathbb{R}$ with $0 < \mathrm{Var}(g(X)) < \infty$ and $0 < \mathrm{Var}(h(Y)) < \infty$, see Gebelein (1941). The

✉ Fernando López-Blázquez
lopez@us.es

Begoña Salamanca-Miño
bsm@us.es

[1] Department of Statistics & OR, Universidad de Sevilla, Avda. Reina Mercedes s/n, 41012 Sevilla, Spain

Published online: 12 April 2021



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Insurance: Mathematics and Economics 86 (2019) 92–97

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

# On a family of risk measures based on largest claims

A. Castaño-Martínez, G. Pigueiras, M.A. Sordo *

*Department of Statistics and Operation Research, University of Cádiz, 11002 Cádiz, Spain*

ARTICLE INFO

*Article history:*
Received March 2018
Received in revised form February 2019
Accepted 4 February 2019
Available online 23 February 2019

*JEL classification:*
G220

*Keywords:*
Risk measure
Premium principle
Order statistics
Stop-loss order
Excess-wealth order
Reinsurance

ABSTRACT

Given a set of $n \geq 2$ independent and identically distributed claims, the expected average of the $n - i$ largest claims, with $0 \leq i \leq n - 1$, is shown to be a distortion risk measure with concave distortion function that can be represented in terms of mixtures of tail value-at-risks with beta mixing distributions. This result allows to interpret the tail value-at-risk in terms of the largest claims of a portfolio of independent claims. As an application, we provide sufficient conditions for stochastic comparisons of premiums in the context of large claims reinsurance.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Consider a random sample $X_1, X_2, \ldots, X_n$ of independent claims with the same distribution $F$ as $X$ and denote by $X_{1:n} \leq \cdots \leq X_{n:n}$ the corresponding order statistics. The total claim amount of the $n - i$ largest claims, namely

$$S_{i,n}(X) = \sum_{j=i+1}^{n} X_{j:n}, \quad 0 \leq i \leq n-1,$$

has received some attention in the actuarial literature, especially in the study of heavy-tailed distributions. For example, Embrechts and Klüppelberg (1993) and Aebi et al. (1994) study a measure of dangerousness of a claim distribution $F$ by using the proportion of $S_{i,n}(X)$ to the aggregate claim amount in the portfolio. In the context of collective risk models, several reinsurance treaties (such as LCR and ECOMOR) are based on the largest claims during a given period of time (see Section 2.5 and Chapter 10 in Albrecher et al., 2017, for an overview).

In this paper, given a random sample of $n$ independent claims with the same distribution $F$ as $X$, we quantify the right-tail risk of $X$ by using the expected average of the $n - i$ largest claims, given by

$$T_{i,n}(X) = \frac{1}{n-i} \sum_{j=i+1}^{n} E[X_{j:n}], \quad 0 \leq i \leq n-1, \quad n \geq 2. \tag{1}$$

The main result of this paper, stated in Section 2, shows that $T_{i,n}(X)$ is a concave distortion risk measure, which means (see Yaari, 1987, and Section 1.1 in Goovaerts et al., 2010, for an overview) that there exists an increasing concave function $h_{i,n} : [0, 1] \longrightarrow [0, 1]$ with $h_{i,n}(0) = 0$ and $h_{i,n}(1) = 1$ (called distortion function) such that

$$T_{i,n}(X) = \int_0^\infty h_{i,n}\left(\overline{F}(x)\right) dx,$$

where $\overline{F} = 1 - F$ is the tail function of $X$. Distortion risk measures have many convenient properties: they are monotone, translation invariant, positively homogeneous and, when the distortion function is concave, subadditive and, therefore, coherent (see Section 2.6 in Denuit et al., 2005). In order to prove the main result, we first rewrite $T_{i,n}(X)$ in terms of mixtures of tail value-at-risks with beta mixing distributions, thus obtaining an alternative representation for the premium principles defined in Sordo et al. (2016). In Section 3, as a consequence, we obtain a convergence result that interprets the tail value-at-risk of $X$, which is defined by

$$\mathrm{TVaR}_p(X) = \frac{1}{1-p} \int_p^1 F^{-1}(t)\, dt, \quad p \in (0, 1),$$

where $F^{-1}(t) = \inf\{x : F(x) \geq t\}$, $t \in (0, 1)$, in terms of the largest claims of a portfolio of independent claims with the same distribution.

We consider in this paper two stochastic orders based on the notion of stop-loss premium: the stop-loss order and the excess-wealth order. For a risk $X$ with deductible $t$, the stop-loss

* Corresponding author.
*E-mail addresses:* antonia.castano@uca.es (A. Castaño-Martínez), gema.pigueiras@uca.es (G. Pigueiras), mangel.sordo@uca.es (M.A. Sordo).

https://doi.org/10.1016/j.insmatheco.2019.02.003
0167-6687/© 2019 Elsevier B.V. All rights reserved.

# Teoría de juegos: 5 publicaciones

J. Math. Anal. Appl. 457 (2018) 1634–1648

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

## On minimax and Pareto optimal security payoffs in multicriteria games

Justo Puerto [a,*], Federico Perea [b]

[a] IMUS, Universidad de Sevilla, Spain
[b] Instituto Tecnológico de Informática, Universitat Politècnica de València, Spain

ARTICLE INFO

*Article history:*
Received 20 July 2016
Available online 5 January 2017
Submitted by A. Barbagallo

*Keywords:*
Multicriteria games
Payoffs
Characterization

ABSTRACT

In this paper, we characterize minimax and Pareto-optimal security payoff vectors for general multicriteria zero-sum matrix games, using properties similar to the ones that have been used in the single criterion case. Our results show that these two solution concepts are rather similar, since they can be characterized with nearly the same sets of properties. Their main difference is the form of consistency that each solution concept satisfies. We also prove that both solution concepts can transform into each other, in their corresponding domains.

© 2017 Elsevier Inc. All rights reserved.

### 1. Introduction

The standard two-person zero-sum game was introduced by von Neumann [14]. Ever since then, extensive research has been devoted to these games. Even nowadays the interest on these games has not decreased, see e.g. [13], who analyze how players' information affect their payoffs. We here elaborate on the Multicriteria Zero-Sum Matrix (MZSM) games, which allow us to handle simultaneous confrontation of two rational agents in several scenarios. The first game theoretical characterizations of minimax values in MZSM games is due to Shapley [20]. In its introductory note, F.D. Rigby remarks the importance of these games: *"The topic of games with vector payoffs is one which could be expected to attract attention on the basis of its intrinsic interest"*. Nevertheless, the development of the theory of multicriteria games has not been as successful as expected. There is a number of references in the literature as we will see along the paper, although its intrinsic difficulty, mainly due to the lack of total orderings among players' payoffs, has diminished the interest of researchers. In spite of that, the goal of further developing the analysis of multicriteria games should not be forgotten. In fact, each competitive situation that can be modeled as a scalar zero-sum game has its counterpart as a multicriteria zero-sum game, when more than one scenario has to be compared

* Corresponding author.
*E-mail addresses:* puerto@us.es (J. Puerto), perea@eio.upv.es (F. Perea).

http://dx.doi.org/10.1016/j.jmaa.2017.01.002
0022-247X/© 2017 Elsevier Inc. All rights reserved.

Computers and Operations Research 112 (2019) 104764

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## A heuristic procedure for computing the nucleolus

Federico Perea [a,*], Justo Puerto [b]

[a] Grupo de Sistemas de Optimización Aplicadas, Instituto Tecnológico de Informática, Ciudad Politécnica de la Innovación, Edificio 8G, Acc. B. Universitat Politècnica de València, Camino de Vera s/n, València, 46021, Spain
[b] Instituto de Matemáticas de la Universidad de Sevilla. Avda. Reina Mercedes, s/n, Sevilla, 41012, Spain

ARTICLE INFO

*Article history:*
Received 13 November 2018
Revised 27 June 2019
Accepted 3 August 2019
Available online 6 August 2019

*Keywords:*
Nucleolus
Game theory
Linear programming
Heuristic

ABSTRACT

This paper introduces a row and column generation algorithm for finding the nucleolus, based on a linear programming model proposed in an earlier research. Since this approach cannot return an allocation for large games, we also propose a heuristic approach, which is based on sampling the coalitions space. Experiments over medium sized games show that the proposed heuristic finds allocations which are close to the true nucleolus, in a reasonable amount of time. Experiments over 100-player games show that the proposed heuristic can be applied to games of large size.

© 2019 Elsevier Ltd. All rights reserved.

### 1. Introduction

A transferable utility (TU) game can be defined by means of:

- A set of *players* $N = \{1, \ldots, n\}$. Players are allowed to cooperate, but their objective is to maximize their own individual benefit.
- For each *coalition* $S \subset N$, the characteristic function $v(S)$ represents the profit that the cooperation of the players in $S$ yields, without the help of the other $N \backslash S$ players. The set of all players $N$ is referred to as the *grand coalition*.

TU-games can therefore be identified by means of their set of players and characteristic function: $(N, v)$.

One of the main challenges in TU-games consists of sharing the profit that the grand coalition can make, among the different players. There are two main ways to address that question:

1. Using solution sets.
2. Using allocation rules.

The most well-known representative of the first type is the *core* (Owen, 1995). The core of a TU-game is defined as

$$C(v) = \{x \in \mathbb{R}^n : x(S) \geq v(S) \ \forall \ S \subset N, \ x(N) = v(N)\}.$$

Core allocations ensure that each coalition $S$ gets a share of the profit obtained by the grand coalition which is, at least, as high as the profit that $S$ can make on their own. Thus, core allocations are well accepted due to the fairness conditions they satisfy.

However, there are games without core allocations. Even for some games which do have core allocations, it might be very difficult to find them or, choose one among them. Therefore, at times one is interested in the second type of solutions: Allocation rules. Allocation rules are procedures that allocate to each player a share of the benefit obtained by the grand coalition. Two important such rules are the Shapley value and the nucleolus, which are well accepted for the properties they satisfy, and are widely used in complex TU-games (e.g. see Yu et al., 2017, where the Shapley value is computed in a pickup and delivery cooperative game). The reader should note that the Shapley value belongs to the core when the game is convex, and the nucleolus belongs to the core whenever the core is non-empty.

In this paper, we focus on the nucleolus (Kohlberg, 1972), which we now introduce for the sake of completeness.

Given a TU-game $(N, v)$, the set of pre-imputations $\tilde{V}$ and imputations $V$ of the game are:

$$\tilde{V} = \left\{ x \in \mathbb{R}^n : \sum_{j=1}^{n} x_j = v(N) \right\},$$

$$V = \left\{ x \in \mathbb{R}^n : \sum_{j=1}^{n} x_j = v(N), x_j \geq v(\{j\}), \ \forall \ j \in N \right\}.$$

Note that $V \subset \tilde{V}$. Given a pre-imputation $x \in \tilde{V}$, the excess vector of $x$ is the vector $\theta(x) \in \mathbb{R}^{2^n-2}$

$$\theta(x) = (e(S, x)), \quad \text{with } e(S, x) = v(S) - \sum_{i \in S} x_i \ \forall \ S \subset N, \ S \neq \emptyset, N.$$

After introducing these two concepts, the nucleolus can be defined.

* Corresponding author.
*E-mail addresses:* perea@eio.upv.es (F. Perea), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.cor.2019.104764
0305-0548/© 2019 Elsevier Ltd. All rights reserved.

European Journal of Operational Research 279 (2019) 93–106

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

## A stochastic approach to approximate values in cooperative games

Stefano Benati [a,*], Fernando López-Blázquez [b], Justo Puerto [b]

[a] Dipartimento di Sociologia e Ricerca Sociale, Università di Trento, Italy
[b] Instituto Universitario de Investigación Matemática (IMUS), Universidad de Sevilla, Spain

ARTICLE INFO

*Article history:*
Received 26 February 2018
Accepted 22 May 2019
Available online 25 May 2019

*Keywords:*
Game theory
Cooperative games
Values
Stochastic games
Statistical sampling

ABSTRACT

Computing additive values in cooperative games, like the Shapley value, is a hard task because, in general, it involves the summation of an exponential number of terms. We propose a new method, based on the stochastic approximation of deterministic games and sampling theory, to calculate a statistic estimate of these values and, at the same time, keeping under control estimation errors. We applied this technique to several well-known games and we show that in many cases we were able to improve previous results.

© 2019 Elsevier B.V. All rights reserved.

### 1. Introduction

Let $(N, v)$ be a cooperative game with set of players $N = \{1, \ldots, n\}$ and characteristic function $v$ with $v(\emptyset) = 0$. For simplicity throughout this paper we will assume that the characteristic function $v$ takes values in the set of real numbers. The extension of our results to the case of vectorial characteristic functions is straightforward, see Puerto, Fernández, and Hinojosa (2008). Let $G^N$ denote the class of $N$ -person cooperative games. The main goal of this class of games is to allocate the grand coalition payoff $v(N)$ among the players in the game. There exist many allocation rules with a number of different properties and any of these rules may be considered as a solution concept of the game.

A well-accepted solution concept in cooperative game theory is the concept of value. A value $\alpha : G^N \to \mathbb{R}^n$ is a $n$-tuple of real numbers (one per agent in the grand coalition) where the $i$th component is the amount allocated to agent $i$ in the game $(N, v)$.

These values are determined by the properties required to them. Linearity is a desirable property that has been required to most of the values considered in the literature. Thus, we will be interested in studying values whose functional form fits into the following formula:

$$\alpha(\boldsymbol{v}_*) = \sum_{S \in \mathcal{U}} a(S) \circ v_*(S), \tag{1}$$

where $\circ$ is the Hadamard product of vectors, $a(S) = (a_1(S), \ldots, a_n(S))^t$, $S \in \mathcal{U}$ are real $n$-vectors depending only on $S$, $\mathcal{U} = 2^N$ the family of all the coalitions of the game and $\boldsymbol{v}_* = (v_*(S))_{S \subseteq U}^t$, where $v_*(S) := (v_{1,*}(S, v), \ldots, v_{n,*}(S, v))^t$ are $n$-vectors of real values depending on $S$, the characteristic function of the game and eventually on each player in the game. There are many examples of these values although the best known may be the Shapley value, the family of Semivalues (Dubey, Neyman, & Weber, 1981), the Least Squares values (Ruiz, Valenciano, & Zarzuelo, 1998) and the Multinomial values (Carreras & Puente, 2015; 2018).

- The Shapley value of the $i$th player, for the cooperative game $(N, v)$, is defined as

$$\phi_i(v) = \sum_{S \subseteq N \backslash \{i\}} \frac{s!\ (n-s-1)!}{n!} (v(S \cup \{i\}) - v(S)), \tag{2}$$

where $|S| = s$ denotes the cardinal of the set $S$. The Shapley value corresponds to the form given in (1) with $a_i(S)$ equal to $\frac{s!(n-s-1)!}{n!}$, if $i \notin S$ and zero otherwise; and $v_{i,*}(S) := v(S \cup \{i\}) - v(S)$.
- The Semivalues, introduced by Dubey et al. (1981), are similar to the Shapley values but with different systems of coefficients:

$$\psi_i(v) = \sum_{S \subseteq N \backslash \{i\}} p_s (v(S \cup \{i\}) - v(S)),$$

with $p_s \geq 0, \ s = 0, \ldots, n-1$, real numbers such that $\sum_{s=0}^{n-1} \binom{n-1}{s} p_s = 1$. The Semivalue $\psi$ corresponds to (1) with $a(S) = p_s$, for $S \subset N \backslash \{i\}$; $a(S) = 0$ otherwise, and $v_{i,*}(S) := v(S \cup \{i\}) - v(S)$. Clearly the Shapley value is also a Semivalue.
- The family of Least Squares values introduced by Dragan (2006) and Ruiz et al. (1998) also fits to this framework (1).

* Corresponding author.
*E-mail addresses:* stefano.benati@unitn.it (S. Benati), lopez@us.es (F. López-Blázquez), puerto@us.es (J. Puerto).

https://doi.org/10.1016/j.ejor.2019.05.027
0377-2217/© 2019 Elsevier B.V. All rights reserved.

Annals of Operations Research
https://doi.org/10.1007/s10479-021-04083-3

ORIGINAL RESEARCH

# Enforcing fair cooperation in production-inventory settings with heterogeneous agents

Luis A. Guardiola[1] · Ana Meca[2] · Justo Puerto[3]

Accepted: 16 April 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Production-Inventory settings with heterogeneous agents appear frequently in the study of supply chain management. For instance, there are Production-Inventory situations in which certain agents are essential as they can reduce the costs of other agents (followers) when they cooperate with each other. The study of such a cooperation can be modelled by means of a cooperative game and studied finding fair cost allocations. These class of cooperative games was introduced in Guardiola et al. (in Games Econ Behav 65:205–219, 2009) where it was also proposed the Owen point. This cost allocation is an appealing solution concept that for Production-Inventory games (PI-games) is always stable, in the sense of the core. The Owen point allows all the players in the game to operate at minimum cost but it does not take into account the cost reduction induced by essential players over their followers. Thus, it may be seen as an altruistic allocation for essential players what can be criticized. The aim of this paper is two-fold: to introduce new core allocations for PI-games improving the weaknesses of the Owen point and to study the structure and complexity of set of stable cost allocations (the core) of PI-games.

**Keywords** Cooperative game · Core · Cost allocation · Production-inventoy setting

**Mathematics Subject Classification** 91A12 · 90B05

The research authors is supported from Spain's Ministerio de Ciencia, Innovación y Universidades (MCIU), from the Agencia Estatal de Investigación (AEI) and from the Fondo Europeo de Desarrollo Regional (FEDER) under the projects MTM2016-74983-C02-01 and PGC2018-097965-BI00. The research of the third author is also partially supported from projects FEDER-US-1256951, CEI-3-FQM331 and NetmeetData: Ayudas Fundación BBVA a equipos de investigación científica 2019.

✉ Ana Meca
ana.meca@umh.es

Luis A. Guardiola
luis.guardiola@ua.es

Justo Puerto
puerto@us.es

1 Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03690 Alicante, Spain

2 I.U. Centro de Investigación Operativa, Universidad Miguel Hernández, 03202 Elche (Alicante), Spain

3 Facultad de Matemáticas, Universidad de Sevilla, 41012 Sevilla, Spain

Published online: 02 May 2021



*mathematics*

MDPI

*Article*

# Unitary Owen Points in Cooperative Lot-Sizing Models with Backlogging

**Luis A. Guardiola [1,*,†], Ana Meca [2,†] and Justo Puerto [3,†]**

1 Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, 03690 Alicante, Spain
2 I.U. Centro de Investigación Operativa, Universidad Miguel Hernández, Edificio Torretamarit, Avda. de la Universidad s.n., 03202 Elche , Spain; ana.meca@umh.es
3 Facultad de Matemáticas, Universidad de Sevilla, 41012 Sevilla, Spain; puerto@us.es
* Correspondence: luis.guardiola@ua.es
† All the authors contributed equally to this work.

**Abstract:** This paper analyzes cost sharing in uncapacitated lot-sizing models with backlogging and heterogeneous costs. It is assumed that several firms participate in a consortium aiming at satisfying their demand over the planning horizon with minimal operating cost. Each individual firm has its own ordering channel and holding technology, but cooperation with other firms consists in sharing that information. Therefore, the firms that cooperate can use the best ordering channels and holding technology among members of the consortium. This mode of cooperation is stable. in that allocations of the overall operating cost exist, so that no group of agents benefit from leaving the consortium. Our contribution in the current paper is to present a new family of cost sharing allocations with good properties for enforcing cooperation: the unitary Owen points. Necessary and sufficient conditions are provided for the unitary Owen points to belong to the core of the cooperative game. In addition, we provide empirical evidence, through simulation, showing that, in randomly-generated situations, the above condition is fulfilled in 99% of the cases. Additionally, a relationship between lot-sizing games and a certain family of production-inventory games, through Owen's points of the latter, is described. This interesting relationship enables easily constructing a variety of coalitionally stable allocations for cooperative lot-sizing models.

**Keywords:** unitary Owen points; cooperation; cost allocation; coalitional stability

check for updates

**Citation:** Guardiola, L.A.; Meca, A.; Puerto, J. Unitary Owen Points in Cooperative Lot-Sizing Models with Backlogging. *Mathematics* **2021**, *9*, 869. https://doi.org/10.3390/math9080869

Academic Editor: David Carfi

Received: 25 February 2021
Accepted: 11 April 2021
Published: 15 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The economic lot-sizing problem (henceforth ELSP) is a production problem in operations management and inventory theory that has been studied by many researchers for more than 50 years. The ELSP is an extension of the economic order quantity model to the case where there are some goods to be produced over a planning horizon, so that the production lots must be decided, in order to meet certain demand over the given finite horizon. Demand is usually generated from forecasts or by customer orders, or often by a combination of both. This production planning model is a common point for most companies or industries: planning what, when, and how much to produce.

To define the feasible production plans, several parameters of the production system are usually taken into account: the resource capacity (with or without restrictions) and all of the inventory costs involved. The simpler production planning model is known as the single-item uncapacitated lot-sizing model. It corresponds to the production planning of a single item to meet some demand over a discretized planning horizon. Despite its simplicity, it already contains most of the modeling elements that are cited above, with the exception of the capacity constraints.

There are also some other modeling elements that can be found in some situations. Those elements usually complicate the models and make them more difficult to solve. For instance, a competitive model for the allocation of capacity from some shared resources.

# Particionamiento: 2 publicaciones

Discrete Optimization 30 (2018) 96–107

Contents lists available at ScienceDirect

Discrete Optimization

www.elsevier.com/locate/disopt

## Uniform and most uniform partitions of trees☆

Isabella Lari[a], Justo Puerto[b], Federica Ricca[a], Andrea Scozzari[c,*]

[a] Sapienza Università di Roma, Italy
[b] IMUS, Spain
[c] Università degli Studi Niccolò Cusano - Roma, Italy

ARTICLE INFO

*Article history:*
Received 6 March 2017
Received in revised form 22 February 2018
Accepted 8 June 2018
Available online 3 July 2018

*Keywords:*
Graph partitioning
Most uniform partition problems
Centered partitions
Equipartition problems

ABSTRACT

This paper addresses centered and non centered equipartition tree problems into $p$ connected components ($p$-partitions). In the former case, each partition must contain exactly one special vertex called center, whereas in the latter, partitions are not required to fulfill this condition. Among the different equipartition problems considered in the literature, we focus on: (1) Most Uniform Partition (MUP) and (2) Uniform Partition (UP). Both criteria are defined either w.r.t. weights assigned to each vertex or to costs assigned to each vertex–center pair. Costs are assumed to be flat, i.e., they are independent of the topology of the tree. With respect to costs, MUP minimizes the difference between the maximum and minimum cost of the components of a partition and UP resorts to optimal min–max or max–min partitions. We provide polynomial time algorithms for centered and non centered versions of the MUP problem on trees when weights are assigned to each vertex. In the non centered case, our results set as polynomial the complexity of this problem which was an open question for several years. On the contrary, we prove that the centered version of MUP with flat costs is NP-complete even on trees. For the UP problem, we develop polynomial time algorithms for the max–min and min–max centered $p$-partition problems on trees both in the case of weight and cost-based objective functions.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

The essence of partitioning problems is to divide a set of units into groups or clusters so that units belonging to the same group are as similar as possible, or the clusters are as balanced as possible. In several applications such as community detection, image processing, wireless sensor networks and territorial districting [1–4], the set of units can be represented by the set of vertices of a graph $G$ where edges represent relations between units [5].

☆ The second, third and fourth authors wish to thank the Spanish Ministry of Economía and Competitividad that has partially supported this research by FEDER grant number MTM2016-74983-C02-01.

* Corresponding author.
*E-mail addresses:* isabella.lari@uniroma1.it (I. Lari), puerto@us.es (J. Puerto), federica.ricca@uniroma1.it (F. Ricca), andrea.scozzari@unicusano.it (A. Scozzari).

https://doi.org/10.1016/j.disopt.2018.06.003
1572-5286/© 2018 Elsevier B.V. All rights reserved.

Discrete Applied Mathematics 285 (2021) 78–92

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

## Locating a discrete subtree of minimum variance on trees: New strategies to tackle a very hard problem

Justo Puerto[a,b], Federica Ricca[c], Andrea Scozzari[d,*]

[a] Institute of Mathematics University of Seville (IMUS), Seville, Spain
[b] Department of Statistics and OR, Universidad de Sevilla, Seville, Spain
[c] MEMOTEF, Sapienza Università di Roma, Italy
[d] Facoltà di Economia, Università degli Studi Niccolò Cusano Roma, Italy

ARTICLE INFO

*Article history:*
Received 9 January 2020
Received in revised form 15 September 2020
Accepted 18 September 2020
Available online 7 October 2020

*Keywords:*
Tree-shaped facility
Variance criterion
Mean absolute deviation criterion
Conic linear programming
Semidefinite programming relaxation

ABSTRACT

In this paper we study the problem of finding a discrete subtree of minimum variance with bounded size of a given tree. We show that the problem is NP-hard on very simple trees and it remains difficult even if we consider the Mean Absolute Deviation criterion on star graphs. Referring to the variance criterion, we first propose a reformulation of the problem as an Integer Quadratic Program that, however, does not have particular structural properties that may help in finding good lower bounds on the optimal value of the problem. Then, basing on recent results on conic optimization, we investigate alternative reformulations of the problem in order to obtain some lower bounds. In particular, we exploit a continuous linear, conic, reformulation of the problem and show that this is, in fact, an exact linear reformulation of the Integer Quadratic Program. Alternatively, by referring to the upper envelope approach, we show how to obtain a lower bound for the optimal objective function value of our discrete tree-shaped facility location problem in pseudo-polynomial time.

© 2020 Elsevier B.V. All rights reserved.

### 1. Introduction

Facility location is one of the central topics in Operations Research and there is a number of papers in the literature which analyze the different (and numerous) aspects of problems arising in this field. Classical reviews on the topic are, for example, [34,45]. Extensive facility location focuses on the location of facilities that are too large to be considered as points. An exhaustive review of these kind of problems can be found in [7,41]. In addition, to the usual classification of these problems based on the space in which facilities must be located (*continuous*, *discrete* and *network* location problems), a distinguished aspect of a location problem is the choice of the objective function. In the last decades an increasing attention has been devoted to equity measures and measures that take into account the dispersion of the clients w.r.t. a facility [5,10,31,40,47]. These criteria are particularly meaningful when locating public services, which, in principle, should be available to everybody at the same accessibility level. The introduction of equity measures into location theory was first discussed in [21]. The notion of equity is also directly related to the idea of taking into account risk in location decision-making problems. In fact, since locating facilities is costly, and the decision about where locating a service may have a considerable impact on the growth of a firm, some authors introduced the notion of volatility in the classical location problems [6]. In this context, the variance of the distances traveled by a customer to a facility is considered as

* Corresponding author.
*E-mail addresses:* puerto@us.es (J. Puerto), federica.ricca@uniroma1.it (F. Ricca), andrea.scozzari@unicusano.it (A. Scozzari).

https://doi.org/10.1016/j.dam.2020.09.015
0166-218X/© 2020 Elsevier B.V. All rights reserved.

# Difusos: 2 publicaciones

Computers & Industrial Engineering 128 (2019) 170–179

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

## On a fully fuzzy framework for minimax mixed integer linear programming

Manuel Arana-Jiménez[a], Víctor Blanco[b,*]

[a] Dept. Statistics and Operational Research, Universidad de Cádiz, Spain
[b] IEMath-GR, Universidad de Granada, Spain

ARTICLE INFO

*Keywords:*
Minimax optimization
Fully fuzzy linear programming problem
Fuzzy numbers
Multiobjective optimization
Facility location

ABSTRACT

In this work, we present a modeling framework for minimax mixed 0–1 fuzzy linear problems. It is based on extending the usual rewriting of crisp minimax problems via auxiliary variables to model the maximum of a finite set of fuzzy linear functions. We prove that the considered problem can be equivalently formulated as a (crips) multiple objective mixed integer programming problem. The framework is applied to a fully fuzzy version of the capacitated center facility location problem.

### 1. Introduction

Minimax optimization is a widely studied modeling strategy in Decision Theory that has been applied in many different fields. Minimax decision policies are desirable when the decision maker is risk-averse or when the average cost is less important than ensuring smaller maximum costs. One of the most popular results involving minimax decisions is the one due to Von Neumann (1928) in which is stated that one can find Nash Equilibria in Zero-Sum games using minimax optimization. That result is considered as the cornerstone of the developments performed in Non Cooperative Game Theory in the last decades. Several applications of minimax optimization can be found in the literature: facility location (Garfinkel, Neebe, & Rao, 1977; Puerto, Rodríguez-Chía, & Tamir, 2009), transportation (Ahuja, 1986), flow (Chen, Campbell, Thomas, & Tamir, 2009), scheduling (Graham, 1966), resource allocation (Kaplan, 1974), portfolio selection (Young, 1998) only to mention a few. In general, minimax optimization problems, although its objective function is not linear, but piecewise linear, it can be rewritten as a problem with the similar shape of the original one but when instead of minimizing the maximum operator of a finite set of functions, one only need to minimize a linear objective function after adding to the problem some extra constraints and a new auxiliary variable. Hence, the complexity of the problem is inherited to minimax problems. In particular, minimax mixed integer linear programming is NP-hard as its minisum version, but polynomially solvable when the number of integer variables is fixed (Lenstra, 1983).

In this paper, we analyze minimax mixed integer linear programming problems in which some kind of imprecision is assumed to the parameters of the models and also to the decision variables. Such an imprecision is modeled by considering that the parameters and variables are fuzzy numbers. Decision making under a fuzzy environment was introduced by Bellman and Zadeh (1970) in the seventies, which has been extended to other frameworks (for instance, see Atanassov, 1986; Dong & Wan, 2018; Wan & Dong, 2014; Wan & Li, 2014, 2015; Wan, Wang, Lin, & Dong, 2015, 2016; Wan, Wang, Xu, Dong, & Tang, 2017; Wan, Qin, & Dong, 2017; Wan, Jin, & Dong, 2018 and bibliography therein for the intuitionistic fuzzy sets and applications to intuitionistic and hesitant linear programming, where decision maker exhibits some hesitation degrees for the assessments).

Such an approach is well-known and has been adopted by researchers in fuzzy optimization problems (see e.g., Abboud, Sakawa, & Inuiguchi, 1997; Campos & Verdegay, 1989; Chanas & Kuchta, 1998; Ebrahimnejad, Nasseri, Lotfi, & Soltanifar, 2010; Ganesan & Veeramani, 2006; Kasperski & Kulej, 2007; Klein, 1991; Maleki, Tata, & Mashinchi, 2000; Maleki, 2002). However, it is usual that not all the elements of a fuzzy linear problem are fuzzy sets (see Herrera, Verdegay, & Zimmermann, 1993; Herrera & Verdegay, 1995, 1991, 1996). In Lotfi, Allahviranloo, Jondabeha, and Alizadeh (2009), the authors point out that there is no general method for finding the fuzzy optimal solution of fully fuzzy linear programming (FFLP) problems, and propose a new method for solving FFLP problems with equality constraints when the parameters and variables are assumed to be symmetric fuzzy numbers. However, in Kumar, Kaur, and Singh (2011) the method proposed in Lotfi et al. (2009) is criticized because its complexity and the authors claimed out that the obtained solutions are approximated, but not exact. A new method for finding the fuzzy optimal solution of (FFLP) problems with equality constraints, with triangular fuzzy numbers involved, although they use ranking function (see Arana-Jiménez, Rufián-

* Corresponding author.
*E-mail addresses:* manuel.arana@uca.es (M. Arana-Jiménez), vblanco@ugr.es (V. Blanco).

https://doi.org/10.1016/j.cie.2018.12.029
Received 21 February 2018; Received in revised form 22 October 2018; Accepted 11 December 2018
Available online 16 December 2018
0360-8352/ © 2018 Elsevier Ltd. All rights reserved.

European Journal of Operational Research 283 (2020) 692–705

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

## On the fuzzy maximal covering location problem

Manuel Arana-Jiménez[a], Víctor Blanco[b,*], Elena Fernández[a]

[a] Department of Statistics & OR, Universidad de Cádiz, Spain
[b] IEMath-GR, Universidad de Granada, Spain

ARTICLE INFO

*Article history:*
Received 27 June 2019
Accepted 18 November 2019
Available online 21 November 2019

*Keywords:*
Location
Maximal Covering
Fuzzy Optimization
Multiobjective Optimization

ABSTRACT

This paper studies the maximal covering location problem, assuming imprecise knowledge of all data involved. The considered problem is modeled from a fuzzy perspective producing suitable fuzzy Pareto solutions. Some properties of the fuzzy model are studied, which validate the equivalent mixed-binary linear multiobjective formulation proposed. A solution algorithm is developed, based on the augmented weighted Tchebycheff method, which produces solutions of guaranteed Pareto optimality. The effectiveness of the algorithm has been tested with a series of computational experiments, whose numerical results are presented and analyzed.

© 2019 Elsevier B.V. All rights reserved.

### 1. Introduction

Covering location models have been extensively studied in the literature. Broadly speaking, in these problems there is a set of users with service demand, which can be satisfied by activating (opening) facilities *sufficiently close* to the demand points. In particular, a given demand user will be served (covered), and its demand *captured* if it is located within the coverage radius of some activated facility; that is, if its distance to some open facility does not exceed a given radius. In particular, if a set of facilities $S$ is opened, the demand of user $i$, $w_i$, will be captured if an only if $d_{ij} < R_j$ for some $j \in S$, where $d_{ij}$ denotes the distance from user $i$ to facility $j$ and $R_j$ is a given coverage radius, which may be different for each potential facility. The two seminal models in this area are the minimum-cover location problem and the maximal covering location problem. In the minimum-cover location problem, introduced by Toregas and ReVelle (1972), the objective is to find a set of facilities that covers all the demand points at minimum cardinality. Khuller, Moss, and Naor (1999) proposed the (budgeted) maximal covering location problem (MCLP), in which there is a set-up cost for each activated facility and a budget that limits the overall set-up cost that can be incurred. Such a budget constraint reduces to a cardinality constraint if all the set-up costs are equal (see Church & ReVelle, 1974). The objective is to find a set of facilities that maximizes the total covered demand.

Applications of covering location models arise in multiple fields and include the location of health care or emergency services, where successful service strongly depends on the distance from facilities to demand points, the location of signal-transmission facilities (TV, radio, cell-phone antennas, etc.), where coverage is only achieved within a certain distance from the facility, or the location of retail facilities, where the attractiveness of a facility for a potential customer clearly depends on its distance from the customer location.

The diversity of applications and the theoretical interest of the underlying optimization models have stimulated active research on the area in the last decades. The interested reader is addressed to Berman, Drezner, and Krass (2010) and references therein for an inspiring overview on the top ic.

*Coverage* and *coverage radius* are two concepts inherent to covering location, which are typically subjected to several modeling assumptions. One of the main assumptions is that a demand user is either fully covered or not covered at all. Another classical assumption is that the coverage radius that determines whether or not a demand user is covered is known. However, as discussed in Berman et al. (2010), in many applications these assumptions may be unrealistic. Examples that illustrate this weakness are, for instance, the location of health care or emergency services, or the location of signal transmission facilities. This has motivated the study of extensions of classical models, more flexible with respect to the meaning and role of *coverage* and *coverage radius*. For instance, gradual covering models mitigate the above concern by extending the *all-or-nothing* coverage assumption to the *gradual coverage* assumption, which is modeled by associating the coverage level of demand points with their distance to open facilities (Berman et al., 2010; Berman & Krass, 2002; Church & Roberts, 1984). Specifically, the captured demand of a given user is computed as $w \cdot f(d)$, where $w$ denotes the demand of the user, $d$

* Corresponding author.
*E-mail addresses:* manuel.arana@uca.es (M. Arana-Jiménez), vblanco@ugr.es (V. Blanco), elena.fernandez@uca.es (E. Fernández).

https://doi.org/10.1016/j.ejor.2019.11.036
0377-2217/ © 2019 Elsevier B.V. All rights reserved.

# Estadísticos ordenados: 3 publicaciones

STATISTICS
2019, VOL. 53, NO. 5, 1052–1081
https://doi.org/10.1080/02331888.2019.1599378

Taylor & Francis
Taylor & Francis Group

Check for updates

## Order statistics from overlapping samples: bivariate densities and regression properties

F. López-Blázquez[a], Nan-Cheng Su [b] and Jacek Wesołowski[c]

[a]Departamento Estadistica e Investigatión Operativa, Universidad de Sevilla, Sevilla, Spain; [b]Department of Statistics, National Taipei University, New Taipei City, Taiwan; [c]Faculty of Mathematics and Information Science, Warsaw University of Technology, Warszawa, Poland

**ABSTRACT**
In this paper, we are interested in the joint distribution of two order statistics from overlapping samples. We give an explicit formula for the distribution of such a pair of random variables under the assumption that the parent distribution is absolutely continuous. We are also interested in the question to what extent conditional expectation of one of such order statistic given another determines the parent distribution. In particular, we provide a new characterization by linearity of regression of an order statistic from the extended sample given the one from the original sample, special case of which solves a problem explicitly stated in the literature. It appears that to describe the correct parent distribution it is convenient to use quantile density functions. In several other cases of regressions of order statistics we provide new results regarding uniqueness of the distribution in the sample.

**ARTICLE HISTORY**
Received 11 June 2018
Accepted 20 March 2019

**KEYWORDS**
Characterization; order statistics; overlapping samples; linearity of regression; uniqueness theorem

**MSC 2010 SUBJECT CLASSIFICATIONS**
Primary: 60E05; Secondary: 62E10

### 1. Introduction

Properties of order statistics (os's) $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ based on the sample $\{X_1, X_2, \ldots, X_n\}$ of independent and identically distributed (iid) random variables with absolutely continuous distribution are widely known, see, e.g. the monographs David and Nagaraja [1] or Arnold et al. [2] for excellent reviews. Much less is known for os's which arise from different samples which have common elements. There are two special cases which until now have been studied in the literature: (1) moving os's, when the subsequent samples are of the same size and have the same size of the overlap – see, e.g. Inagaki [3], David and Rogers [4], Ishida and Kvedaras [5] or Balakrishnan and Tan [6]; (let us mention that moving samples have a long history in quality control and time series analysis – in particular, the moving median is a simple robust estimator of location and the moving range is a current measure of dispersion complementing the moving average); (2) special cases of os's from the original and extended sample which except the original sample contains a number of additional observations – see, e.g. Siddiqui [7], Tryfos and Blackmore [8], Ahsanullah and Nevzorov [9] or López-Blázquez and Salamanca-Miño [10].

**CONTACT** Nan-Cheng Su sunanchen@gmail.com

© 2019 Informa UK Limited, trading as Taylor & Francis Group

Metrika (2020) 83:699–712
https://doi.org/10.1007/s00184-019-00754-1

Check for updates

## Increasing concave orderings of linear combinations of order statistics with applications to social welfare

Antonia Castaño-Martínez[1] · Gema Pigueiras[1] · Georgios Psarrakos[2] · Miguel A. Sordo[1]

Received: 13 May 2019 / Published online: 4 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

**Abstract**
We provide in this paper sufficient conditions for comparing, in terms of the increasing concave order, some income random variables based on linear combinations of order statistics that are relevant in the framework of social welfare. The random variables under study are weighted average incomes of the poorest and, for some particular weights, their expectations are welfare measures whose integral representations are weighted areas underneath Bonferroni curves.

**Keywords** Increasing concave order · Generalized Lorenz order · Welfare measurement · Order statistics

**JEL Classification** G220

### 1 Introduction

A number of well-known cardinal measures of inequality and social welfare can be written as moments of a linear combination of order statistics [see the book by Arnold and Sarabia (2018), for plenty of examples]. Since the ranking of income distributions based on these measures may depend significantly on the particular moment or selected coefficients, sometimes it may be preferable to rank them by comparing, by using stochastic orderings, the underlying linear combinations of order statistics, which is the focus of this paper. Many results involving stochastic comparisons of functions of order statistics can be found in the books by Shaked and Shanthikumar (2007) and Belzunce et al. (2016).

✉ Miguel A. Sordo
mangel.sordo@uca.es

[1] Department of Statistics and Operations Research, University of Cádiz, Cádiz, Spain

[2] Department of Statistics and Insurance Science, University of Piraeus, Piraeus, Greece



*mathematics*

MDPI

*Article*

## On the Increasing Convex Order of Relative Spacings of Order Statistics

**Antonia Castaño-Martínez, Gema Pigueiras * and Miguel A. Sordo**

Dpto. Estadística e Investigación Operativa, Universidad de Cádiz, 11510 Puerto Real, Spain; antonia.castano@uca.es (A.C.-M.); mangel.sordo@uca.es (M.A.S.)
* Correspondence: gema.pigueiras@uca.es

**Abstract:** Relative spacings are relative differences between order statistics. In this context, we extend previous results concerning the increasing convex order of relative spacings of two distributions from the case of consecutive spacings to general spacings. The sufficient conditions are given in terms of the expected proportional shortfall order. As an application, we compare relative deprivation within some parametric families of income distributions.

**Keywords:** order statistics; expected proportional shortfall order; relative spacings; deprivation

check for updates

**Citation:** Castaño-Martínez, A.; Pigueiras, G.; Sordo, M.A. On the Increasing Convex Order of Relative Spacings of Order Statistics. *Mathematics* **2021**, *9*, 618. https://doi.org/10.3390/math9060618

Academic Editors: Antonio Di Crescenzo and Lev Klebanov

Received: 19 February 2021
Accepted: 11 March 2021
Published: 15 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

### 1. Introduction

Let $X_1, X_2, ..., X_n$ be $n$ independent and identically distributed (i.i.d.) random variables, and let $X_{1:n}, ..., X_{n:n}$ be the corresponding order statistics, which are the $X_i$s arranged in a non-decreasing order (that is, $X_{1:n} \le ... \le X_{n:n}$). Differences between order statistics are called spacings. The sequences of spacings, given by

$$S^X_{i,j,n} = X_{j:n} - X_{i:n}, \quad 1 \le i < j \le n, \tag{1}$$

and relative spacings, given by

$$R^X_{i,j,n} = \frac{X_{j:n} - X_{i:n}}{X_{i:n}}, \quad 1 \le i < j \le n, \tag{2}$$

have found applicability in fields such as reliability, auction theory, insurance, economics and other areas of applied mathematics. Given two probability models, it is often of interest to study which one has larger spacings under different probabilistic criteria (see, e.g., Kochar [1], Kochar and Rojo [2], Hu and Wei [3], Kochar et al. [4], Kochar and Xu [5], Genest et al. [6], Mao and Hu [7], Zhao and Zhang [8], Torrado and Lillo [9], Belzunce et al. [10] and Ding et al. [11]). There has been less research conducted on the comparison of relative spacings, which is the focus of this paper. As far as we know, the earliest reference on this subject is the work of Belzunce et al. [12], who gave conditions for the comparison of relative spacings between consecutive generalized order statistics from two random vectors in terms of several stochastic orders. Some of these conditions were given in terms of the expected proportional shortfall order and the increasing convex order, which are defined below.

**Definition 1.** *Let $X$ and $Y$ be two random variables with respective distribution functions $F$ and $G$ and survival functions $\overline{F}$ and $\overline{G}$, respectively. Let $F^{-1}$ and $G^{-1}$ be the corresponding quantile functions given by $F^{-1}(u) = \inf\{x : F(x) \ge u\}$ and $G^{-1}(u) = \inf\{x : G(x) \ge u\}$, for $0 \le u \le 1$. Then,*

*(a) We say that $X$ is less than $Y$ in the increasing convex order (denoted by $X \le_{icx} Y$) if*

$$E[\phi(X)] \le E[\phi(Y)],$$

# Portfolio: 3 publicaciones

Computers and Operations Research 117 (2020) 104891

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## Clustering and portfolio selection problems: A unified framework

Justo Puerto[a,b], Moisés Rodríguez-Madrena[a,b], Andrea Scozzari[c,*]

[a] Institute of Mathematics University of Seville (IMUS), Seville, Spain
[b] Department of Statistics and OR, Universidad de Sevilla, 41012 Seville, Spain
[c] Faculty of Economics, Università degli Studi Niccolò Cusano Roma, Italy

ARTICLE INFO

*Article history:*
Received 17 June 2019
Revised 27 November 2019
Accepted 13 January 2020
Available online 15 January 2020

*Keywords:*
Portfolio selection
Clustering
*p*-median problem on networks
Conditional Value at Risk
Multicretiria optimization

ABSTRACT

Given a set of assets and an investment capital, the classical portfolio selection problem consists in determining the amount of capital to be invested in each asset in order to build the most profitable portfolio. The portfolio optimization problem is naturally modeled as a mean-risk bi-criteria optimization problem where the mean rate of return of the portfolio must be maximized whereas a given risk measure must be minimized. Several mathematical programming models and techniques have been presented in the literature in order to efficiently solve the portfolio problem. A relatively recent promising line of research is to exploit clustering information of an assets network in order to develop new portfolio optimization paradigms. In this paper we endow the assets network with a metric based on correlation coefficients between assets' returns, and show how classical location problems on networks can be used for clustering assets. In particular, by adding a new criterion to the portfolio selection problem based on an objective function of a classical location problem, we are able to measure the effect of clustering on the selected assets with respect to the non-selected ones. Most papers dealing with clustering and portfolio selection models solve these problems in two distinct steps: cluster first and then selection. The innovative contribution of this paper is that we propose a Mixed-Integer Linear Programming formulation for dealing with this problem in a unified phase. The effectiveness of our approach is validated reporting some computational experiments on some real financial datasets.

© 2020 Elsevier Ltd. All rights reserved.

### 1. Introduction and motivation

Portfolio selection problem is among most the widely studied problems in financial literature since the seminal paper of Markowitz in 1952 Markowitz (1952). He formalized the notion of diversification in investing and used the variance of asset prices as a proxy of risk. In spite of its success, the model proposed by Markowitz has received many criticisms, first and foremost due to the large estimation errors on the vector of expected returns and on the covariance matrix Chopra and Ziemba (2013); Jobson and Korkie (1980); Merton (1980). Hence, different directions have been proposed in the literature exploiting several strategies that consider improved estimates of correlation coefficients Elton et al. (2006), alternative risk measures Artzner et al. (1999); Bellini and Di Bernardino (2017), effective diversification approaches (see, e.g., Cesarone and Tardella (2017); De Miguel et al. (2007), Maillard et al., (2010) and the references therein). As a consequence, new mathematical programming models and techniques were proposed in the literature in order to efficiently solve the portfolio selection problems incorporating the above mentioned different strategies (the interested reader can refer to Mansini et al. (2015)).

At the early 2000s, some authors started in tackling the asset allocation problem via graph theory in which assets are identified by vertices of a complete graph $G$ and distances $d_{ij}$ assigned to pair of assets $i$ and $j$ incorporate the dependency structure of returns. Mantegna Mantegna (1999) was one of the first to construct asset graphs based on stock price correlations in order to detect the hierarchical organization inside a stock market. He showed that a minimal spanning tree of $G$ provides an arrangement of stocks which selects the most relevant connections among them. Moreover, the minimal spanning tree provides, in a direct way, the subdominant ultrametric hierarchical organization of the assets in a market. Based on this idea, Onnela et al. J.-P. Onnela et al. (2004) study the topology of stock networks that can be constructed by using return correlations. The aim is to disclose the relevant correlations from the correlation matrix. They start from an empty graph with no edges where the vertices correspond to assets. Then, one by one, they insert edges between the vertices according to the rank of their correlation strength, thus determining different stock graphs

* Corresponding author.
*E-mail addresses:* puerto@us.es (J. Puerto), madrena@us.es (M. Rodríguez-Madrena), andrea.scozzari@unicusano.it (A. Scozzari).

https://doi.org/10.1016/j.cor.2020.104891
0305-0548/© 2020 Elsevier Ltd. All rights reserved.

European Journal of Operational Research 284 (2020) 712–727

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

## Portfolio problems with two levels decision-makers: Optimal portfolio selection with pricing decisions on transaction costs☆

Marina Leal[a,b,*], Diego Ponce[a,c], Justo Puerto[a,b]

[a] Institute of Mathematics of the University of Seville (IMUS), Spain
[b] Department of Statistics and Operations Research, Faculty of Mathematics, University of Seville, Spain
[c] CIRRELT and Department of Mechanical, Industrial and Aerospace Engineering, Faculty of Engineering and Computer Science, Concordia University, Montreal, Canada

ARTICLE INFO

*Article history:*
Received 31 July 2019
Accepted 26 December 2019
Available online 9 January 2020

*Keywords:*
Finance
Portfolio optimization
Bilevel optimization
Transaction costs
Conditional Value at Risk (CVaR)

ABSTRACT

This paper presents novel bilevel leader-follower portfolio selection problems in which the financial intermediary becomes a decision-maker. This financial intermediary decides on the unit transaction costs for investing in some securities, maximizing its benefits, and the investor chooses his optimal portfolio, minimizing risk and ensuring a given expected return. Hence, transaction costs become decision variables in the portfolio problem, and two levels of decision-makers are incorporated: the financial intermediary and the investor. These situations give rise to general Nonlinear Programming formulations in both levels of the decision process. We present different bilevel versions of the problem: financial intermediary-leader, investor-leader, and social welfare; besides, their properties are analyzed. Moreover, we develop Mixed Integer Linear Programming formulations for some of the proposed problems and effective algorithms for some others. Finally, we report on some computational experiments performed on data taken from the Dow Jones Industrial Average, and analyze and compare the results obtained by the different models.

© 2019 Elsevier B.V. All rights reserved.

### 1. Introduction

The classical model in portfolio optimization was originally proposed by Markowitz (1952). This model has served as the initial point for the development of modern portfolio theory. Over time, portfolio optimization problems have become more realistic, incorporating real-life aspects that make the resulting portfolios more cost-effective than the alternatives that do not consider them (Castro, Gago, Hartillo, Puerto, & Ucha, 2011; Kolm, Tütüncü, & Fabozzi, 2014; Lynch & Tan, 2011; Mansini, Ogryczak, & Speranza, 2014; 2015b). Transaction costs can be seen as one of these important actual features to be included in portfolio optimization. These costs are those incurred by the investors when buying and selling assets on financial markets, charged by the brokers, the financial institutions or the market makers playing the role of intermediary. Transaction costs usually include banks and brokers' commissions, fees, etc. These commissions or fees have a direct impact on the portfolio, especially for individual or small investors, since they will determine the net returns, reducing them and decreasing also the budget available for future investments (Baule, 2010; Baumann & Trautmann, 2013; Liu & Loewenstein, 2002).

To the best of our knowledge, in the existing literature, transaction costs are assumed to be given (Davis & Norman, 1990; Korn, 1998; Lobo, Fazel, & Boyd, 2007; Magill & Constantinides, 1976; Mansini et al., 2014; 2015b; Morton & Pliska, 1995). They can be a fixed cost applied to each selected security in the portfolio; or a variable cost to be paid which depends on the amount invested in each security included in the portfolio (see e.g. (A. Valle, Meade, & E. Beasley, 2014; Baule, 2010; Baumann & Trautmann, 2013; Kellerer, Mansini, & Speranza, 2000; Mansini et al., 2014; 2015b; Woodside-Oriakhi, Lucas, & Beasley, 2013) and the references therein). This dependence can be proportional to the investment or given by a fixed cost that is only charged if the amount invested exceeds a given threshold, or by some other functional form (see e.g. Baule, 2010; Konno, Akishino, & Yamamoto, 2005; Mansini et al., 2014; Mansini, Ogryczak, & Speranza, 2015b; Thi, Moeini, & Dinh, 2009 and the references therein). But in any case, unit transaction costs are known and predetermined in the optimization process. Nevertheless, it is meaningful to analyze the situations where transaction costs can be decision variables set by financial institutions so that they are trying to maximize its profit

☆ This research has been funded by the Spanish Ministry of Science and Technology project MTM2016-74983-C2-1-R.
* Corresponding author at: Institute of Mathematics of the University of Seville (IMUS), Spain.

https://doi.org/10.1016/j.ejor.2019.12.039
0377-2217/© 2019 Elsevier B.V. All rights reserved.

Omega 102 (2021) 102353

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

## Global optimization for bilevel portfolio design: Economic insights from the Dow Jones index☆

Julio González-Díaz[a,b,c,*], Brais González-Rodríguez[a,b], Marina Leal[d,e], Justo Puerto[e]

[a] Department of Statistics, Mathematical Analysis and Optimization and IMAT, University of Santiago de Compostela, Spain
[b] MODESTYA Research Group, Santiago de Compostela, Spain
[c] ITMATI (Technological Institute for Industrial Mathematics), Santiago de Compostela, Spain
[d] ALOP, Trier University, Germany
[e] IMUS, University of Seville, Spain

ARTICLE INFO

*Article history:*
Received 2 January 2020
Accepted 20 October 2020
Available online 25 October 2020

*Keywords:*
Portfolio design
Nonlinear programming
Bilevel optimization
Global Optimization

ABSTRACT

This paper deals with a portfolio selection problem with transaction costs and two levels of decision-making. It is assumed that the decision making structure is twofold: there is a broker-dealer that controls the fees to be charged on the different securities in order to maximize his benefit and there is an investor who chooses his portfolio trying to minimize risk while ensuring a minimum level of return. This structure gives rise to an implicit hierarchical competition that consists in anticipating the rational decision of the other agent in order to optimize the decision-makers' own criteria. We analyze different situations depending on who is first in the hierarchy: the broker-dealer or the investor. We present different nonlinear and nonconvex mathematical programming models for the different situations and develop an extensive computational study in which we discuss the ensuing economic insights for the models based on Dow Jones index data.

© 2020 Elsevier Ltd. All rights reserved.

### 1. Introduction

Mathematical optimization problems are pervasive in the fields of economics and management science, and the importance of developing realistic models to delve into the understanding of complex economic settings has long been recognized. In particular, bilevel optimization models have attracted a lot of attention since the pioneering work by von Stackelberg [40]. However, realistic models often result in difficult optimization problems and, thus, a compromise is required between realism and solvability of the model. Quite often the analyst is confronted with a nonlinear and nonconvex optimization problem, in which finding globally optimal solutions may be an extremely challenging task.

One of the contributions of this paper is to illustrate how proper modeling skills may allow to efficiently solve complex optimization problems, enabling the development of qualitative and quantitative economic analysis in problems that, otherwise, would be hard to tackle. In order to do so, we formally study two novel bilevel portfolio design problems and show how one can rely on well-established approaches to reformulate them as single-level problems that can be fully solved with state-of-the-art optimization software.

Historically, the main criterion to design an optimal portfolio was to find the configuration of assets that generated the highest expected return. However, this perspective changed in 1952, when Harry Markowitz introduced a new variable along with the expected return: the risk of each portfolio [28]. Thereafter, analysts began to incorporate a risk-return trade-off in their models.

The model proposed by Markowitz only focuses on finding an optimal portfolio from the investor's point of view. The literature is plenty of papers with similar approaches, as for instance [6,7,12–14,20,26,31,32], to mention a few. However, real markets are generally more complex, since there is another decision-maker who can set fees on the transactions of the securities to profit from anticipating the rational behavior of investors. Transaction costs, those incurred by the investors when buying and selling assets on financial markets, have been widely studied in papers such as [3,4,18,26,43,45], among others. In this work we study these situations in which the investor, when deciding his optimal portfolio, has to consider the transaction fee to pay to an agent: the broker-dealer, hereafter, the broker. The broker makes decisions regard-

☆ Area: Optimization; Health Care Systems and Applications. This manuscript was processed by Associate Editor Jay Rosenberger.
* Corresponding author at: Department of Statistics, Mathematical Analysis and Optimization and IMAT, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain.
*E-mail address:* julio.gonzalez@usc.es (J. González-Díaz).

https://doi.org/10.1016/j.omega.2020.102353
0305-0483/© 2020 Elsevier Ltd. All rights reserved.

# Scheduling: 2 publicaciones

Expert Systems With Applications 141 (2020) 112959

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

## GRASP algorithm for the unrelated parallel machine scheduling problem with setup times and additional resources

Juan C. Yepes-Borrero [a,*], Fulgencia Villa [a], Federico Perea [a], Juan Pablo Caballero-Villalobos [b]

[a] Grupo de Sistemas de Optimización Aplicada, Instituto Tecnológico de Informática, Ciudad Politécnica de la Innovación, Edifico 8G, Acc. B., Universitat Politècnica de València, Camino de Vera s/n, València, 46021, Spain
[b] Departamento de Ingeniería Industrial, Pontificia Universidad Javeriana, Calle 40 No. 5-50 Ed. José Gabriel Maldonado, S.J. Bogotá, Colombia

ARTICLE INFO

*Article history:*
Received 3 December 2018
Revised 16 September 2019
Accepted 17 September 2019
Available online 18 September 2019

*Keywords:*
Unrelated parallel machines
Scheduling
Sequence dependent setup times
Makespan
Additional resources
GRASP

ABSTRACT

This paper provides practitioners with new approaches for solving realistic scheduling problems that consider additional resources, which can be implemented on expert and intelligent systems and help decision making in realistic settings. More specifically, we study the unrelated parallel machine scheduling problem with setup times and additional limited resources in the setups (UPMSR-S), with makespan minimization criterion. This is a more realistic extension of the traditional problem, in which the setups are assumed to be done without using additional resources (e.g. workers). We propose three metaheuristics following two approaches: a first approach that ignores the information about additional resources in the constructive phase, and a second approach that takes into account this information about the resources. Computational experiments are carried out over a benchmark of small and large instances. After the computational analysis we conclude that the second approach shows an excellent performance, overcoming the first approach.

© 2019 Elsevier Ltd. All rights reserved.

### 1. Introduction

Nowadays, companies face more changing and volatile environments, where increased competitiveness and personalization of products play a key role. Therefore, companies need smart tools that help their decision-making process to become more efficient and effective. In this paper, we propose intelligent methods to solve hard decision-making problems using optimization techniques, in order to contribute to the smart factory concept,[1] that is, sustainable and intelligent industries. Scheduling plays a very important role in industry. There are many different scheduling problems modeling different types of production processes. In many cases, factories need flexibility in their productive processes to achieve a higher personalization of their products. This flexibility may include the need of additional resources, which makes scheduling problems much more difficult to solve.

Among the scheduling problems that appear in industrial processes, one of them is the so called Unrelated Parallel Machines scheduling problem (UPM), where a set of jobs have to be processed by a set of parallel machines. As the machines are unrelated, the processing time of a job may be different depending on the machine the job is assigned to. Recently, many studies have been conducted on the Unrelated Parallel Machine scheduling problem with Setup times between jobs (UPMS), which is an extension of the UPM problem. The UPMS arises when machines need to be reconfigured after the processing of one job, and before the processing of the next one.

In the literature of the UPMS, no constraint is normally assumed made on the number of setups that can be done at the same time. In other words, at any point in time one may do as many setups as needed. Arguably, it is common that machines process jobs automatically, without the help of extra resources. However, we want to underline that in manufacturing environments, the machine setups between jobs is usually done by additional resources (e.g. workers). Since the number of these available resources is typically limited, the number of setups that can be done at the same time is limited. Therefore, an extension, and more realistic approach to the UPMS is the Unrelated Parallel Machine scheduling problem with setup times and additional Resources in the Setups (UPMSR-S), which is the problem introduced in this paper.

Among the variety of objectives considered in scheduling, one of the most studied is the minimization of the makespan, denoted by $C_{max}$. The makespan is defined as the completion time of the

* Corresponding author.
*E-mail addresses:* juancamiloyepes@gmail.com (J.C. Yepes-Borrero), mfuvilju@eio.upv.es (F. Villa), perea@eio.upv.es (F. Perea), juan.caballero@javeriana.edu.co (J.P. Caballero-Villalobos).

[1] https://www.capgemini.com/resources/preparing-for-smart-factories/.

https://doi.org/10.1016/j.eswa.2019.112959
0957-4174/© 2019 Elsevier Ltd. All rights reserved.

European Journal of Operational Research 292 (2021) 443–455

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

## Bi-objective parallel machine scheduling with additional resources during setups

Juan C. Yepes-Borrero [a,*], Federico Perea [b], Rubén Ruiz [a], Fulgencia Villa [a]

[a] Grupo de Sistemas de Optimización Aplicada, Instituto Tecnológico de Informática, Ciudad Politécnica de la Innovación, Edifico 8G, Acc. B. Universitat Politècnica de València, Camino de Vera s/n, 46021 València, Spain
[b] Departamento de Matemática Aplicada II, Universidad de Sevilla, Escuela Politécnica Superior, Calle Virgen de África 7, Sevilla 41011, Spain

ARTICLE INFO

*Article history:*
Received 27 April 2020
Accepted 31 October 2020
Available online 7 November 2020

*Keywords:*
Scheduling
Parallel machine
Sequence dependent setup times
Multicriteria optimization
Resources

ABSTRACT

We present a bi-objective parallel machine scheduling problem with machine and job sequence dependent setup times, with the additional consideration of resources needed during setups. The availability of such resources is limited. This models many practical situations where setup times imply, for example, cleaning and/or the reconfiguration of productive equipment. These setups are performed by personnel, who are of course limited in number. The objectives considered are the minimization of the makespan and the minimization of the number of resources. Fewer available resources reduce production costs but inevitably increase the makespan. On the contrary, a greater number of resources increase costs but allow for more setups to be done in parallel and a reduced makespan. An algorithm based on iterated greedy approaches is proposed to search for the Pareto front of the problem. This algorithm is compared with state-of-the art methods adapted to the problem. Computational experiments, supported by statistical analyses, indicate that the proposed approach outperforms all other tested procedures.

© 2020 Elsevier B.V. All rights reserved.

### 1. Introduction

Operations research methods are being used more and more in the management of a variety of production planning and scheduling problems. Roughly speaking, scheduling problems consist of deciding which machines should process which jobs, and when such processing should be done. Out of the many scheduling problems proposed in the literature (flowshop, job shop, etc.), in this paper we deal with a generalized version of the unrelated parallel machine scheduling problem.

Many practical manufacturing settings can be modelled as a parallel machine scheduling problem. More specifically, multi-stage production systems frequently have a bottleneck stage that can be adequately solved with parallel scheduling models. Most of the literature, however, is centered around simplified models that lack many of the features needed for successful practical application. That is why more realistic models, and suitable algorithms to solve them, are needed. The basic unrelated parallel machine scheduling problem (referred to as UPM) deals with the scheduling of a set of jobs that need to be processed by exactly one of the available machines. Note that in the UPM, the processing time of a job depends on the machine to which it is assigned. This is already a generalization of the simpler identical parallel machine scheduling problem. The interested reader is referred to Fanjul-Peyro and Ruiz (2010) for a more detailed explanation of this UPM problem.

Machines are not the only resources employed in production shops. Many manufacturing environments require additional resources, typically machine operators, but also molds, jigs, fixtures, etc. For example, when workers are needed to be present during the processing of jobs, a new and more challenging problem than the UPM appears, called the Unrelated Parallel Machine scheduling problem with additional Resources in the Processes, denoted as UPMR-P.

Although Allahverdi (2015) shows that the literature including setup times is quite extensive, there still is a large portion of the scheduling literature that assumes that setup times are not sequence dependent, and are included in the processing times or just ignored. Nevertheless, setup times are ubiquitous in real settings. Both the UPM and UPMR-P assume that machines are ready to process a job right after they have finished processing the job before. This is not always realistic since machines may need reconfiguration, adjustments, cleaning, etc. between the processing of any two consecutive jobs. Furthermore, setup times often depend on the machine and on the job sequence. When setups are explicitly considered we have the UPMS (Unrelated Parallel Machine scheduling problem with Setups). A recent work is

* Corresponding author.
*E-mail addresses:* juayebor@posgrado.upv.es (J.C. Yepes-Borrero), perea@us.es (F. Perea), rruiz@eio.upv.es (R. Ruiz), mfuviljul@eio.upv.es (F. Villa).

https://doi.org/10.1016/j.ejor.2020.10.052
0377-2217/© 2020 Elsevier B.V. All rights reserved.

# Minimax regret: 3 publicaciones

Computers & Operations Research 77 (2017) 11–19

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

## A minimum expected regret model for the shortest path problem with solution-dependent probability distributions☆

Eduardo Conde

*Department of Statistics and Operations Research, Faculty of Mathematics, Campus Reina Mercedes, 41011 University of Seville, Seville, Spain*

ARTICLE INFO

*Article history:*
Received 13 July 2015
Received in revised form
21 April 2016
Accepted 11 July 2016
Available online 29 July 2016

*Keywords:*
Robustness and sensitivity analysis
Stochastic programming

ABSTRACT

We consider an optimization problem in which the cost of a feasible solution depends on a set of unknown parameters (scenario) that will be realized. In order to assess the cost of implementing a given solution, its performance is compared with the optimal one under each feasible scenario. The positive difference between the objective values of both solutions defines the regret corresponding to a fixed scenario. The proposed optimization model will seek for a compromise solution by minimizing the expected regret where the expectation is taken respect to a probability distribution that depends on the same solution that is being evaluated, which is called solution-dependent probability distribution. We study the optimization model obtained by applying a specific family of solution-dependent probability distributions to the shortest path problem where the unknown parameters are the arc lengths of the network. This approach can be used to generate new models for robust optimization where the degree of conservatism is calibrated by using different families of probability distributions for the unknown parameters.

© 2016 Elsevier Ltd. All rights reserved.

### 1. Introduction

Optimization problems with uncertainty in the objective function or in the set of constraints have received increasing attention over the last decades due to the importance of their practical applications (see, e.g., [5–8,10,11]). Robustness Analysis and Stochastic Programming are two of the theoretical frameworks developed in Mathematical Programming to manage uncertainty or risk in order to find compromise decisions with a good behavior under any likely input data.

In the stochastic approach, it is assumed the knowledge of the probability distribution of the vector of unknown parameters. Every realization of this random vector defines a scenario associated to a deterministic instance of the optimization problem. Under this paradigm, the decision maker adopts a compromise solution by minimizing the expected cost.

In the robust optimization models, the scenarios are described without any information about how likely they can occur. The worst-case approach is one of the techniques within this framework. Its objective is to find a solution that performs reasonably well for all the scenarios. In this context, the minmax regret optimization model aims at obtaining a solution minimizing the maximum deviation (regret) between the cost of a given solution and the optimal cost under any possible scenario.

The stochastic approach is suitable when the decision maker wants to implement the obtained solution in repetitive situations. On the other hand, the robust optimization model is more conservative trying to hedge the system against the unknown scenario that may happen in specially sensitive situations needing precautionary measures (environmental interventions, public health actions, emergency contingency plans,…).

However, in some cases, the expected cost and the maximum regret can be viewed as particular instances of a family of objective functions obtained by taking the expectation according to distributions of probability that depend on the solution which is evaluated. In other words, this model introduces the possibility of using a sort of subjective probabilities that model the behavior of the uncertain vector of parameters from the point of view of the feasible solution that is considered to be implemented. The resulting approach enables us to generate new optimization models with decision criteria that reflect certain compromise between the above two perspectives and could give rise to new models used in practical situations where none of them may fit the decision maker needs (see e.g. [9]).

In the context of robust optimization, there exists a precedent to this idea and was proposed by Averbakh in [4], where a new minmax regret model is studied with a set of scenarios that is *induced* by the choice of the feasible solution. This paper considers subset-type optimization problems, where the feasible solutions

☆This research has been funded by the Spanish Ministry of Science and Technology project MTM2013-46962-C2-1-P.
*E-mail address:* educon@us.es

http://dx.doi.org/10.1016/j.cor.2016.07.007
0305-0548/© 2016 Elsevier Ltd. All rights reserved.

Computers and Operations Research 85 (2017) 1–11

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

## Minmax regret combinatorial optimization problems with investments☆

Eduardo Conde*, Marina Leal

*Department of Statistics and Operations Research, Faculty of Mathematics. Campus Reina Mercedes, 41011. University of Seville. Seville, Spain*

ARTICLE INFO

*Article history:*
Received 18 June 2016
Revised 16 March 2017
Accepted 16 March 2017
Available online 24 March 2017

*Keywords:*
Minmax regret models
Robustness and sensitivity analysis
Shortest path problem

ABSTRACT

A new minmax regret optimization model in a system with uncertain parameters is proposed. In this model it is allowed to make investments before a minmax regret solution is implemented in order to modify the source or the nature of the existing uncertainty. Therefore, it is allowed to spend resources in order to change the basic cost structure of the system and take advantage of the modified system to find a robust solution. Some properties of this model allow us to have proper Mathematical Programming formulations that can be solved by standard optimization packages. As a practical application we consider the shortest path problem in a network in which it is possible to modify the uncertainty intervals for the arc costs by investing in the system. We also give an approximate algorithm and generalize some existing results on constant factor approximations.

© 2017 Elsevier Ltd. All rights reserved.

### 1. Introduction

In this paper a new minmax regret optimization model (Kasperski, 2008; Kouvelis and Yu, 1997) is studied. The optimization model must take into account a set of unknown parameters which determine the efficiency of the proposed solution. Each realization of these parameters defines, what is called in what follows, a scenario. We assume that these parameters, usually costs, take values on a given set of intervals and it is not supposed any probabilistic information about their behavior. In order to assess the cost of implementing a given solution, its performance is compared with the optimal one under each feasible scenario. The positive difference between the objective values of both solutions defines the regret corresponding to a fixed scenario. The optimization model seeks a compromise solution by minimizing the maximum regret.

Most optimization models considered in the literature (see e.g. Aissi et al., 2009b and the references therein) rule out the fact of having some kind of control over the set of unknown parameters. However, the source of uncertainty could be modified or the degree of knowledge about the possible scenarios could change as a consequence of particular actions carried out in the system. For instance, if we consider a problem in which the demand of a customer is modeled as an uncertain parameter and we assume that this parameter can take on any value from an interval fixed beforehand, then the length of this uncertainty interval can be reduced by performing a prospective study and obtaining some extra information about its behavior.

It can be logically assumed that this kind of actions that modify the uncertainty set of possible scenarios involves certain costs in terms of a set of limited resources. On the one hand, these costs can be considered as an investment in the system with the purpose of gaining insight into the mechanism involved in its behavior and obtaining, in this way, more information about the uncertain elements (costs, capacities, lengths, times, etc.). On the other hand, the investment can also have a different purpose, the available resources could be spent to modify the nature of the system itself. For example, if we consider a communication network with uncertain parameters, like transfer speed or capacity of the connections, one can invest in order to increase this transfer speed or the capacity of a given connection or even invest in the construction of new connections. In all these cases we are investing in the system in order to modify the range of possible values of the uncertain parameters.

Then, the problem will consist in determining how to invest the available resources in order to obtain a robust solution *as close as possible* to the optimal one under whatever scenario that may occur in the transformed system. Therefore, the purpose of the investments is to protect the decision-maker, in some way, against the uncertainty of implementing a feasible solution without the knowledge of the effects that the behavior of the system could have on it.

☆ The research activities have been supported by projects FQM-5849 (Junta de Andalucía/FEDER) and MTM2013-46962-C2-1-P, MTM2016-74983-C2-1-R (MICINN, Spain). This support is gratefully acknowledged.
* Corresponding author.
*E-mail addresses:* educon@us.es (E. Conde), mleal3@us.es (M. Leal).

http://dx.doi.org/10.1016/j.cor.2017.03.007
0305-0548/© 2017 Elsevier Ltd. All rights reserved.

Discrete Applied Mathematics 239 (2018) 193–199

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Note

## Improved complexity results for the robust mean absolute deviation problem on networks with linear vertex weights

Igor Averbakh [a,b,*], Oded Berman [b], Marina Leal [b,c]

[a] *Department of Management, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C 1A4, Canada*
[b] *Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S3E6, Canada*
[c] *Department of Statistics and Operations Research, Faculty of Mathematics, Campus Reina Mercedes, 41011, University of Seville, Seville, Spain*

ARTICLE INFO

*Article history:*
Received 28 November 2016
Received in revised form 3 August 2017
Accepted 28 December 2017
Available online 1 February 2018

*Keywords:*
Minmax regret
Robust location
Facility location under uncertainty
Network algorithms

ABSTRACT

In a recent paper Lopez-de-los-Mozos et al. (2013), an algorithmic approach was presented for the robust (minmax regret) absolute deviation single-facility location problem on networks with node weights which are linear functions of an uncertain or dynamically changing parameter. The problem combines the mean absolute deviation criterion, which is the weighted average of the absolute deviations of individual customer–facility distances from the mean customer–facility distance and is one of the standard measures of "inequality" between the customers, with the minmax regret approach to optimization under uncertainty. The uncertain data are node weights (demands) which are assumed to change in a correlated manner being linear functions of a single uncertain parameter. The analysis in Lopez-de-los-Mozos et al. (2013) presented complexity bounds that are polynomial but too high to be of practical value. In this note, we present improvements of the analysis that significantly reduce the computational complexity bounds for the algorithm.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

In the last two decades, much attention has been focused on the minmax regret approach to optimization problems with data uncertainty (e.g., [2,3]), where it is required to find a solution with an objective function value reasonably close to the optimal one for all possible realizations of uncertain parameters. The literature on minmax regret facility location largely focuses on minmax regret versions of classical location objectives such as the median or center objectives. In [4], the minmax regret version of a location problem with an equity-based objective was considered. The mean absolute deviation is the weighted average of the absolute deviations of customer–facility distances from the mean distance between the facility and the customers, and is a measure of "inequality" between the customers. Lopez-de-los-Mozos et al. [4] consider this objective in the context of dynamic evolution (or linearly-correlated uncertainty) of node weights which are assumed to be linear functions of a changing parameter (time, for instance), and study the minmax and minmax regret versions of the problem. For the minmax regret version, they develop a polynomial time algorithm, which, in spite of being polynomial, cannot be considered practically useful due to very high order of complexity obtained by the analysis in [4] (slightly higher than $O(n^8 \log n)$ for trees and slightly higher than $O(m^3 n^8 \log n)$ for general networks, where $n$ and $m$ are the numbers of vertices and edges, respectively). In this paper, we present improvements of the analysis that reduce the complexity bounds to $O(m^2 n^3 (\log^* n)^2 \log n)$ for general networks, $O(n^4 (\log^* n)^2 \log n)$ for trees, and $O(n^3 (\log^* n)^2 \log n)$ for paths, where $\log^* n$ is

* Corresponding author at: Department of Management, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C 1A4, Canada.
*E-mail addresses:* averbakh@utsc.utoronto.ca (I. Averbakh), berman@rotman.utoronto.ca (O. Berman), mleal3@us.es (M. Leal).

https://doi.org/10.1016/j.dam.2017.12.036
0166-218X/© 2018 Elsevier B.V. All rights reserved.

# Tomografía computerizada: 1 public.

FULL PAPER

Particle & Particle Systems Characterization
www.particle-journal.com

## Optimization of STEM-HAADF Electron Tomography Reconstructions by Parameter Selection in Compressed Sensing Total Variation Minimization-Based Algorithms

*Juan M. Muñoz-Ocaña, Ainouna Bouziane, Farzeen Sakina, Richard T. Baker, Ana B. Hungría, Jose J. Calvino, Antonio M. Rodríguez-Chía, and Miguel López-Haro**

**A novel procedure to optimize the 3D morphological characterization of nanomaterials by means of high angle annular dark field scanning-transmission electron tomography is reported and is successfully applied to the analysis of a metal- and halogen-free ordered mesoporous carbon material. The new method is based on a selection of the two parameters ($\mu$ and $\beta$) which are key in the reconstruction of tomographic series by means of total variation minimization (TVM). The parameter-selected TVM reconstructions obtained using this approach clearly reveal the porous structure of the carbon-based material as consisting of a network of parallel, straight channels of ≈6 nm diameter ordered in a honeycomb-type arrangement. Such an unusual structure cannot be retrieved from a TVM 3D reconstruction using default reconstruction values. Moreover, segmentation and further quantification of the optimized 3D tomographic reconstruction provide values for different textural parameters, such as pore size distribution and specific pore volume that match very closely with those determined by macroscopic physisorption techniques. The approach developed can be extended to other reconstruction models in which the final result is influenced by parameter choice.**

### 1. Introduction

3D morphological analysis of nanostructured materials by means of electron tomography (ET) is currently a well-established approach, which provides meaningful information out of the reach of conventional 2D transmission electron microscopy/scanning transmission electron microscopy (TEM/STEM) analysis.[1]

In the ET experiment, a series of 2D images is acquired at different tilts of the sample under analysis with respect to the incoming electron beam, usually around only one axis. Back-projection of the information contained in each of these images by different mathematical procedures (algorithms) allows the retrieval of the 3D morphology of the object, which produced the set of recorded 2D projections. The size of the tilt range and the total number of projections (tilt increment) influence the quality of the reconstruction in terms of resolution and contribution of image artifacts, like elongation of the object in the direction perpendicular to the tilting axis (the so-called missing wedge effect).[2] Therefore, maximizing within the achievable experimental limits, these two parameters were initially considered as the essential ingredients of the recipe to obtain optimum tomographic reconstructions.

Nevertheless, the development of analytical STEM tomography[3] as well as the need to minimize, especially in the case of beam sensitive materials, the influence of beam damage, have contributed to focus most of the research in the field on the development of methodologies. The main goal of these methodologies is to keep or even to improve the quality of the reconstructed volumes in experiments in which the amount of available information becomes, often severely, reduced.

Different strategies have been proposed to reach this goal, such as limiting the tilt range, decreasing the number of projections, or decreasing the number of sampled image points in each projection, for example, by scanning only a random selection of image points.[4]

To compensate for the loss of information involved by any of these strategies, an enhancement of the reconstruction step of the tomographic experiment is required. Methods based on compressed sensing (CS), which exploit the minimization of the total variation ($l_1$-norm or $l_2$-norm) of the intensity gradient of the whole set of images in the tilt series, have proven to be quite efficient at this respect.[5] Therefore, though the more conventional reconstruction approaches such as SIRT

J. M. Muñoz-Ocaña, Prof. A. M. Rodríguez-Chía
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias
Universidad de Cádiz
Campus Puerto Real, Puerto Real, Cádiz 11510, Spain
A. Bouziane, Dr. A. B. Hungría, Prof. J. J. Calvino, Dr. M. López-Haro
Departamento de Ciencia de los Materiales e Ingeniería Metalúrgica y Química Inorgánica
Facultad de Ciencias
Universidad de Cádiz
Campus Puerto Real, Puerto Real, Cádiz 11510, Spain
E-mail: miguel.lopezharo@uca.es
Dr. F. Sakina, Dr. R. T. Baker
EaStChem
School of Chemistry
University of St Andrews
Fife, St Andrews KY16 9ST, UK
The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/ppsc.202000070.

DOI: 10.1002/ppsc.202000070


© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

# Nuevo proyecto concedido

## PID2020-114594GB-C21

## OPTIMIZATION ON DATA SCIENCE AND NETWORK DESIGN PROBLEMS: LARGE SCALE NETWORK MODELS MEET OPTIMIZATION AND DATA SCIENCE TOOLS

**FECHA DE INICIO**

**1 DE SEPTIEMBRE DE 2021**

# RESUMEN:

**Subproyecto US: 60**

**Subproyecto UCA: 20**

**Ambos subproyectos: 6**

Muchas gracias

| Cuartiles | |
|---|---|
| Q1 | 46 |
| Q2 | 20 |
| Q3 | 8 |
| Q4 | 6 |

https://grupo.us.es/fqm331/