

Bridging k -sum and CVaR optimization in MILP

Carlo Filippi^{a,*}, Włodzimierz Ogryczak^b, M. Grazia Speranza^a

^a Department of Economics and Management, University of Brescia, Contrada S. Chiara 50, Brescia 25122, Italy

^b Warsaw University of Technology, Institute of Control and Computation Engineering, Warsaw, Poland

ARTICLE INFO

Article history:

Received 17 November 2017

Revised 20 December 2018

Accepted 22 January 2019

Available online 30 January 2019

Keywords:

Integer programming

k -sum optimization

Minimax/Maximin

Inequality

Fairness

Equitability

Conditional Value-at-Risk

ABSTRACT

Mixed-Integer Linear Programming models often optimize the sum, or average, of different outcomes. In a deterministic setting, each outcome may be associated with an agent, for example a customer, an employee, or a time period. In a stochastic setting, each outcome may be associated with a discrete scenario. The average approach optimizes the overall efficiency of the solution, but neglects the possible unfair distribution of outcomes among agents or the risk of very bad scenarios. In this paper, we exploit the analogies of the two settings to derive a common optimization paradigm bridging the gap between k -sum optimization in the deterministic setting and Conditional Value-at-Risk optimization in the stochastic setting. We show that the proposed paradigm satisfies properties that make it an attractive criterion. To illustrate the proposed paradigm, we apply it to the multidimensional knapsack problem and the p -median/ p -center problem.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Mixed-Integer Linear Programming (MILP) is one of the most relevant tools in Operations Research, and is applied to a huge spectrum of situations. In many cases, the objective is the sum, or average, of outcomes (costs or profits), each associated with a distinct agent, for example a customer, an employee, or a time period.

The optimization of the average outcome is aimed at improving efficiency, i.e., the overall system performance. However, a most efficient solution may imply large variability on the individual outcomes. In a MILP model aimed at capturing uncertainty, discretized through a set of scenarios, an agent may be a scenario and the average over all scenarios as optimization criterion is usually adopted. An optimal solution does not take into account the variability of the outcome.

A classical remedy to avoid the worst outcomes and control the variability on individual outcomes is using the Minimax (or Maximin) criterion (Wald, 1950). This criterion looks for solutions where the worst outcome is at least as good as the worst outcome of any other feasible solution. The Minimax criterion is consistent with the Rawlsian theory of justice (Rawls, 1971) and is one of the most important models in robust decision making in general. How-

ever, the robustness of the decisions taken according to this principle comes with a price in terms of efficiency, and this price may be high (see Bertsimas and Sim, 2004).

Attempts have been made to trade-off between the search for efficiency of the average criterion and the search for control on the extreme outcomes of the Minimax criterion. The first one is to consider a convex combination of the average and the Minimax objective. This approach was proposed and analyzed in the context of location theory (Halpern, 1976; 1978), and treats the average and the Minimax as two conflicting objectives. However, it is well-known that in bi-objective MILP models there may be nondominated solutions that cannot be captured in this way (see Ehrgott, 2006).

A different approach is based upon the observation that both the average and the Minimax criteria are extreme cases of a more general concept, called k -sum optimization (Gupta and Punnen, 1990). In k -sum optimization, the adopted criterion is the average of the k worst outcomes, where k is a positive integer chosen by the decision maker. The Minimax criterion is in fact a 1-sum criterion, whereas, if S is the total number of agents generating outcomes, the average criterion is an S -sum criterion. A natural formulation of the k -sum criterion for a general MILP model involves an exponential number of constraints (Gupta and Punnen, 1990). This is probably the reason why k -sum optimization has been applied to specific combinatorial problems where MILP models are not used. The criterion of the k largest distances minimization was introduced in location problems in Slater (1978) as the so called k -centrum model. If $k = 1$, the model reduces to the standard

* Corresponding author.

E-mail addresses: carlo.filippi@unibs.it (C. Filippi), ogryczak@ia.pw.edu.pl (W. Ogryczak), grazia.speranza@unibs.it (M.G. Speranza).

p -center model while with $k=S$ it is equivalent to the classical p -median model. Early works on the concept (Andreatta and Mason, 1985a; 1985b; Slater, 1978) were focused on the case of the discrete single facility location on tree graphs. Later, Tamir (2001) has presented polynomial time algorithms for solving the multiple facility k -centrum on path and tree graphs, while Ogryczak and Tamir (2003) have shown that the criterion can be modeled with an auxiliary set of simple linear inequalities, thus simplifying several k -centrum models. Later, Puerto and Tamir (2005) have analyzed the problem of locating a tree-shaped facility using a k -centrum criterion on a tree network. Computational aspects of related problems have been studied in Grygiel (1981), Punnen (1992), and Punnen and Aneja (1996). In machine scheduling, the problem of minimizing the sum of the k , among S , maximum tardiness jobs on a single machine without preemption was shown to be polynomially solvable for any fixed value k (Woeginger, 1991). The partial sum criterion was also considered for Steiner trees in graphs and shortest paths (Duin and Volgenant, 1997; Garfinkel et al., 2006), for fair resource allocation in networks Ogryczak and Śliwiński (2002), as well as for matroids and minimum spanning trees (Gupta and Punnen, 1990). More recently, Puerto et al. (2017) have provided a general tool for solving a variety of k -sum optimization problems and improving the complexity bounds of several ad-hoc algorithms previously reported in the literature for specific combinatorial problems (see Puerto et al., 2017 for additional references). Finally, k -sum optimization can also be seen as a particular case of *ordered median* optimization, a concept that has been especially applied in location analysis (see Nickel and Puerto, 2005; Puerto and Rodríguez-Chía, 2015 for further details).

In this paper, we exploit the potential of the k -sum criterion for general MILP models applied to a deterministic setting where the average outcome over the k worst agents is optimized or a stochastic setting, discretized through scenarios, where the average outcome is computed over the k worst scenarios. We show that the k -sum criterion is strictly related to the Conditional Value-at-Risk (CVaR) criterion, a very popular measure in risk management, applied to a discrete distribution (Rockafellar and Uryasev, 2000).

The contributions of this paper are the following:

- we extend the k -sum criterion to a general MILP model with different weights on the outcomes, and for this reason we call it β -average;
- using only concepts of LP theory, we derive a compact linear formulation for the corresponding optimization problem;
- we establish a precise connection between the β -average and the CVaR, and we prove that, as a risk measure, the β -average is coherent according to Artzner et al. (1999);
- we show that the β -average is also a suitable inequality measure.

A few authors have investigated the relation between optimization of a coherent risk measure and robust optimization. Bertsimas and Sim (2004) consider a linear programming model where the constraint matrix coefficients are uncertain. They propose a robust approach where it is assumed that, for each constraint i , at most Γ_i coefficients will vary with respect to the expected values. The values of Γ_i control the conservativeness of the approach, and the resulting robust formulation retains linearity. The connection between Bertsimas and Sim's approach, where only a fraction of the uncertain coefficients may assume a bad behaviour, and optimization of risk measures where only a fraction of the worst outcomes is taken into account, is recognized by Bertsimas and Brown (2009), who show that the optimization of a coherent risk measure is equivalent to robust optimization under an appropriately defined uncertainty set. More specifically, Bertsimas and Brown (2009) suggest that a coherent risk mea-

sure can guide the choice of the uncertainty set in robust optimization. A similar result is given in Ogryczak (2014), where for a general class of optimization problems including MILP it is shown that CVaR optimization is equivalent to a properly defined robust optimization approach. Though these results establish a relation between two seemingly different fields, there is no evidence that a robust optimization approach is computationally more efficient than the direct optimization of CVaR.

The paper is structured as follows. In Section 2, we introduce the general MILP setting we study and the notation. The β -average measure is presented in Section 3, and a compact MILP formulation is derived for both the minimization and the maximization case. In Section 4, we prove relevant properties of the β -average and its strong relation to the CVaR measure. In Section 5, to illustrate the proposed model, we apply the β -average optimization to classical combinatorial optimization problems. Finally, some conclusions are drawn in Section 6.

2. Average and Minimax in MILP

We consider a general MILP model in minimization form:

$$\begin{aligned} \min \quad & c^\top x \\ \text{subject to} \quad & Ax = b \\ & x_B \in \mathbb{Z}_+^{|B|} \\ & x_N \geq 0, \end{aligned} \quad (1)$$

where c is a real n -vector, A is a real $m \times n$ matrix, b is a real m -vector, x is the n -vector of variables. The index set $\{1, 2, \dots, n\}$ is partitioned in two subsets B and N , where B contains the indices of the integer variables and N contains the indices of the continuous variables. If B is empty, the model is a plain LP model. Notice that we consider linear constraints in standard form, but any other form would work as well. For the sake of simplicity, we assume that the feasible set

$$\mathcal{X} = \{x \in \mathbb{R}^n : Ax = b, x_B \in \mathbb{Z}_+^{|B|}, x_N \geq 0\} \quad (2)$$

is nonempty and bounded.

We assume that c is the average over a finite number S of real cost vectors c^ℓ of appropriate dimension, $\ell = 1, \dots, S$. A cost vector is associated with an *agent*, for example a customer, a job, or a period of a discretized time horizon. We assume that all cost vectors have the same importance, i.e., they have the same (unitary) weight. This assumption is often satisfied in practical applications. Then,

$$c = \frac{1}{S} \sum_{\ell=1}^S c^\ell. \quad (3)$$

For any $x \in \mathcal{X}$, we define the ℓ th *outcome* of x as

$$y_\ell(x) = (c^\ell)^\top x,$$

with $\ell = 1, \dots, S$. Since outcomes represent costs, lower outcomes are preferred to higher ones. In order to rank the feasible solutions, the outcomes can be assessed according to different criteria. The most common criterion is the average (or total) outcome, taken over all the agents. We denote the average outcome of a solution x as

$$M(x) = \frac{1}{S} \sum_{\ell=1}^S y_\ell(x) = \frac{1}{S} \sum_{\ell=1}^S (c^\ell)^\top x = c^\top x.$$

We call the model where $M(x)$ is minimized the *Minavg model*. Here, we search for a solution that has the best possible average outcome. This corresponds to maximizing the system overall efficiency. An optimal solution x^{avg} of the Minavg model is such that

$$M(x^{\text{avg}}) = \min\{M(x) : x \in \mathcal{X}\}.$$

As \mathcal{X} is nonempty and bounded, the above expression is well defined.

If the Minimax criterion is adopted, then the objective function becomes

$$\mu(x) = \max\{(c^\ell)^\top x : \ell = 1, \dots, S\} = \max\{y_\ell(x) : \ell = 1, \dots, S\}.$$

The corresponding minimization model, known as *Minimax model*, selects a solution x^{\max} such that

$$\mu(x^{\max}) = \min\{\mu(x) : x \in \mathcal{X}\}.$$

The Minimax model is only focused on the worst outcome and, thus, takes into account only a small part of the available information. On the other hand, the Minavg model uses all the available information but it does not consider variability in the outcomes and the existence of very poor possible outcomes. The trade-off between the Minavg and the Minimax models could be shown in several different situations. We focus here on a simple assignment problem.

Example 1. To give an example without entering into excessive technical details, we consider a simple case. Consider the problem of assigning n tasks to S operators. Assume the operators are not identical, and let $c_{i\ell}$ be the time required by operator ℓ to complete task i . The goal is to minimize the total (average) working time of operators. The feasible set of this problem is described as

$$\mathcal{X} = \left\{ x \in \{0, 1\}^{n \times S} : \sum_{\ell=1}^S x_{i\ell} = 1, i = 1, \dots, n \right\}. \quad (4)$$

Model (1) turns out to be the following simple semi-assignment problem:

$$\text{ASSIGN: } \min \left\{ \sum_{i=1}^n \sum_{\ell=1}^S c_{i\ell} x_{i\ell} : x \in \mathcal{X} \right\}.$$

Notice that the constraint coefficient matrix in \mathcal{X} is totally unimodular (TUM). The integrality requirements can then be relaxed and model (1) is equivalent to a LP model. The ASSIGN model is a Minavg model, where the agents are the operators.

In the ASSIGN objective, the contribution of every operator is taken into account, but the optimal solution may be such that some of the operators work much more than others. Thus, the solution may be unfair and unsatisfactory. To take this concern into account, a makespan objective may be adopted, where the working time of the most loaded operator is minimized. This implies using a Minimax model, that can be formulated as

$$\text{MINMAXASSIGN: } \min \left\{ u : u \geq \sum_{i=1}^n c_{i\ell} x_{i\ell}, \ell = 1, \dots, S; x \in \mathcal{X} \right\}.$$

Notice that in MINMAXASSIGN the integrality requirements cannot be relaxed. The problem corresponds to the basic unrelated parallel machine scheduling problem $R||C_{\max}$ (Pinedo, 2012).

Whereas the ASSIGN model pursues the efficiency objective, the MINMAXASSIGN model pursues the objective of fairness among operators. As a drawback, every operator which is not the most loaded does not contribute to the objective function, and thus the workload of most of the agents is neglected.

The discussion about Minavg and Minimax models can be extended to problems where the coefficients of the objective function are uncertain and uncertainty is modeled through a finite number of scenarios with equal probability (see, e.g., Kaut and Wallace, 2007). We identify scenarios with agents. In this way, model (1) minimizes the average system performance with respect to the scenarios. The outcome $y_\ell(x)$ measures the performance of solution x under the ℓ th scenario. In the Minavg model, an optimal solution

x^{avg} is such that the average outcome, taken over all the scenarios, is optimized. In the Minimax model, an optimal solution x^{\max} is such that the outcome in the worst scenario is optimized. Again, the trade-off between the Minavg and the Minimax models could be shown in several different situations. We focus here on a classic problem that we use also in the computational experiments.

Example 2. Let us consider the multidimensional knapsack problem (MKP) (Kellerer et al., 2004). A set of n activities is given. A profit c_j is associated with each activity j , $j = 1, \dots, n$. A set of m resources is used by the activities, each available in a maximum amount b_i , $i = 1, \dots, m$. Each activity j requires a quantity a_{ij} of resource i . The problem is to select a subset of the activities, with maximum total profit, while guaranteeing that the constraints on the maximum availability of the resources are satisfied. The problem is formulated as a binary linear program, with a binary variable x_j associated with each activity j . Its feasible set may be expressed as follows

$$\mathcal{X} = \left\{ x \in \{0, 1\}^m : \sum_{j=1}^n a_{ij} x_j \leq b_i, i = 1, \dots, m \right\}, \quad (5)$$

where all parameters a_{ij} and b_i are positive integers. The MKP is then formulated as (Kellerer et al., 2004)

$$\text{MKP: } \max \left\{ \sum_{j=1}^n c_j x_j : x \in \mathcal{X} \right\},$$

where all coefficients c_j are integers. The classical knapsack problem is the MKP with one constraint only.

Assume that activities correspond to investments and the profits c_j are the returns of these investments. It is often the case that such returns are not known with precision. Combining market forecasts, historical data, opinions of experts, one may build up a set of S equiprobable scenarios describing the uncertainty. In other words, we assume that S vectors $c^\ell = (c_{\ell 1}, c_{\ell 2}, \dots, c_{\ell n})$ are given, for $\ell = 1, \dots, S$, and that a probability $\pi_\ell = 1/S$ is assigned to each vector. For consistency, we define $c_j = \sum_{\ell=1}^S \pi_\ell c_{\ell j}$ for $j = 1, \dots, n$. In this case, the MKP model maximizes the average total profit on all scenarios. Hence, it is a Minavg model with reversed sign in the objective function, i.e., a *Maxavg* model.

Activities with average high profit may have a very variable profit (very low in some scenarios and very high in others). Conversely, activities with average low profit may have quite stable profit. For a risk-averse decision maker, the goal is to control the worst possible scenarios. A minimax approach leads to maximize the minimum profit among the scenarios and, considering the reversed sign of the profits, to a *Maximin* model:

$$\text{MKP-MAXMIN: } \max \left\{ u : u \leq \sum_{j=1}^n c_{\ell j} x_j, \ell = 1, \dots, S; x \in \mathcal{X} \right\}.$$

In summary, in the uncertainty setting the Minavg model implies risk indifference, whereas the Minimax model implies as high as possible risk aversion. In the variability setting, the Minavg model implies that the only goal is system efficiency and there is no concern about fairness, whereas opposite concerns are taken into account in the Minimax model.

3. The Minimax(β) model

Let \mathcal{X} be defined as in (2), let $x \in \mathcal{X}$, and let $y(x) = (y_1(x), y_2(x), \dots, y_S(x))^\top$ be the corresponding vector of outcomes. Notice that $M(x)$ is the average outcome. Moreover, $\mu(x)$ is the largest outcome associated with an agent, i.e., the largest

average outcome associated with a subset of agents with cardinality 1. These observations suggest the following definition.

Definition 1. For any fraction β , with $\beta \in (0, 1]$, and for any $x \in \mathcal{X}$, the β -average $M_\beta(x)$ of the outcome vector $y(x)$ is the largest average outcome attainable by a subset of $\lceil \beta S \rceil$ agents.

The β -average optimization problem, that we also call $\text{Minimax}(\beta)$ model, aims at finding a solution x^* such that

$$M_\beta(x^*) = \min\{M_\beta(x) : x \in \mathcal{X}\}. \tag{6}$$

When $\beta = 1$ the $\text{Minimax}(\beta)$ model becomes the Minavg model, whereas for $0 < \beta \leq 1/S$ the model is the Minimax model. When $\beta = k/S$, with $k \in \{1, 2, \dots, S\}$, then the β -average is the largest average outcome attainable by a subset of k agents. In a deterministic setting, the $\text{Minimax}(k/S)$ model corresponds to k -sum optimization. In a stochastic setting, the $\text{Minimax}(\beta)$ model is strictly related to CVaR optimization on discrete distributions, as we shall see later.

Let

$$\tau(1), \tau(2), \dots, \tau(S)$$

be a valid permutation of the outcome indices, that is a permutation such that

$$y_{\tau(1)}(x) \geq y_{\tau(2)}(x) \geq \dots \geq y_{\tau(S)}(x). \tag{7}$$

The permutation τ actually depends on $x \in \mathcal{X}$, but, to ease notation, we take such a dependence for granted. Let $y^\tau(x)$ be the ordered vector of outcomes. Whereas $y_j(x)$ indicates the j th component of the original vector $y(x)$, $y_{\tau(j)}(x)$ indicates the j th component of the ordered vector $y^\tau(x)$.

We can now express the β -average, with $\beta \in (0, 1]$, in formula as

$$M_\beta(x) = \frac{1}{\lceil \beta S \rceil} \sum_{j=1}^{\lceil \beta S \rceil} y_{\tau(j)}(x). \tag{8}$$

Example 3. Suppose that $S = 10$ and that, for a given solution $x \in \mathcal{X}$, we have:

$$y(x) = (12, 3, 1, 7, 18, 9, 4, 12, 15, 13)^\top.$$

Then, we consider the valid permutation $\tau: 5, 9, 10, 1, 8, 6, 4, 7, 2, 3$, which orders the vector of outcomes. The corresponding ordered vector is

$$y^\tau(x) = (18, 15, 13, 12, 12, 9, 7, 4, 3, 1)^\top.$$

We note that an alternative valid permutation $\tau': 5, 9, 10, 8, 1, 6, 4, 7, 2, 3$, where indices 1 and 8 are swapped would result in the same ordered vector.

Using formula (8), we can compute $M_\beta(x)$ for different values of β . If $\beta = 2/10$, then $M_{0.2}(x) = (y_5 + y_9)/2 = 16.5$. If $\beta = 4/10$, then $M_{0.4}(x) = (y_5 + y_9 + y_{10} + y_1)/4 = 14.5$. If β is not a multiple of $1/S$, we round the value up to the closest multiple. For example, if $\beta = 0.35$, which is included between $3/10$ and $4/10$, then $\lceil \beta S \rceil = 4$ and we obtain $M_{0.35}(x) = M_{0.4}(x) = 14.5$.

The β -average is not affected by the specific valid permutation of outcomes. Indeed, directly from Definition 1 and expression (8), the following statement follows.

Proposition 1. For any given $x \in \mathcal{X}$, and for any $\beta \in (0, 1]$, the β -average is a symmetric function of $y_1(x), y_2(x), \dots, y_S(x)$, i.e., given two valid permutations τ' and τ'' , the value $M_\beta(x)$ obtained with permutation τ' is identical to the value obtained with permutation τ'' .

3.1. Optimizing the β -average

Solving the $\text{Minimax}(\beta)$ model defined in (6) with $M_\beta(x)$ formulated as in (8) is not practical, since it relies on a permutation of the outcomes, that is, a function of $x \in \mathcal{X}$. In this section, we show how to model the β -average optimization problem with an efficient LP expansion of the original MILP model, i.e., by adding a limited number of linear constraints and continuous variables.

Proposition 2. For any β , with $\beta \in (0, 1]$, the β -average is given by

$$M_\beta(x) = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, z_\ell \in \{0, 1\}, \ell = 1, \dots, S \right\}. \tag{9}$$

The above proposition is straightforward. Indeed, the binary variables describe the incidence vectors of the outcome subsets. The unique constraint imposes that only subsets of cardinality $\lceil \beta S \rceil$ are chosen. The objective function maximizes the average outcome of the selected subset.

Proposition 2 allows a practical formulation for the $\text{Minimax}(\beta)$ model. As the matrix of the coefficients of the constraint in (9) is totally unimodular, we can relax the integrality constraints, obtaining the equivalent linear program

$$M_\beta(x) = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\}. \tag{10}$$

Problem (10) is feasible and bounded for any $\beta \in (0, 1]$ and $x \in \mathcal{X}$. Thus, we can switch to the dual, obtaining

$$M_\beta(x) = \min \left\{ \lceil \beta S \rceil u + \sum_{\ell=1}^S v_\ell : \lceil \beta S \rceil (u + v_\ell) \geq y_\ell(x), v_\ell \geq 0, \ell = 1, \dots, S \right\}.$$

Replacing the outcomes with their explicit expression in terms of $x \in \mathcal{X}$, we are able to express the $\text{Minimax}(\beta)$ model for the minimization problem (1) as

$$\begin{aligned} \min \quad & \lceil \beta S \rceil u + \sum_{\ell=1}^S v_\ell \\ \text{subject to} \quad & \lceil \beta S \rceil (u + v_\ell) \geq (c^\ell)^\top x \quad \ell = 1, \dots, S \\ & Ax = b \\ & v_\ell \geq 0 \quad \ell = 1, \dots, S \\ & x_B \in \mathbb{Z}_+^{|B|}, x_N \geq 0. \end{aligned} \tag{11}$$

We remark that (11) is still a MILP model. With respect to the original (1), it has $S + 1$ additional continuous variables and S additional inequality constraints. Moreover, the objective function and the new constraints have integer coefficients, provided the outcomes are integer valued.

3.2. Case of maximization

We consider here the case where the original MILP problem is formulated in maximization form as

$$\begin{aligned} \max \quad & c^\top x \\ \text{subject to} \quad & Ax = b \\ & x_B \in \mathbb{Z}_+^{|B|} \\ & x_N \geq 0. \end{aligned} \tag{12}$$

In this case, the vectors c^ℓ , $\ell = 1, \dots, S$, represent profits, and a higher outcome is preferred to a lower one. Then, the β -average is the smallest average outcome attainable by a subset of $\lceil \beta S \rceil$ agents. The simplest way to derive the $\text{Maximin}(\beta)$ optimization model is to transform problem (12) in minimization form and then use the $\text{Minimax}(\beta)$ optimization model (11). With some algebraic

calculations, considering that u is a free variable, we obtain the following Maximin(β) optimization problem:

$$\begin{aligned} \max \quad & [\beta S]u - \sum_{\ell=1}^S v_\ell \\ \text{subject to} \quad & [\beta S](u - v_\ell) \leq (c^\ell)^\top x \quad \ell = 1, \dots, S \\ & Ax = b \\ & v_\ell \geq 0 \quad \ell = 1, \dots, S \\ & x_B \in \mathbb{Z}_+^{|B|}, x_N \geq 0. \end{aligned} \tag{13}$$

4. Properties of the β -average

In this section, we describe the relation between the β -average criterion and the well-known CVaR risk measure (Rockafellar and Uryasev, 2000). Then, we show that, in a stochastic setting, the β -average is a coherent risk measure and that, in a deterministic setting, the β -average is an appropriate measure of inequality.

4.1. Relation between β -average and CVaR

Several approaches have been proposed in finance to shape a loss distribution associated with a portfolio of assets. The Value-at-Risk (VaR) and the CVaR have become very popular concepts. Given a specified probability level α , the α -VaR of a loss distribution is the lowest amount v such that, with probability α , the loss will not exceed v . For continuous distributions, the α -CVaR is the conditional expectation of losses above that amount v (see Rockafellar and Uryasev, 2000). For distributions with discontinuities, e.g., discrete distributions defined by scenarios, the definition of α -CVaR is more subtle and requires some notation (Rockafellar and Uryasev, 2002).

Let $F_Y(\xi)$ be the cumulative distribution function of the loss Y . For any $\alpha \in (0, 1)$, we have α -VaR(Y) = $\min\{\xi : F_Y(\xi) \geq \alpha\}$. Moreover, let

$$F_Y^\alpha(\xi) = \begin{cases} 0, & \text{if } \xi < \alpha\text{-VaR}(Y), \\ \frac{F_Y(\xi) - \alpha}{1 - \alpha}, & \text{if } \xi \geq \alpha\text{-VaR}(Y). \end{cases}$$

Then, we have

$$\alpha\text{-CVaR}(Y) = \int_{-\infty}^{+\infty} \xi dF_Y^\alpha(\xi).$$

We can also define the “lower” CVaR (Sarykalin et al., 2008) as $\alpha\text{-CVaR}^-(Y) = \mathbb{E}[Y|Y \geq \alpha\text{-VaR}(Y)]$,

where \mathbb{E} is the expectation operator. For general distributions,

$$\alpha\text{-CVaR}^-(Y) \leq \alpha\text{-CVaR}(Y),$$

where the inequality holds strictly on every discontinuity of $F_Y(\xi)$ (see Rockafellar and Uryasev, 2002 for details). While the VaR has some undesirable features, the CVaR has all the desirable characteristics needed to be defined a coherent risk measure (see Artzner et al., 1999).

To highlight the relation between the β -average and the CVaR, we interpret any vector $x \in \mathcal{X}$ as a feasible portfolio, and the associated distribution of outcomes $y_1(x), y_2(x), \dots, y_S(x)$ as a probabilistic distribution of losses, each with probability $1/S$. We need to define some functions, starting from the right-continuous function:

$$F_{y(x)}(\xi) = \frac{1}{S} \sum_{\ell=1}^S \delta_\ell(\xi), \quad \text{where } \delta_\ell(\xi) = \begin{cases} 1 & \text{if } y_\ell(x) \leq \xi \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

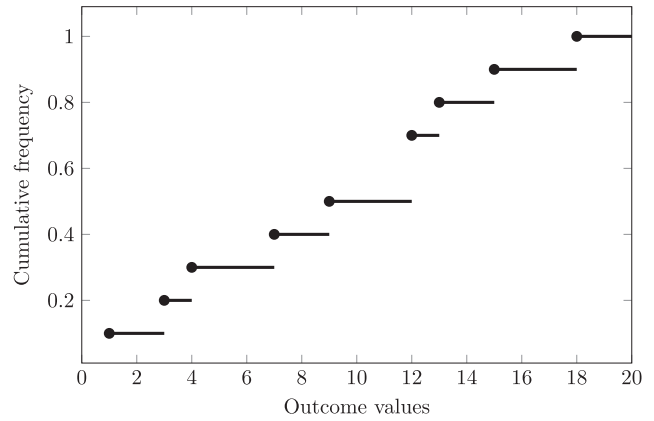


Fig. 1. Plot of function $F_{y(x)}(\xi)$ for $y(x)$ as in Example 3.

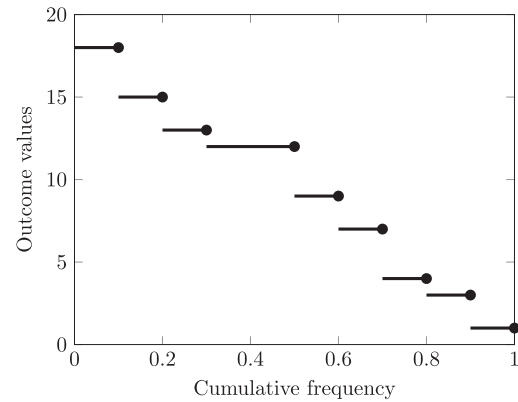


Fig. 2. Plot of function $\bar{F}_{y(x)}^{(-1)}(\xi)$ for $y(x)$ as in Example 3.

which, for any real value ξ , provides the fraction of outcomes smaller than or equal to ξ . Note that function $F_{y(x)}(\xi)$ is, in fact, the cumulative distribution function of a discrete random variable that takes the values $y_1(x), y_2(x), \dots, y_S(x)$, each with the same probability $1/S$. In Fig. 1, we show function $F_{y(x)}(\xi)$ for Example 3.

Similarly, the left-continuous right tail function

$$\bar{F}_{y(x)}(\xi) = \frac{1}{S} \sum_{\ell=1}^S \bar{\delta}_\ell(\xi), \quad \text{where } \bar{\delta}_\ell(\xi) = \begin{cases} 1 & \text{if } y_\ell(x) \geq \xi \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

can be defined which, for any real value ξ , provides the fraction of outcomes greater than or equal to ξ . Note that $\bar{F}_{y(x)}(\xi) = 1 - F_{y(x)}(\xi)$ for $\xi \notin \{y_1(x), y_2(x), \dots, y_S(x)\}$.

Next, we introduce the quantile function $F_{y(x)}^{(-1)}$ as the left-continuous inverse of $F_{y(x)}(x)$, i.e., $F_{y(x)}^{(-1)}(\alpha) = \inf\{\xi : F_{y(x)}(\xi) \geq \alpha\}$ for $\alpha \in (0, 1]$. The quantile function $F_{y(x)}^{(-1)}$ is the nondecreasing stepwise function $F_{y(x)}^{(-1)}(\alpha) = y_{\tau(S-j+1)}(x)$ for $(j-1)/S < \alpha \leq j/S$. Similarly, we introduce the right tail quantile function $\bar{F}_{y(x)}^{(-1)}$ as the left-continuous inverse of $\bar{F}_{y(x)}$, i.e., $\bar{F}_{y(x)}^{(-1)}(\alpha) = \sup\{\xi : \bar{F}_{y(x)}(\xi) \geq \alpha\}$ for $\alpha \in (0, 1]$. Actually, $\bar{F}_{y(x)}^{(-1)}(\alpha) = F_{y(x)}^{(-1)}(1 - \alpha)$. Note that the function $\bar{F}_{y(x)}^{(-1)}$ is the nonincreasing stepwise function $\bar{F}_{y(x)}^{(-1)}(\alpha) = y_{\tau(j)}(x)$ for $(j-1)/S < \alpha \leq j/S$. In Fig. 2, we show function $\bar{F}_{y(x)}^{(-1)}(\xi)$ for Example 3.

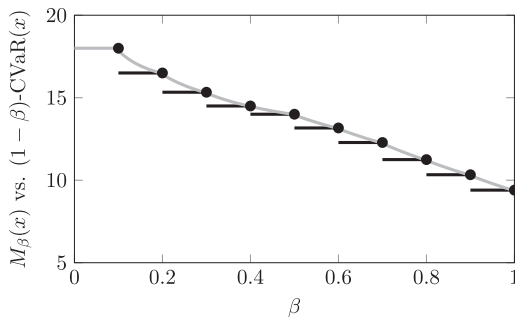


Fig. 3. Plot of $M_\beta(x)$ (black line) vs. $(1 - \beta)$ -CVaR(x) (gray line) for $y(x)$ as in Example 3.

Recalling expression (8), we obtain

$$M_\beta(x) = \frac{1}{\lceil \beta S \rceil} \sum_{j=1}^{\lceil \beta S \rceil} y_{\tau(j)}(x) = \frac{1}{\lceil \beta S \rceil} \sum_{j=1}^{\lceil \beta S \rceil} \bar{F}_{y(x)}^{(-1)}\left(\frac{j}{S}\right) \\ = \frac{S}{\lceil \beta S \rceil} \int_0^{\frac{\lceil \beta S \rceil}{S}} \bar{F}_{y(x)}^{(-1)}(v) dv. \tag{16}$$

Given a specified probability level β , we denote as $(1 - \beta)$ -CVaR(x) the $(1 - \beta)$ -CVaR value of the distribution $y_1(x), y_2(x), \dots, y_S(x)$. This right tail mean can be expressed as (see Rockafellar and Uryasev, 2000):

$$(1 - \beta)\text{-CVaR}(x) = \frac{1}{\beta} \int_0^\beta \bar{F}_{y(x)}^{(-1)}(v) dv. \tag{17}$$

Since $S/\lceil \beta S \rceil \leq 1/\beta$ and $(1 - \beta)$ -CVaR(x) is a nonincreasing function of β , (16) and (17) imply the following.

Proposition 3. For all $\beta \in (0, 1]$, we have $M_\beta(x) \leq (1 - \beta)$ -CVaR(x). In particular,

$$M_\beta(x) = \left(1 - \frac{\lceil \beta S \rceil}{S}\right) \text{CVaR}(x),$$

that is, the β -average coincides with the $\left(1 - \frac{\lceil \beta S \rceil}{S}\right)$ -CVaR.

In Fig. 3, we plot functions $M_\beta(x)$ and $(1 - \beta)$ -CVaR(x) with respect to parameter β for the outcome vector of Example 3.

The above result implies, in case of a uniform discrete distribution,

$$M_\beta(x) = (1 - \beta) \text{CVaR}^-(x). \tag{18}$$

4.2. The β -average is a coherent risk measure

It is well known that the CVaR is a coherent risk measure. According to Rockafellar and Uryasev (2002), this is a “formidable advantage not shared by any other widely applicable measure of risk yet proposed.” We have seen above that, in case of a uniform discrete distribution, the β -average is equivalent to the “lower CVaR” $(1 - \beta) \text{CVaR}^-(x)$. An interesting point is that, for general discrete distributions, CVaR^- is not a coherent risk measure (Rockafellar and Uryasev, 2002). In this section, we prove that the β -average is a coherent measure. As a corollary, we have that CVaR^- is coherent on uniform discrete distributions.

We first note that the β -average of $x \in \mathcal{X}$ is actually a composite function. To stress this point, we write $M_\beta(x) = \mu_\beta(y(x))$. By Proposition 2 and the total unimodularity of the constraints in

problem (9), we have:

$$\mu_\beta(y(x)) = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\}. \tag{19}$$

Thus, the β -average is described by the optimal value of a feasible and bounded LP in maximization form, where the outcome vector is the objective coefficient vector. In the following, we use $y(x) \leq y(x')$ to denote the existence of two valid permutations τ and τ' such that $y_{\tau(\ell)}(x) \leq y_{\tau'(\ell)}(x')$ for $\ell = 1, \dots, S$. Let $\mathbf{1}$ denote a vector of S ones.

Theorem 1. For any $x, x' \in \mathcal{X}$ and $\beta \in (0, 1]$, function μ_β satisfies the following properties:

- i. $\mu_\beta(y(x) + c \cdot \mathbf{1}) = \mu_\beta(y(x)) + c$ for any constant c (translation-invariance);
- ii. $\mu_\beta(y(x) + y(x')) \leq \mu_\beta(y(x)) + \mu_\beta(y(x'))$ (subadditivity);
- iii. $\mu_\beta(cy(x)) = c\mu_\beta(y(x))$ for all $c > 0$ (positive homogeneity);
- iv. $y(x) \leq y(x') \Rightarrow \mu_\beta(y(x)) \leq \mu_\beta(y(x'))$ (monotonicity);

Proof. All properties essentially follow from characterization (19).

Property i. We have:

$$\mu_\beta(y(x) + c \cdot \mathbf{1}) \\ = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S (y_\ell(x) + c) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} \\ = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell + c : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} \\ = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} + c \\ = \mu_\beta(y(x)) + c.$$

Property ii. We have:

$$\mu_\beta(y(x) + y(x')) \\ = \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S (y_\ell(x) + y_\ell(x')) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} \\ \leq \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x) z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} \\ + \max \left\{ \frac{1}{\lceil \beta S \rceil} \sum_{\ell=1}^S y_\ell(x') z_\ell : \sum_{\ell=1}^S z_\ell = \lceil \beta S \rceil, 0 \leq z_\ell \leq 1, \ell = 1, \dots, S \right\} \\ = \mu_\beta(y(x)) + \mu_\beta(y(x')).$$

Property iii. It follows immediately from characterization (19).

Property iv. By the symmetry of the feasible region in problem (19), if $y_{\tau(\ell)}(x) \leq y_{\tau'(\ell)}(x')$ for $\ell = 1, \dots, S$, then we have:

$$\mu_\beta(y(x)) = \mu_\beta(y^\tau(x)) \leq \mu_\beta(y^{\tau'}(x')) = \mu_\beta(y(x')).$$

□

In the framework of Artzner et al. (1999), a risk measure is a functional on a linear space of random variables expressing losses. In our framework, for any $x \in \mathcal{X}$, the role of the random variable is taken by the outcome vector $y(x)$, equivalent to a discrete random variable $Y(x)$ taking value $y_\ell(x)$ with probability $1/S$, for $\ell = 1, \dots, S$. Then, the relation $y(x) \leq y(x')$ is equivalent to first-order stochastic dominance of $Y(x)$ over $Y(x')$. The role of the risk measure is taken by the β -average, in the form $\mu_\beta(y(x))$. It turns out that the above properties (i)–(iv) are the four axioms required by Artzner et al. (1999) to define coherency. This leads to the following statement.

Corollary 1. The β -average is a coherent risk measure.

As a consequence of (18), we also have the following.

Corollary 2. In case of a uniform discrete distribution, $(1 - \beta) \text{CVaR}^-$ is a coherent risk measure.

4.3. The β -average is a measure of inequality

We have introduced the β -average to trade-off between system efficiency and outcome variability. Since in a deterministic setting large variability means unequal distribution of outcomes among agents, it is pertinent to wonder whether the β -average is an appropriate measure of inequality.

Inequality measures have a central role in applied welfare economics, a branch of economics devoted to the development of analytical tools to evaluate the level of inequality in the distribution of resources and the application of these tools to actual data. It is known that Schur-convexity is the key property for an inequality measure, since from the Hardy–Littlewood–Polya’s theorem it is equivalent to monotonicity with respect to the Lorenz order or the Pigou–Dalton principle of transfers (Marshall et al., 2011). In other contexts, inequality measures are also called fairness measures (Ogryczak et al., 2014) or equitability measures (Kostreva et al., 2004).

We recall that a function $f : \mathbb{R}^S \mapsto \mathbb{R}$ is Schur-convex if, for any $y, y' \in \mathbb{R}^S$ such that there exist two valid permutations τ, τ' with $y_{\tau(\ell)} \leq y'_{\tau'(\ell)}$, we have $f(y) \leq f(y')$. This corresponds to Property iv of Theorem 1, and thus we have the following.

Corollary 3. For any $\beta \in (0, 1]$, the β -average is a Schur-convex function of the outcome vector $y(x)$.

5. Computational experiments

The aim of this section is to illustrate the concepts presented above by using two classic combinatorial optimization problems. The first one is the MKP, which gives rise to MILP models with a maximization objective and will be used as an example where the risk involved by the uncertainty in the objective function coefficients is an issue. The second one is the p -median/ p -center problem, which gives rise to MILP models with a minimization objective, and will be used as an example where the fairness of the outcome distribution is an issue. We test the different models on benchmark instances taken from the OR-library (Beasley, 1990), as detailed below. All used instances are available at <http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files>.

All tests have been performed on a 64-bit Windows machine, with Intel Xeon processor E5-1650, 3.50 GHz, and 16 GB of RAM. All models have been built and solved using IBM ILOG CPLEX Optimization Studio 12.6 (64 bit version) with default solver settings, except for the maximum running time, set to 7200 s for each model-instance pair.

5.1. Example of uncertainty: the multidimensional knapsack problem

In Example 2, we have introduced the MKP model and its Maximin counterpart MKP-MAXMIN. In this context, a MaxMin(β) model aims at maximizing the total profit earned under the $k = \lceil \beta S \rceil$ worst scenarios. This leads to the following model

MKP(β):

$$\max \left\{ ku - \sum_{\ell=1}^S v_{\ell} : ku - kv_{\ell} \leq \sum_{j=1}^n c_{\ell j} x_j, \ell = 1, \dots, S; v_{\ell} \geq 0, \ell = 1, \dots, S; x \in \mathcal{X} \right\},$$

where set \mathcal{X} is defined in (5). In our computational tests, the value of βS is integer. In this case, β -average optimization and (1β) -CVaR optimization are in fact equivalent (see Proposition 3).

We compare the solutions obtained by the three models MKP, MKP-MAXMIN, and MKP(β), on a set of medium size benchmark instances taken from the OR-library, namely the instances contained in file `mknapcb1.txt`. These are 30 randomly generated instances where $n = 100$ and $m = 5$, the profit c_j is correlated to $\sum_{i=1}^m a_{ij}$ for all $j = 1, \dots, n$, and $b_i = \alpha \sum_{j=1}^n a_{ij}$, $i = 1, \dots, m$, with

α being the *tightness ratio*. In the first 10 instances, $\alpha = 0.25$; in the next 10 instances, $\alpha = 0.50$; in the last 10 instances, $\alpha = 0.75$ (see Chu and Beasley, 1998 for details). As we will see, the hardness of the instances is inversely correlated with α .

For each instance, S scenarios are generated, for different S values. We assume that for each object j , the original profit c_j is actually the maximum possible one, and we generate independently S values by drawing uniformly distributed integers in the interval $[-0.75c_j, c_j]$. As a consequence, the expected profit of object j is $0.125 \cdot c_j$, for all $j = 1, \dots, 100$. A similar scheme has been used in Ogryczak and Śliwiński (2003) to generate rates of return in a portfolio optimization problem.

We first analyse in detail the results concerning instances #11–20, where $\alpha = 0.50$, fixing $S = 100$. The four models MKP-MAXMIN, MKP(0.10), MKP(0.30), and MKP are run on each instance. To get a fair comparison, in this experiment we identify the distribution of profits with the set of generated scenarios. Thus, for each instance, the average profit vector used in model MKP is the average over the generated scenarios, see (3). In Table 1, the solution returned by each model is evaluated with respect to the objective of the other models. More precisely, we have four rows per instance, each row corresponding to a model. Column *Min* contains the percentage differences between the minimum total profit obtained under the worst scenario and the optimal value of the MKP-MAXMIN model. Similarly, column $M_{0.10}$ (resp. $M_{0.30}$) contains the percentage differences between the average profit under the $\lceil 0.10 \times S \rceil$ (resp. $\lceil 0.30 \times S \rceil$) worst scenarios and the optimal value of the MKP(0.10) (resp. MKP(0.30)) model. Column *Avg* contains the percentage differences between the average profit under all scenarios and the optimal value of the MKP model. In other words, these four columns contain the percentage *regrets* that the solution of each model implies with respect to every considered objective. Such regrets are computed in negative form, to emphasize the fact that they correspond to a reduction with respect to the optimal value. For example, it turns out that in instance #11 the worst possible profit for the solution returned by the MKP(0.30) is 92.74% lower than the worst possible profit for the solution returned by the MKP-MAXMIN.

Finally, column *Time* reports the CPU time in rounded seconds required to obtain the solutions, and column *Gap* that contains percentage optimality gap for instances not solved to optimality within the time limit of 7200 seconds. These are all instances of the MKP-MAXMIN model, which turns out to be very hard to solve for CPLEX. On the other extreme, CPLEX is very fast in solving the classical MKP model. MKP(0.10) and MKP(0.30) instances are all solved to optimality with a CPU time ranging from 22 to 2994 s. Figures in column *Time* suggest a strong inverse correlation between the β value and the hardness of the instance for CPLEX. It is also interesting to note that for instance #11, models MKP-MAXMIN and MKP(0.10) return the same solution, but while CPLEX required 1553 s to prove optimality with respect to the former model, it requires just 35 s to prove optimality with respect to the latter model.

Column *Min* suggests a strong negative correlation between the β value and the value of the worst possible profit, though the magnitude of the regret value is clearly related to the adopted perturbation. The positive correlation between the β value and the value of the average profit is less pronounced, and there are some cases of non-monotonicity.

To give a better view of the differences among the solutions returned by the different models, in Fig. 4 we show the cumulative frequency function of the four solutions for instance #15, taken as an example. More specifically, the values on the horizontal axis are the possible profits, while the values on the vertical axis are fractions of the scenarios. For example, in the MKP-MAXMIN solution, 65% of the scenarios imply a profit not greater than 5000. In

Table 1
Multidimensional knapsack instances with uncertain profits and 100 scenarios: percentage deviations on the optimal objective values.

Instance	Model type	Min	$M_{0.10}$	$M_{0.30}$	Avg	Time	Gap
11	MKP-MAXMIN	-	0.00	-6.15	-13.26	1553	-
...	MKP(0.10)	0.00	-	-6.15	-13.26	35	-
...	MKP(0.30)	-92.74	-11.40	-	-9.69	22	-
...	MKP	-129.56	-60.40	-20.13	-	0	-
12	MKP-MAXMIN	-	-1.45	-9.98	-13.82	3184	-
...	MKP(0.10)	-16.20	-	-5.20	-7.02	141	-
...	MKP(0.30)	-103.73	-13.55	-	-9.11	27	-
...	MKP	-249.80	-77.95	-26.57	-	0	-
13	MKP-MAXMIN	-	-2.16	-7.90	-15.01	7201	4.11
...	MKP(0.10)	-23.03	-	-1.19	-10.51	249	-
...	MKP(0.30)	-64.67	-14.43	-	-10.42	69	-
...	MKP	-139.48	-65.92	-25.90	-	0	-
14	MKP-MAXMIN	-	-1.31	-6.87	-9.07	5796	-
...	MKP(0.10)	-18.23	-	-4.63	-10.78	237	-
...	MKP(0.30)	-94.52	-21.91	-	-6.71	22	-
...	MKP	-130.27	-52.82	-19.26	-	0	-
15	MKP-MAXMIN	-	-2.10	-8.22	-18.74	7201	6.07
...	MKP(0.10)	-6.34	-	-3.96	-14.50	2994	-
...	MKP(0.30)	-32.29	-14.32	-	-9.98	40	-
...	MKP	-130.93	-67.78	-32.40	-	0	-
16	MKP-MAXMIN	-	-1.36	-8.21	-12.64	7200	4.42
...	MKP(0.10)	-15.93	-	-3.97	-12.77	547	-
...	MKP(0.30)	-48.82	-12.45	-	-8.16	60	-
...	MKP	-132.76	-55.08	-20.84	-	0	-
17	MKP-MAXMIN	-	-1.98	-7.13	-14.14	7201	5.03
...	MKP(0.10)	-1.94	-	-4.49	-13.72	476	-
...	MKP(0.30)	-74.33	-14.90	-	-11.69	22	-
...	MKP	-167.20	-71.10	-24.71	-	0	-
18	MKP-MAXMIN	-	-2.00	-6.05	-12.38	7201	5.82
...	MKP(0.10)	-10.14	-	-3.85	-12.55	2140	-
...	MKP(0.30)	-81.91	-17.64	-	-8.40	225	-
...	MKP	-183.47	-68.97	-28.57	-	0	-
19	MKP-MAXMIN	-	-1.56	-6.88	-14.12	3540	-
...	MKP(0.10)	-7.00	-	-4.89	-13.47	150	-
...	MKP(0.30)	-32.34	-7.82	-	-11.34	26	-
...	MKP	-205.90	-83.99	-33.68	-	1	-
20	MKP-MAXMIN	-	-2.37	-3.86	-15.02	7201	5.90
...	MKP(0.10)	-6.95	-	-3.75	-15.73	809	-
...	MKP(0.30)	-81.51	-18.95	-	-10.95	53	-
...	MKP	-111.51	-74.29	-31.74	-	0	-

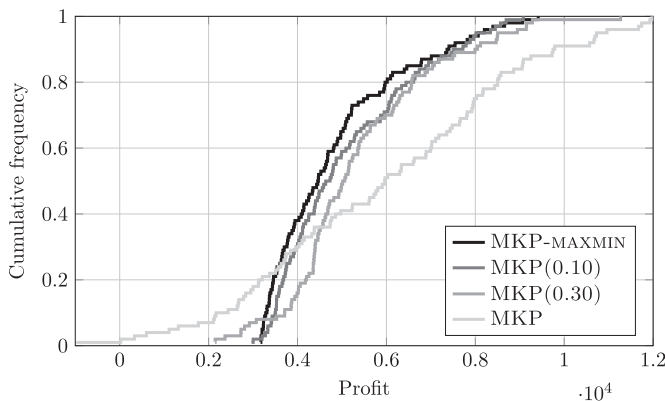


Fig. 4. Cumulative frequency plots for instance #15 in mknapcb1.

the MKP(0.10) solution, the percentage reduces to 59% and in the MKP(0.30) solution a 49% is reached. Finally, in the MKP solution, 41% of the scenarios imply a profit not exceeding 5000. To improve readability, we fill the discontinuities with vertical segments.

The plot shows a common pattern: the distributions for models MKP-MAXMIN and MKP(0.10) are quite similar, with a large number

of payoffs concentrated on the right of the worst payoff, though the solutions of the MKP(β) models tend to slightly dominate the solution of the MKP-MAXMIN, in the sense that for any value δ , the percentage of scenarios with profit not exceeding δ is smaller in the MKP(0.10) solution. The distribution corresponding to the MKP is very different, with a bold left tail that includes a fraction of bad payoffs. The distribution corresponding to the MKP(0.30) tends to mediate between these two patterns.

Having assessed the quality of the obtained solutions, we evaluate more closely the computational effort required by MKP(β). We fix $\beta = 0.30$ and consider all 30 instances in mknapcb1.txt, generating an increasing number of scenarios, from 100 to 500. The results are summarized in Table 2, where α is the tightness ratio of the instances and *t.l.* indicates that the time limit of 7200 s is reached. Notice that, along each row of Table 2, the increment of S is obtained by adding scenarios to the existing ones. We see that, at least for CPLEX, the hardness of the instances is increasing with S and decreasing with α . None of the instances with $\alpha = 0.25$ are solved to optimality for $S = 400$ and $S = 500$. Moreover, the optimality gap may be as high as 23.6%. Conversely, all instances with $\alpha = 0.75$ are solved to optimality well before the time limit.

In general, the computational effort increases with S , but the relation is not monotone. For example, in instance #12, $S = 400$

Table 2
Multidimensional knapsack instances with uncertain profits and increasing number of scenarios: times, gaps and objective values with model MKP(0,30).

N. of scenarios S =		100			200			300			400			500		
α	Instance	Time	Gap	$M_{0,30}$	Time	Gap	$M_{0,30}$	Time	Gap	$M_{0,30}$	Time	Gap	$M_{0,30}$	Time	Gap	$M_{0,30}$
0.25	1	288	–	1753.20	t.l.	3.87	–28.53	t.l.	4.54	–35.25	t.l.	15.01	–44.53	t.l.	19.08	–48.48
	2	40	–	1995.53	962	–	–34.15	1352	–	–38.84	t.l.	3.03	–44.97	t.l.	11.24	–54.12
	3	32	–	1808.23	5887	–	–29.34	t.l.	2.31	–44.17	t.l.	11.72	–49.13	t.l.	10.20	–50.58
	4	28	–	2249.93	235	–	–32.06	3331	–	–40.68	t.l.	2.44	–45.40	t.l.	14.59	–53.62
	5	65	–	1883.73	574	–	–24.33	2589	–	–34.43	t.l.	4.57	–41.93	t.l.	13.61	–49.28
	6	51	–	1778.00	545	–	–24.54	t.l.	2.71	–34.63	t.l.	9.64	–41.00	t.l.	11.30	–45.64
	7	27	–	2178.37	360	–	–19.39	2110	–	–31.80	t.l.	6.11	–41.07	t.l.	4.21	–43.18
	8	49	–	2039.97	5722	–	–37.69	t.l.	4.41	–48.30	t.l.	16.79	–59.92	t.l.	23.63	–61.76
	9	30	–	1824.97	4352	–	–30.48	6107	–	–34.11	t.l.	6.49	–39.46	t.l.	13.11	–44.76
	10	31	–	2193.83	772	–	–29.79	3115	–	–40.58	t.l.	11.05	–48.10	t.l.	10.80	–48.71
0.5	11	22	–	3705.37	65	–	–17.15	587	–	–28.49	6394	–	–31.82	t.l.	1.17	–30.45
	12	122	–	3513.10	1162	–	–18.59	2237	–	–20.51	4026	–	–24.59	1459	–	–26.18
	13	56	–	3886.27	398	–	–14.99	3169	–	–23.85	t.l.	1.66	–27.15	t.l.	1.05	–32.72
	14	30	–	4414.87	155	–	–23.95	460	–	–30.60	t.l.	0.49	–32.77	t.l.	2.01	–35.72
	15	46	–	3610.20	616	–	–17.45	685	–	–21.64	1384	–	–24.40	t.l.	2.00	–28.99
	16	27	–	3529.37	464	–	–12.22	279	–	–17.96	832	–	–21.08	t.l.	2.43	–25.78
	17	42	–	3869.30	329	–	–19.39	2663	–	–27.56	944	–	–28.00	4916	–	–32.09
	18	125	–	3751.53	1051	–	–9.03	3678	–	–18.94	t.l.	2.10	–21.75	t.l.	3.53	–24.60
	19	115	–	4115.47	1116	–	–21.49	t.l.	1.16	–27.42	3357	–	–27.85	t.l.	2.40	–33.59
	20	20	–	3917.50	521	–	–18.35	705	–	–24.85	5024	–	–27.55	t.l.	2.29	–32.35
0.75	21	1	–	5089.63	14	–	–16.62	83	–	–20.58	78	–	–22.60	142	–	–22.43
	22	3	–	5763.67	20	–	–12.97	127	–	–21.84	1952	–	–24.01	818	–	–26.31
	23	3	–	5152.53	4	–	–9.66	31	–	–11.22	125	–	–17.10	579	–	–21.35
	24	2	–	4639.60	1	–	–4.69	21	–	–12.55	24	–	–16.19	66	–	–19.03
	25	2	–	5236.70	22	–	–9.55	25	–	–21.38	121	–	–25.08	180	–	–24.87
	26	1	–	5475.20	4	–	–15.49	33	–	–25.37	29	–	–30.15	15	–	–34.05
	27	1	–	5747.90	18	–	–19.94	88	–	–27.76	193	–	–31.94	373	–	–32.47
	28	2	–	4686.43	4	–	–5.42	32	–	–12.29	57	–	–12.25	200	–	–13.18
	29	3	–	4807.63	15	–	–7.62	250	–	–12.96	59	–	–11.75	77	–	–13.62
	30	2	–	5029.83	13	–	–14.92	4	–	–21.48	9	–	–20.65	44	–	–24.82

requires 4026 s, whereas $S = 500$ requires 1459 s to be solved. Moreover, in instance #19, $S = 300$ cannot be solved within the time limit, but $S = 400$ is solved in 3357 s. For this class of problems, we may conclude that the number of scenarios (agents) is not the main factor determining the hardness of an instance.

In order to evaluate the effect of the increasing number of scenarios on the objective value, we take the case of $S = 100$ scenarios, always solved to optimality by model MKP(0,30), as the base case. For the base case, we report in column $M_{0,30}$ the optimal objective values (bold figures). For all the other cases, we report in column $M_{0,30}$ the percentage difference with respect to the base case. We see that the differences are remarkable. The figures referred to the instances with $\alpha = 0.75$ are more significant, as in this case all instances are resolved to optimality. For these instances, we see that the differences are increasing with the number of scenarios, but the rate of increment is clearly decreasing. This can be explained as follows. In this example, β -average optimization is used on a set of scenarios that is a sample from a theoretical distribution. Thus, the optimal value obtained by model MKP(0,30) is an approximation of the real value of $M_{0,30}$. As the cardinality of the sample increases, the quality of the approximation improves, converging to the real value. For a discussion on scenario generation techniques, we refer the reader to the specialized literature (see, e.g., Guastaroba et al., 2009). Concerning the results of the optimization process, we mention that not only the objective values, but also the returned solutions differ considerably. The interested reader may find details in the supplementary material.

5.2. Example of variability: a facility location problem

Let us consider a set of S customers, each with unit demand, a set of m potential locations for p facilities, and an $m \times S$ matrix

$C = [c_{i\ell}]$ of distances from potential locations to customers. The p -median problem is to select p potential facilities in order to minimize the total distance from customers to selected facilities, assuming that each customer is supplied from the closest selected facility (Drezner and Hamacher, 2004). The feasible set of the (uncapacitated) p -median problem is described as

$$XY = \left\{ x \in \{0, 1\}^{m \times S}, y \in \{0, 1\}^m : \sum_{i=1}^m x_{i\ell} = 1, \ell = 1, \dots, S; \sum_{i=1}^m y_i = p, x_{i\ell} \leq y_i, i = 1, \dots, m, \ell = 1, \dots, S \right\}.$$

The classic p -median problem is described as follows:

$$p\text{-MEDIAN: } \min \left\{ \sum_{i=1}^m \sum_{\ell=1}^S c_{i\ell} x_{i\ell} : (x, y) \in XY \right\}.$$

To include the p -median in our framework, we have to consider the distance matrix C as the average over S different cost functions, each one corresponding to an agent, in this case a customer. Formally, this can be done by expanding C into S cost vectors c^ℓ of mS entries, with $\ell = 1, \dots, S$, defined as

$$c_k^\ell = \begin{cases} c_{i\ell} & \text{if } k = (\ell - 1)m + i \\ 0 & \text{otherwise,} \end{cases}$$

where $i = 1, \dots, m$ and $k = 1, \dots, mS$. In other words, all entries of c_k^ℓ are set to zero, except for the entries from $k = (\ell - 1)m + 1$ to $k = \ell m$, that correspond to the ℓ th column of C , i.e., the ℓ th customer. To avoid cumbersome notation, we shall not use explicitly this expansion.

The p -center problem requires to minimize the maximum distance between a customer and its closest facility. This corresponds

Table 3
Facility location instances with 100 customers from the OR-library.

Instance	Model	Max	$M_{0.05}$	$M_{0.10}$	Avg	Time
pmed1	P -CENTER	–	7.08	11.33	47.45	7
	P -MEDIAN(0.05)	0.00	–	1.50	21.12	251
	P -MEDIAN(0.10)	0.79	0.34	–	21.60	199
	P -MEDIAN	4.72	2.02	2.04	–	1
pmed2	P -CENTER	–	4.68	7.92	44.40	70
	P -MEDIAN(0.05)	1.56	–	2.02	21.81	935
	P -MEDIAN(0.10)	7.81	2.68	–	22.09	936
	P -MEDIAN	19.53	17.56	14.51	–	5
pmed3	P -CENTER	–	2.13	2.96	35.22	4
	P -MEDIAN(0.05)	1.57	–	0.42	24.40	143
	P -MEDIAN(0.10)	0.00	0.00	–	27.18	159
	P -MEDIAN	59.06	27.00	15.13	–	1
pmed4	P -CENTER	–	1.68	5.10	40.15	5
	P -MEDIAN(0.05)	3.70	–	0.72	21.96	238
	P -MEDIAN(0.10)	11.85	1.38	–	18.44	267
	P -MEDIAN	31.85	6.88	2.95	–	1
pmed5	P -CENTER	–	6.51	10.98	43.07	5
	P -MEDIAN(0.05)	2.59	–	0.79	25.59	382
	P -MEDIAN(0.10)	8.62	2.79	–	21.40	376
	P -MEDIAN	29.31	13.20	7.62	–	1

to using a Minimax model, formulated as

$$p\text{-CENTER: } \min \left\{ u : u \geq \sum_{i=1}^m c_{i\ell} x_{i\ell}, \ell = 1, \dots, S; (x, y) \in XY \right\}.$$

In this context, a Minimax(β) model aims at minimizing the total distance travelled by the $k = \lceil \beta S \rceil$ customers with largest distances from their closest facilities. This leads to the following model

$$p\text{-MEDIAN}(\beta): \min \left\{ ku + \sum_{\ell=1}^n v_{\ell} : ku + kv_{\ell} \geq \sum_{i=1}^m c_{i\ell} x_{i\ell}, \ell = 1, \dots, S; v_{\ell} \geq 0, \ell = 1, \dots, S; (x, y) \in XY \right\}.$$

We compare the solutions obtained by the three models above on a set of 20 benchmark instances taken from the OR-library, namely those described in files pmed z .txt with $z = 1, 2, \dots, 20$. In all instances, the customer set and the potential location set are the same. To assess the quality of the obtained solutions, we first focus on the 5 smallest instances, having 100 customers, i.e., $S = 100$. On each instance, we test four models: P -CENTER, P -MEDIAN(0.05), P -MEDIAN(0.10), and P -MEDIAN. The value of p is fixed to 5 for all instances. The obtained solutions are compared in Table 3.

The structure of Table 3 is similar to Table 1, except for column *Gap*, that has been omitted as all instances are solved within the time limit of 7200 s. Columns *Max*–*Avg* are filled similarly to columns *Min*–*Avg* in Table 3, but taking into account that we have a minimization problem. The objective of the P -CENTER model is to minimize the maximum possible distance of a customer from the closest active facility, the objective of the P -MEDIAN(β) is to minimize M_{β} , i.e., the average of the $\lceil \beta S \rceil$ worst distances from the closest facility, and the objective of the P -MEDIAN model is to minimize the distance from the closest facility averaged over all customers. To emphasize the fact that in this case the regrets correspond to increments with respect to the optimal value, they are computed in positive form.

According to the results, models P -MEDIAN(0.05) and P -MEDIAN(0.10) always offer a compromise solution between P -CENTER and P -MEDIAN in terms of regrets. More precisely, the

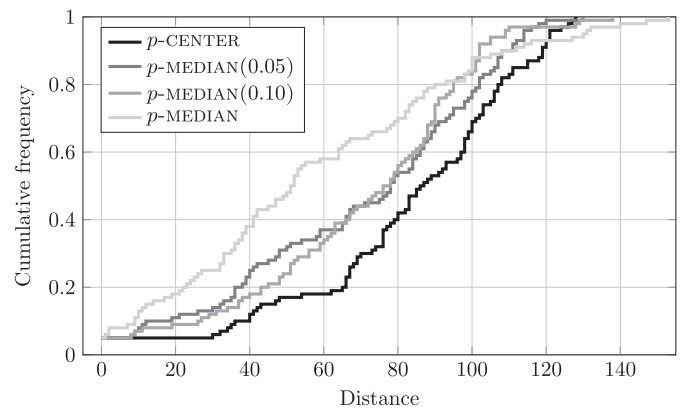


Fig. 5. Cumulative frequency plots for instance pmed2.

P -MEDIAN(0.05) and P -MEDIAN(0.10) models ensure a much better performance than the P -CENTER model in terms of average objective at the expense of a reasonable increment of the largest distance from an active facility.

From Table 3 it is clear that the P -MEDIAN and the P -CENTER models are much easier to solve for CPLEX. This phenomenon might be explained by the fact that CPLEX default parameters and settings are optimized over sets of benchmark instances for classical models like P -MEDIAN and P -CENTER.

In Fig. 5, we plot the cumulative frequency distribution (14) for instance pmed2, taken as an example. The pattern is found in all instances: for most distances δ from the closest facility, the solution of the P -MEDIAN model dominates the solution of the P -CENTER model, in the sense that the percentage of customers with distance not exceeding δ is larger in the P -MEDIAN solution. However, for the largest distances, the role is swapped, and the P -MEDIAN solution allows a certain fraction of customers to have a very large distance from their facilities. From the figures it is also clear that the P -MEDIAN(0.05) and the P -MEDIAN(0.10) solutions mediate between the Minimax criterion and the Average criterion, not only in terms of values (average and maximum) but also in terms of distribution.

To better evaluate the computational effort required by P -MEDIAN(β), we fix $\beta = 0.10$ and consider instances pmed z with $z = 1, \dots, 20$. Here, the number of customers S increases from 100 to 400. The results are summarized in Table 4, where the

Table 4
Facility location instances from the OR-library: times and gaps.

Instance	S	Time	Gap
pmed1	100	199	–
pmed2	100	936	–
pmed3	100	159	–
pmed4	100	267	–
pmed5	100	376	–
pmed6	200	80	–
pmed7	200	3255	–
pmed8	200	774	–
pmed9	200	540	–
pmed10	200	1093	–
pmed11	300	3087	–
pmed12	300	4370	–
pmed13	300	2221	–
pmed14	300	6885	–
pmed15	300	3979	–
pmed16	400	t.l.	1.71
pmed17	400	t.l.	6.88
pmed18	400	t.l.	5.06
pmed19	400	t.l.	6.53
pmed20	400	t.l.	5.94

positive correlation between S and the computational effort is apparent. In fact, we also tried to solve instances $p_{med21}, \dots, p_{med25}$, where $S = 500$, but none of them could be solved within the time limit. We notice however that the optimality gaps returned for such instances were not significantly larger than those returned for $p_{med16}, \dots, p_{med20}$ and reported in Table 4.

6. Conclusions

In this paper, the role of k -sum optimization as an attractive criterion to trade-off between a conservative Minimax/Maximin criterion and an aggressive minimum total cost/maximum total profit criterion in MILP models has been revised. Starting from the definition of β -average, that generalizes the concept of k -sum, and using only LP theory and basic statistical concepts, it has been shown how to embed the β -average measure in a MILP model. We have shown several nice properties of the β -average criterion, and clarified its strong connection with the Conditional Value-at-Risk. The proposed concepts and models have a great potential for application in many different areas where fairness or uncertainty is a concern.

Several research directions remain to be explored. First, the β -average measure could be extended to the case of weighted agents, enlarging its applicability. Second, as the illustrative examples suggest, the β -average optimization problem poses new computational issues, concerning the most effective exact solution method and the most appropriate valid inequalities. Benders decomposition (or L-shaped method) seems a natural starting point (see, e.g., Künzi-Bay and Mayer, 2006 and Sarin et al., 2014). Moreover, the analysis of local search techniques for β -average models is also worth of investigation, for their impact on the design of efficient heuristics. Finally, it would be interesting to embed the β -average optimization model in specific optimization problems.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.cor.2019.01.010](https://doi.org/10.1016/j.cor.2019.01.010)

References

- Andreatta, G., Mason, F., 1985a. k -eccentricity and absolute k -centrum of a probabilistic tree. *Eur. J. Oper. Res.* 19 (1), 114–117. doi:[10.1016/0377-2217\(85\)90315-7](https://doi.org/10.1016/0377-2217(85)90315-7).
- Andreatta, G., Mason, F., 1985b. Properties of the k -centra in a tree network. *Networks* 15 (1), 21–25. doi:[10.1002/net.3230150103](https://doi.org/10.1002/net.3230150103).
- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Math. Finance* 9 (3), 203–228. doi:[10.1111/1467-9965.00068](https://doi.org/10.1111/1467-9965.00068).
- Beasley, J.E., 1990. OR-library: distributing test problems by electronic mail. *J. Oper. Res. Soc.* 41 (11), 1069–1072.
- Bertsimas, A., Sim, M., 2004. The price of robustness. *Oper. Res.* 52, 35–53.
- Bertsimas, D., Brown, D.B., 2009. Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57 (6), 1483–1495.
- Chu, P.C., Beasley, J.E., 1998. A genetic algorithm for the multidimensional knapsack problem. *J. Heurist.* 4 (1), 63–86. doi:[10.1023/A:1009642405419](https://doi.org/10.1023/A:1009642405419).
- , 2004. In: Drezner, Z., Hamacher, H.W. (Eds.), *Facility Location: Applications and Theory*. Springer, Berlin.
- Duin, C., Volgenant, A., 1997. The partial sum criterion for steiner trees in graphs and shortest paths. *Eur. J. Oper. Res.* 97 (1), 172–182. doi:[10.1016/S0377-2217\(96\)00113-0](https://doi.org/10.1016/S0377-2217(96)00113-0).
- Ehrgott, M., 2006. *Multicriteria Optimization*. Springer Science & Business Media, Berlin.
- Garfinkel, R., Fernández, E., Lowe, T.J., 2006. The k -centrum shortest path problem. *TOP* 14 (2), 279–292. doi:[10.1007/BF02837564](https://doi.org/10.1007/BF02837564).
- Grygiel, G., 1981. Algebraic Σ_k assignment problem. *Control Cybern.* 10 (3–4), 155–165.
- Guastaroba, G., Mansini, R., Speranza, M.G., 2009. On the effectiveness of scenario generation techniques in single-period portfolio optimization. *Eur. J. Oper. Res.* 192 (2), 500–511.
- Gupta, S.K., Punnen, A.P., 1990. k -sum optimization problems. *Oper. Res. Lett.* 9 (2), 121–126. doi:[10.1016/0167-6377\(90\)90051-6](https://doi.org/10.1016/0167-6377(90)90051-6).
- Halpern, J., 1976. The location of a center-median convex combination on an undirected tree. *J. Reg. Sci.* 16 (2), 237–245.
- Halpern, J., 1978. Finding minimal center-median convex combination (cent-dian) of a graph. *Manage. Sci.* 24 (5), 535–544.
- Kaut, M., Wallace, S.W., 2007. Evaluation of scenario-generation methods for stochastic programming. *Pacific J. Optim.* 3 (2), 257–271.
- Kellerer, H., Pferschy, U., Pisinger, D., 2004. *Knapsack Problems*. Springer-Verlag, Berlin.
- Kostreva, M.M., Ogryczak, W., Wierzbicki, A., 2004. Equitable aggregations and multiple criteria analysis. *Eur. J. Oper. Res.* 158 (2), 362–377. doi:[10.1016/j.ejor.2003.06.010](https://doi.org/10.1016/j.ejor.2003.06.010).
- Künzi-Bay, A., Mayer, J., 2006. Computational aspects of minimizing conditional value-at-risk. *Comput. Manag. Sci.* 3, 3–27.
- Marshall, A.W., Olkin, I., Arnold, B.C., 2011. *Inequalities: Theory of Majorization and Its Applications*. Springer, New York.
- Nickel, S., Puerto, J., 2005. *Location Theory: A Unified Approach*. Springer-Verlag, Berlin.
- Ogryczak, W., 2014. Tail mean and related robust solution concepts. *Int. J. Syst. Sci.* 45 (1), 29–38.
- Ogryczak, W., Luss, H., Pióro, M., Nace, D., Tomaszewski, A., 2014. Fair optimization and networks: a survey. *J. Appl. Math.* 2014 (Article ID 612018), 1–25.
- Ogryczak, W., Śliwiński, T., 2002. On equitable approaches to resource allocation problems: the conditional minimax solutions. *J. Telecommun. Inf.Technol.* 2002 (3), 40–48.
- Ogryczak, W., Śliwiński, T., 2003. On solving linear programs with the ordered weighted averaging objective. *Eur. J. Oper. Res.* 148 (1), 80–91. doi:[10.1016/S0377-2217\(02\)00399-5](https://doi.org/10.1016/S0377-2217(02)00399-5).
- Ogryczak, W., Tamir, A., 2003. Minimizing the sum of the k largest functions in linear time. *Inf. Process. Lett.* 85 (3), 117–122. doi:[10.1016/S0020-0190\(02\)00370-8](https://doi.org/10.1016/S0020-0190(02)00370-8).
- Pinedo, M., 2012. *Scheduling – Theory, Algorithms, and Systems*. Springer, New York.
- Puerto, J., Rodríguez-Chía, A.M., 2015. Ordered median location problems. In: Laporte, G., Nickel, S., Saldanha da Gama, F. (Eds.), *Location Science*, chapter 10. Springer International Publishing, Cham, pp. 249–288.
- Puerto, J., Rodríguez-Chía, A.M., Tamir, A., 2017. Revisiting k -sum optimization. *Math. Program.* 165, 579–604.
- Puerto, J., Tamir, A., 2005. Locating tree-shaped facilities using the ordered median objective. *Math. Program.* 102, 313–338.
- Punnen, A.P., 1992. k -sum linear programming. *J. Oper. Res. Soc.* 43 (4), 359–363.
- Punnen, A.P., Aneja, Y.P., 1996. On k -sum optimization. *Oper. Res. Lett.* 18 (5), 233–236. doi:[10.1016/0167-6377\(95\)00059-3](https://doi.org/10.1016/0167-6377(95)00059-3).
- Rawls, J., 1971. *A Theory of Justice*. Harvard University Press, Cambridge, MA.
- Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. *J. Risk* 2, 21–41.
- Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. *J. Banking Finance* 26 (7), 1443–1471.
- Sarin, S., Sherali, H., Liao, L., 2014. Minimizing conditional-value-at-risk for stochastic scheduling problems. *J. Schedul.* 17, 5–15.
- Sarykalin, S., Serraino, G., Uryasev, S., 2008. Value-at-risk vs. conditional value-at-risk in risk management and optimization. In: Chen, Z.-L., Raghavan, S. (Eds.), *Tutorials in Operations Research. INFORMS*, pp. 270–294.
- Slater, P.J., 1978. Centers to centroids in graphs. *J. Graph Theory* 2 (3), 209–222. doi:[10.1002/jgt.3190020304](https://doi.org/10.1002/jgt.3190020304).
- Tamir, A., 2001. The k -centrum multi-facility location problem. *Discrete Appl. Math.* 109 (3), 293–307. doi:[10.1016/S0166-218X\(00\)00253-5](https://doi.org/10.1016/S0166-218X(00)00253-5).
- Wald, A., 1950. *Statistical Decision Functions*. John Wiley, New York.
- Woeginger, G., 1991. On minimizing the sum of k tardiness. *Inf. Process. Lett.* 38 (5), 253–256. doi:[10.1016/0020-0190\(91\)90067-R](https://doi.org/10.1016/0020-0190(91)90067-R).