Université Libre de Bruxelles

CoDE - SMG

Inverse optimization applied to PROMETHEE II weights to assess ranking stability

CoDE-SMG – Technical Report Series

N.A.V. Doan, Y. De Smet

CoDE-SMG – Technical Report Series

Technical Report No.
TR/SMG/2016-001

May 2016

CoDE-SMG – Technical Report Series
ISSN 2030-6296

Published by:
CoDE-SMG, CP 210/01
Université Libre de Bruxelles
Bvd du Triomphe
1050 Ixelles, Belgium

Technical report number TR/SMG/2016-001

Inverse optimization applied to PROMETHEE II weights to assess ranking stability

CoDE-SMG – Technical Report Series

N.A.V. Doan nguyen.anh.vu.doan@ulb.ac.be
Y. De Smet yves.de.smet@ulb.ac.be
CoDE-SMG, Université Libre de Bruxelles, Brussels, Belgium

May 2016

Inverse optimization applied to PROMETHEE II weights to assess ranking stability

N.A.V. Doan[a,*], Y. De Smet[a]

[a] *CoDE-SMG lab, Université libre de Bruxelles, 50 Avenue F.D. Roosevelt, 1050 Brussels, Belgium*

Abstract

Most strategic decision problems involve the simultaneous optimization of several conflicting criteria. Among multicriteria decision aid methods, PROMETHEE has earned a lot of attention during the past decades. This method provides a complete ranking of a set of alternatives, given some parameters given by a decision maker, such as the weights of the criteria. In order to assess the stability of the ranking, weights stability intervals (WSI) have been developed. However, this method only focuses on one criterion at a time (changes are assumed to be applied uniformly to the other criteria in order to remain normalized). In this work, we analyze how weights stability intervals can be redefined while allowing any modifications on any criterion weight, by applying inverse optimization based on mixed integer linear programming (MILP). The problem formulation can be stated as follows: what would be the minimum modification on the weights such that a given alternative a_i of rank $n > 1$ becomes first? This methodology has been applied on two case studies, and compared to the WSI. The results show that it is always possible to find narrower intervals while it is never required to modify all the weights simultaneously. Additionally, when using the WSI, only few alternatives can be ranked first. By taking the inverse optimization point of view, all (non-dominated) alternatives can also be considered successively for the stability analysis of the first-ranked. It is therefore possible to apply this methodology for the research of a consensus when different decision makers are involved.

Keywords: multicriteria decision aid, ranking stability, inverse optimization, PROMETHEE

1. Introduction

Most strategic decision problems involve the simultaneous optimization of several conflicting criteria. For instance, in a procurement conducted by a transport company, the buyer (looking for new trucks) wants to simultaneously optimize: the investment and operational costs, both the quality of the vehicle and the supplier, the time of delivery, the mean time before failure, etc. Another example can be the design of integrated circuits: a manufacturer will try to simultaneously maximize the performance while minimize the cost of the chip. However, one can already guess that those two objectives are conflicting. Also, producing high-end integrated circuits can be subject to more difficulties in terms of thermal dissipation. In addition, a criterion based on ecological standards may have impacts on the cost and the performance [1].

Many methods have been developed to address multicriteria problems. One can cite the Multi-Attribute Utility Theory [2], Analytical Hierarchy Process [3], ELECTRE methods [4], MACBETH [5], PROMETHEE methods [6], etc.

In this contribution, we will focus on PROMETHEE methods. These have been applied in hundreds of applications in finance, health care, environmental management, transport, sports, hydrology and water management, production, etc. [7]. This success is due to their simplicity and the existence of user-friendly softwares such as D-SIGHT [8].

*Corresponding author

Email addresses: Nguyen.Anh.Vu.Doan@ulb.ac.be (N.A.V. Doan), Yves.De.Smet@ulb.ac.be (Y. De Smet)

PROMETHEE II aims to provide a complete ranking of a set of alternatives. In order to assess the robustness of this ranking, weights stability intervals (WSI) have been developed [9]. This method allows to analyze the modifications that can be applied to criteria weights without modifying the ranking. However, it only focuses on one criterion at a time (changes are assumed to be applied uniformly to the other criteria in order to remain normalized). In this work, we will analyze how weights stability intervals can be redefined while allowing any modifications on any criterion weight, by applying an inverse optimization methodology based on linear programming.

This approach will be applied on two case studies: the best cities ranking published by the Economist Intelligence Unit [10] and the Environmental Performance Index provided by two research centres of the Columbia University [11].

This paper is organized as follows: in Section 2 we will recall the main steps required to compute the PROMETHEE II ranking and the WSI. Next, in Section 3, we will define the mixed integer linear program that has been developed to solve the inverse optimization problem and apply it to an example in Section 4. Finally, in Section 5 we will analyze the performance of this approach using the previously-described case studies.

2. Computing PROMETHEE II rankings and weights stability intervals

In this section, we remind the basic steps to compute the PROMETHEE II ranking and the weights stability intervals. Of course, a detailed description of these approaches goes beyond the scope of this contribution. Therefore we refer the interested reader to [12] (PROMETHEE) and [9] (WSI) for detailed analyses.

2.1. PROMETHEE II

Let $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ be a set of n alternatives and $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ be a set of m criteria. Without loss of generality, we assume that all criteria have to be maximized. The PROMETHEE methods are based on pairwise comparisons. At first, each pair of alternatives $a_i, a_j \in \mathcal{A}$ is compared on each criterion f_k:

$$d_k(a_i, a_j) = f_k(a_i) - f_k(a_j) \tag{1}$$

The quantity $d_k(a_i, a_j)$ represents the *advantage* of a_i over a_j for criterion f_k. On the one hand, when $d_k(a_i, a_j)$ is small enough, there is no good reason to say that a_i is better than a_j regarding criterion f_k. On the other hand, when $d_k(a_i, a_j)$ exceeds a certain limit, the decision maker (DM) may express that a_i is strictly preferred to a_j for f_k. In order to model these statements, the difference $d_k(a_i, a_j)$ is transformed into a unicriterion preference degree, denoted $P_k(a_i, a_j)$, by using a non-decreasing function H_k;

$$P_k(a_i, a_j) = H_k(d_k(a_i, a_j)), \ \forall\, a_i, a_j \ \in \ \mathcal{A} \tag{2}$$

The quantity $P_k(a_i, a_j) \in [0, 1]$ and $P_k(a_i, a_j) = 0$ when $d_k(a_i, a_j) < 0$. There are plenty of functions that can be considered to compute the unicriterion preference degrees. In most software implementing the PROMETHEE method, 6 main functions are considered [8]. Figure 1 represents the so-called linear preference function. Two thresholds characterize it:

- q_k plays the role of an *indifference* threshold. When the difference $d_k(a_i, a_j) \leq q_k$, it is considered to be so small that the unicriterion preference is equal to zero;
- p_k plays the role of a *preference* threshold, When the difference $d_k(a_i, a_j) \geq p_k$, it is considered to be important enough to state that a_i is strongly preferred to a_j for this criterion.

Between these two thresholds, the unicriterion preference is assumed to increase linearly.

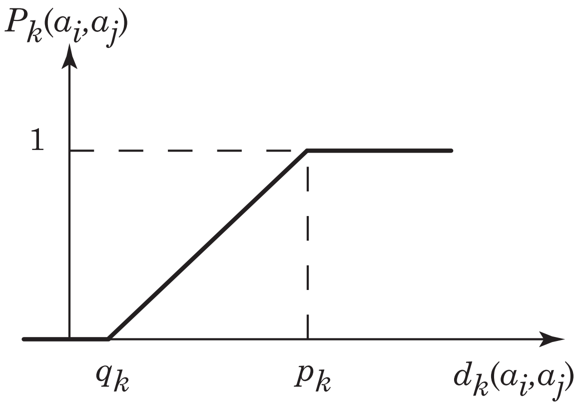


Figure 1: Generalized criterion of type 5

Once the unicriterion preference degrees between two actions a_i and a_j have been computed for every criterion, one has to aggregate these marginal contributions to obtain $P(a_i, a_j)$ i.e. a global measure of the preference of a_i over a_j:

$$P(a_i, a_j) = \sum_{k=1}^{m} w_k \cdot P_k(a_i, a_j) \tag{3}$$

where w_k represents the relative importance of criterion f_k. These weights are assumed to be positive and normalized. Obviously, we have $P(a_i, a_j) \geq 0$ and $P(a_i, a_j) + P(a_j, a_i) \leq 1$.

The PROMETHEE I and II rankings are based on the exploitation of the P matrix. Therefore, three flow scores are built.; the positive flow ϕ^+, the negative flow ϕ^- and the net flow ϕ scores:

$$\phi^+(a_i) = \frac{1}{n-1} \sum_{a_j \in \mathcal{A}, i \neq j} P(a_i, a_j) \tag{4}$$

$$\phi^-(a_i) = \frac{1}{n-1} \sum_{a_j \in \mathcal{A}, i \neq j} P(a_j, a_i) \tag{5}$$

$$\phi(a_i) = \phi^+(a_i) - \phi^-(a_j) \tag{6}$$

The PROMETHEE I ranking is obtained as being the intersection of the rankings induced by ϕ^+ and ϕ^-. The PROMETHEE II ranking is given by the ranking based on ϕ.

Finally, it is worth noting that:

$$\phi(a_i) = \frac{1}{n-1} \sum_{k=1}^{m} \sum_{a_j \in \mathcal{A}} [P_k(a_i, a_j) - P_k(a_j, a_i)] \cdot w_k = \sum_{k=1}^{m} \phi_k(a_i) \cdot w_k \tag{7}$$

where $\phi_k(a_i)$ is called the k^{th} unicriterion net flow score assigned to action a_i. At this point, the multicriteria problem can be considered as matrix containing the unicriterions net flow score evaluations. These values already integrates criterion preference information and all lie between -1 and 1.

The PROMETHEE I and II rankings provide prescriptive tools for decision making. The GAIA [13] tool complements them with a descriptive approach. The idea is to represent each alternative by its evaluations in the unicriterion net flow space:

$$\Phi(a_i) = [\phi_1(a_i), \phi_2(a_i), \ldots, \phi_m(a_i)] \tag{8}$$

GAIA is the result of a principal component analysis applied on this dataset. Therefore, the decision maker is able to visualize the decision problem on a plane and compare:

- the relative positions of alternatives (in order to identify groups of similar or distinct alternatives profiles);
- the relative positions of criteria (in order to identify conflicts or redundancies);
- the relative positions of alternatives with respect to a given criterion (in order to identify the best and worst alternatives for the different points of views);
- the relative positions of alternatives with respect to the so-called *decision stick* (in order the identify the best compromise solutions).

2.2. WSI

The determination of the precise weight values is often a tedious cognitive task for a decision maker. Uncertainties may exist and a question can be raised: how a change in the weights values can impact the ranking? The aim of weight stability intervals is to analyze how the weight value on one given criterion can be modified, while all the other weights change uniformly (in order to remain normalized), without altering the ranking.

In PROMETHEE II, the preference structure (P: preference, I: indifference) between a pair of alternatives (a_i, a_j) is defined as follows:

$$\begin{cases} a_i P a_j & \text{iff } \phi(a_i) > \phi(a_j) \\ a_i I a_j & \text{iff } \phi(a_i) = \phi(a_j) \end{cases} \tag{9}$$

The WSI will define values on the weights such that this (P, I) structure is not altered. Let us denote (P', I') the structure for modified weights w'_k. In the following, we will focus on a partial stability, namely a modification of the first-ranked alternative. We refer the interested readers to [9] for complete explanations.

The new set of weights (keeping them normalized) is defined as follows:

$$w'_k = (1+\beta)w_k, \quad \beta \geq -1, \quad \text{for the modified criterion } k \tag{10}$$

$$w'_l = \alpha w_l, \quad 0 \leq \alpha \leq \frac{1}{1-w_k}, \quad \forall l \neq k \tag{11}$$

$$\tag{12}$$

We define the following notations:

$$\Delta(a_i, a_j) = \phi(a_i) - \phi(a_j) \tag{13}$$

$$\Delta_k(a_i, a_j) = \phi_k(a_i) - \phi_k(a_j) \tag{14}$$

$$\Delta'(a_i, a_j) = \phi'(a_i) - \phi'(a_j) \tag{15}$$

It is then easy to show that:

$$\Delta'(a_i, a_j) = \alpha\Delta(a_i, a_j) + (1-\alpha)\Delta_k(a_i, a_j) \tag{16}$$

A condition for stability is then:

$$\Delta(a_i, a_j)\Delta'(a_i, a_j) > 0 \text{ s.t. } \Delta(a_i, a_j) \neq 0 \tag{17}$$

Using Equation (16), this condition becomes:

$$\alpha[\Delta(a_i, a_j)\Delta_k(a_i, a_j) - \Delta^2(a_i, a_j)] < \Delta(a_i, a_j)\Delta(a_i, a_j) \tag{18}$$

3 different cases can then occur:

- if f_k is strongly compatible with the ranking of a_i and a_j, we have $\Delta(a_i,a_j)\Delta_k(a_i,a_j) > \Delta^2(a_i,a_j)$ so the condition becomes:
$$\alpha < \frac{\Delta(a_i,a_j)\Delta_k(a_i,a_j)}{\Delta(a_i,a_j)\Delta_k(a_i,a_j) - \Delta^2(a_i,a_j)}, \quad \alpha > 1 \tag{19}$$
giving an upper bound for the stability interval;

- if f_k is not compatible with the ranking of a_i and a_j, we have $\Delta(a_i,a_j)\Delta_k(a_i,a_j) < 0$ so the condition becomes:
$$\alpha > \frac{\Delta(a_i,a_j)\Delta_k(a_i,a_j)}{\Delta(a_i,a_j)\Delta_k(a_i,a_j) - \Delta^2(a_i,a_j)}, \quad \alpha < 1 \tag{20}$$
giving a lower bound for the stability interval;

- if $0 \le \Delta(a_i,a_j)\Delta_k(a_i,a_j) \le \Delta^2(a_i,a_j)$, then no inversion is possible.

For the stability of a subset $\mathcal{M} \subset \mathcal{A}$ (for instance the m first ranked alternatives), let us define:

$$\Omega^0_{\mathcal{M}} = \{(a_i,a_j) \in \mathcal{M}\times\mathcal{A}, \text{ s.t. } \Delta(a_i,a_j) = 0 \text{ and } \Delta_k(a_i,a_j) < 0\} \tag{21}$$
$$\Omega^-_{\mathcal{M}} = \{(a_i,a_j) \in \mathcal{M}\times\mathcal{A}, \text{ s.t. } \Delta(a_i,a_j)\Delta_k(a_i,a_j) < 0\} \tag{22}$$
$$\Omega^+_{\mathcal{M}} = \{(a_i,a_j) \in \mathcal{M}\times\mathcal{A}, \text{ s.t. } \Delta(a_i,a_j)\Delta_k(a_i,a_j) > \Delta^2(a_i,a_j)\} \tag{23}$$
$$\alpha^-_k = \max_{(a_i,a_j)\in\Omega^-_{\mathcal{M}}} \frac{\Delta(a_i,a_j)\Delta_k(a_i,a_j)}{\Delta(a_i,a_j)\Delta_k(a_i,a_j) - \Delta^2(a_i,a_j)} \tag{24}$$
$$\alpha^+_k = \min_{(a_i,a_j)\in\Omega^+_{\mathcal{M}}} \frac{\Delta(a_i,a_j)\Delta_k(a_i,a_j)}{\Delta(a_i,a_j)\Delta_k(a_i,a_j) - \Delta^2(a_i,a_j)} \tag{25}$$

If $\Omega^0_{\mathcal{M}} \neq \emptyset$, then the stability interval for α is reduced to 1 and no change of weights is allowed. If $\Omega^0_{\mathcal{M}}$ is empty, then the stability interval is obtained for:

$$\alpha^-_k < \alpha < \alpha^+_k \tag{26}$$

Finally, the WSI for the k^{th} criterion is:

$$\text{WSI}_k = [w^-_k; w^+_k] = [1 + (1-w_k)\alpha^+_k; 1 + (1-w_k)\alpha^-_k] \tag{27}$$

As we can observe, this provides stability intervals where no modification can occur in the subset $\mathcal{M}$. However, as mentioned previously, this only considers one criterion at a time. In this contribution, we will analyze the stability of the first-ranked alternative when all the criteria can be simultaneously considered for a modification of their weight values.

The problem can thus be formulated as follows: for a PROMETHEE II instanciation (hence, a given related ranking), what would be the minimum modification of the weights such that a given alternative a_i of rank $n > 1$ becomes first. This can thus be considered as an inverse optimization problem on the PROMETHEE II ranking. This problem can be formulated as linear equations. Therefore, we can use mixed integer linear programming (MILP) to solve it.

3. MILP model

In this section, we detail the linear program that has been developed to address to aforementioned question. Let us note that in addition to the objective of minimizing the weights modification, we will also consider the minimization of the number of modified criteria. This will be introduced as an additional constraint (see Equations (38-40))

The decision variables are the set of new weights values $\mathcal{W}' = \{w'_1, w'_2, \ldots, w'_m\}$. The objective is to minimize the sum of (L_1) distances of these new weights from the initial ones: $\sum_{k=1}^{m} |w_k - w'_k|$. In order to linearize this absolute value, we define 3 other sets of variables: $\Delta = \{\delta_1, \delta_2, \ldots, \delta_m\}, \mathcal{D}_1 = \{d_{1,1}, d_{2,1}, \ldots, d_{m,1}\}, \mathcal{D}_2 = \{d_{1,2}, d_{2,2}, \ldots, d_{m,2}\}$ such that, $\forall k = 1, 2, \ldots, m$:

$$\delta_k = \begin{cases} 0 & \text{if } w_k - w'_k < 0 \\ 1 & \text{otherwise} \end{cases}, \quad \delta_k \in \{0,1\} \tag{28}$$

$$w_k - w'_k = \begin{cases} d_{k_1} & \text{if } w_k - w'_k \geq 0 \\ -d_{k_2} & \text{otherwise} \end{cases}, \quad d_{k_1}, d_{k_2} \geq 0 \tag{29}$$

In order to introduce a constraint on the number of allowed modified criteria, we also define the set $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_m\}$ such that, $\forall k = 1, 2, \ldots, m$:

$$\gamma_k = \begin{cases} 0 & \text{if } d_{k_1} + d_{k_2} = 0 \\ 1 & \text{otherwise} \end{cases}, \quad \gamma_k \in \{0,1\} \tag{30}$$

The constants of the problem are:

- the set of m initial weights for the criteria: $\mathcal{W} = \{w_1, w_2, \ldots, w_m\}$;
- the unicriterion net flow scores table:

a	$\phi_1(\cdot)$	$\phi_2(\cdot)$	$\ldots$	$\phi_k(\cdot)$	$\ldots$	$\phi_m(\cdot)$
a_1	$\phi_1(a_1)$	$\phi_2(a_1)$	$\ldots$	$\phi_k(a_1)$	$\ldots$	$\phi_m(a_1)$
a_2	$\phi_1(a_2)$	$\phi_2(a_2)$	$\ldots$	$\phi_k(a_2)$	$\ldots$	$\phi_m(a_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_i	$\phi_1(a_i)$	$\phi_2(a_i)$	$\ldots$	$\phi_k(a_i)$	$\ldots$	$\phi_m(a_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_n	$\phi_1(a_n)$	$\phi_2(a_n)$	$\ldots$	$\phi_k(a_n)$	$\ldots$	$\phi_m(a_n)$

;

- M, a constant so that $M \geq \frac{1}{d_{k_1}+d_{k_2}}$, $\forall k = 1, 2, \ldots, m$;
- $N \in \{2, 3, \ldots, m\}$, a constant for the constraint on the number of modified criteria).

The MILP can therefore be formalized as follows:

$$\min z = \sum_{k=1}^{m} |w_k - w'_k| = \sum_{k=1}^{m} d_{k_1} + d_{k_2} \tag{31}$$

s.t.

$$\sum_{k=1}^{m} w'_k = 1 \qquad \text{(weights constraint)} \tag{32}$$

$$\delta_k > \frac{w_k - w'_k}{2}, \forall k = 1, 2, \ldots, m \qquad \text{(absolute value linearization)} \tag{33}$$

$$\delta_k \leq \frac{w_k - w'_k}{2} + 1, \forall k = 1, 2, \ldots, m \tag{34}$$

$$w_k - w'_k = d_{k_1} - d_{k_2}, \forall k = 1, 2, \ldots, m \tag{35}$$

$$0 \leq d_{k_1} \leq \delta_k, \forall k = 1, 2, \ldots, m \tag{36}$$

$$0 \leq d_{k_2} \leq 1 - \delta_k, \forall k = 1, 2, \ldots, m \tag{37}$$

$$\gamma_k \geq d_{k_1} + d_{k_2}, \forall k = 1, 2, \ldots, m \qquad \text{(number of modified criteria)} \tag{38}$$

$$\gamma_k \leq M(d_{k_1} + d_{k_2}), \forall k = 1, 2, \ldots, m \tag{39}$$

$$\sum_{k=1}^{m} \gamma_k \leq N \qquad \text{(N allowed modified criteria)} \tag{40}$$

$$\phi'(a_i) = \sum_{k=1}^{m} w'_k \phi_k(a_i) \qquad \text{(net flow scores computation)} \tag{41}$$

$$\phi'(a_i) > \phi'(a_j), \forall j \neq i \qquad \text{(rank change of a_i)} \tag{42}$$

$$w_k, d_{k_1}, d_{k_2} \geq 0, \forall k = 1, 2, \ldots, m \qquad \text{(domain)} \tag{43}$$

$$\delta_k, \gamma_k \in \{0, 1\}, \forall k = 1, 2, \ldots, m \tag{44}$$

Let us note that in the best case, $z = 0$ and in the worst case, $z = 2$.

The MILP model has been implemented in Python using the formalism from the PuLP package and the Gurobi solver. The simulations have been carried out on an Intel® Core$^{\text{TM}}$ i5-5200U Processor (dual core, 2.7 GHz).

4. Illustrative example

Best cities ranking subset

In this section, we will show the results obtained with the MILP on an illustrative example and compared them to the classical WSI method. We will use a subset from the best cities ranking published by the Economist Intelligence Unit [10] shown in Table 1. It is composed of 5 cities evaluated on 6 criteria: stability, healthcare, culture and environment, education, infrastructure, and spatial characteristics, with weights respectively equal to 18.75%, 15%, 18.75%, 7.5%, 15%, 25% (in the following, we will refer to the weights in the same order). In our case, as the optimization only requires to use the unicriterion net flow scores and we focus on the performance of the optimization problem rather than the modeling, we arbitrarily choose to make use of usual preference functions for all the criteria for the sake of simplicity. Let us note that we will also make use of linear preference functions for the EPI ranking case study.

The obtained PROMETHEE II ranking is: Hong Kong, Stockholm, Rome, Atlanta, New York, and the associated weight stability intervals are shown in Table 2.

Each of the extreme values outside of the WSI will lead to another first-ranked city:

- the set of weights {0.61%, 18.35%, 22.94%, 9.17%, 18.35%, 30.58% } will rank Rome first, $z = 0.3628$;
- the set of weights {18.55%, 15.9%, 18.55%, 7.42%, 14.84%, 24.74% } will rank Stockholm first, $z = 0.018$;
- the set of weights {18.07%, 14.46%, 21.69%, 7.23%, 14.46%, 24.1% } will rank Stockholm first, $z = 0.0588$;

Table 1: Best cities ranking subset - Evaluation table

City	Stability	Healthcare	Culture and Environment	Education	Infrastructure	Spatial Characteristics
Hong Kong	95	87.5	85.9	100	96.4	75
Stockholm	95	95.8	91.2	100	96.4	58.9
Rome	80	87.5	91.7	100	92.9	67.3
New York	70	91.7	91.7	100	89.3	65.2
Atlanta	85	91.7	91.7	100	92.9	42.9

Table 2: Best cities ranking subset - Weight stability intervals

Criteria	Minimum weight	Current weight	Maximum weight
Stability	0.61%	18.75%	100%
Healthcare	0%	15%	15.9%
Culture and Environment	0%	18.75%	21.69%
Education	0%	7.5%	100%
Infrastructure	0%	15%	100%
Spatial Characteristics	24.05%	25%	100%

- the set of weights {24.05%, 15.19%, 7.59%, 18.99%, 15.19%, 18.99% } will rank Stockholm first, $z = 0.019$.

With the MILP, we have:

- when Rome is ranked first and at most 2 criteria are allowed to change: {9.44%, 15%, 28.06%, 7.5%, 15%, 25% }, $z = 0.18612$;
- when Rome is ranked first and at most 3 criteria are allowed to change (minimum value for z achieved): {11.27%, 13.51%, 27.72%, 7.5%, 15%, 25% }, $z = 0.1795$;
- when Stockholm is ranked first and at most 2 criteria are allowed to change (minimum value for z achieved): {18.75%, 15.58%, 18.75%, 7.5%, 15%, 24.42% }, $z = 0.0116$.

We can observe that, in all cases, these modifications do not impact all the criteria simultaneously. Besides, the distances from the initial sets are always lower than the WSI. This shows that the ranking robustness can considered to actually be narrower than what is defined by the WSI. The computation time for this illustrative case is 90 ms.

GAIA representation

Additionally, we can represent these results using the GAIA plane [13], as shown in Figure 2. The black dots on the graph represent the location of the decision stick using the weight sets from the MILP. These delimit the area where the ranking is stable with respect to the MILP.

Multiple decision makers - reaching a consensus

It is also worth noting that, when using the WSI, only 2 alternatives can be ranked first. By taking the inverse optimization point of view, other (non-dominated) alternatives can also be considered for the stability analysis of the first-ranked. An interesting application can be the research of a consensus between

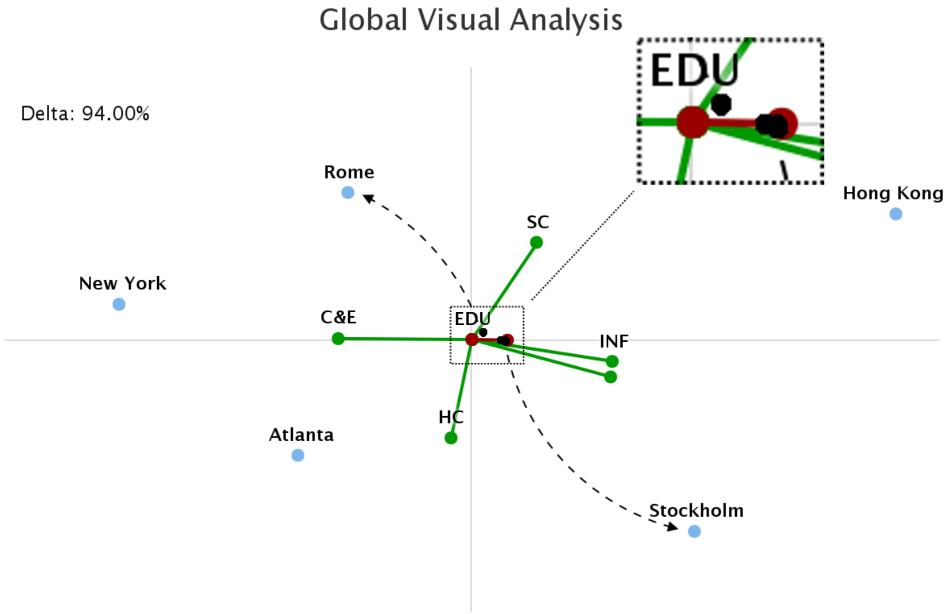


Figure 2: MILP weights stability

different decision makers. Let us suppose that we have q actors in a decision process, with each having their own set of weights. Let us also consider that each DM accepts small changes on their weights. The problem can therefore be formulated as follows: what would be the minimum modification for all decision makers such that the same alternative is ranked first in all the q rankings? Solving this will only require to generalize the MILP to consider multiple actors and to analyze which of all the alternatives can be ranked first while minimizing the distance to the initial weights sets:

$$\min z = \sum_{p=1}^{q} \sum_{k=1}^{m} |w_{k,p} - w'_{k,p}| = \sum_{k=1}^{m} d_{k_1,p} + d_{k_2,p} \tag{45}$$

s.t.

$$\sum_{k=1}^{m} w'_{k,p} = 1, \forall p = 1, 2, \ldots, q \quad \text{(weights constraint)} \tag{46}$$

$$\delta_{k,p} > \frac{w_{k,p} - w'_{k,p}}{2}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \quad \text{(absolute value linearization)} \tag{47}$$

$$\delta_{k,p} \leq \frac{w_{k,p} - w'_{k,p}}{2} + 1, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{48}$$

$$w_{k,p} - w'_{k,p} = d_{k_1,p} - d_{k_2,p}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{49}$$

$$0 \leq d_{k_1,p} \leq \delta_{k,p}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{50}$$

$$0 \leq d_{k_2,p} \leq 1 - \delta_{k,p}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{51}$$

$$\gamma_{k,p} \geq d_{k_1,p} + d_{k_2,p}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \quad \text{(number of modified criteria)} \tag{52}$$

$$\gamma_{k,p} \leq M(d_{k_1,p} + d_{k_2,p}), \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{53}$$

$$\sum_{k=1}^{m} \gamma_{k,p} \leq N_p, \forall p = 1, 2, \ldots, q \quad \text{(N_p allowed modified criteria)} \tag{54}$$

$$\phi'_p(a_i) = \sum_{k=1}^{m} w'_{k,p} \phi_{k,p}(a_i), \forall p = 1, 2, \ldots, q \quad \text{(net flow scores computation)} \tag{55}$$

$$\phi'_p(a_i) > \phi'_p(a_j), \forall j \neq i; \forall p = 1, 2, \ldots, q \quad \text{(rank change of a_i)} \tag{56}$$

$$w_{k,p}, d_{k_1,p}, d_{k_2,p} \geq 0, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \quad \text{(domain)} \tag{57}$$

$$\delta_{k,p}, \gamma_{k,p} \in \{0, 1\}, \forall k = 1, 2, \ldots, m; \forall p = 1, 2, \ldots, q \tag{58}$$

For our illustrative case, we will consider two decision makers. Let us suppose that the first DM has the set of weights used in the Economist Intelligence Unit's study: 18.75%, 15%, 18.75%, 7.5%, 15%, 25%; and the second DM has the following set: 5%, 20%, 45%, 10%, 10%, 10%. This will lead to the following rankings:

- first decision maker: Hong Kong, Stockholm, Rome, Atlanta, New York (associated net flow scores: 0.203125, 0.184375, -0.025, -0.15625, -0.20625);
- second decision maker: Atlanta, New York, Rome Stockholm, Hong Kong (associated net flow scores: 0.15, 0.125, 0.075, 0.0375, -0.3875).

We can observe that the rankings are rather different from one DM to the other. The simulation results show that the best consensus is Stockholm with a distance of 0.0115 for the first DM and 0.1125 for the second DM. The complete consensus ranking is given in Table 3, and graphically represented in Figure 3 (the red point being the best consensus). We can observe that, while it represents the best global compromise, Stockholm requires the second DM to modify its weights more than the first DM, whereas the second consensus, Rome, would lead to the opposite situation.

Let us point out that we are considering all the decision makers share the same importance. Of course, in the case weights can be set on the relative importance of the different actors, another compromise alternative might be ranked first in the consensus ranking.

Table 3: Best cities ranking subset - 2 decision makers consensus ranking

	Consensus ranking	DM1 distance	DM2 distance	Total distance
1	Stockholm	0.0116	0.1125	0.1241
2	Rome	0.1795	0.0652	0.2447
3	New York	0.3543	0.025	0.3793
4	Atlanta	0.3844	0	0.3844
5	Hong Kong	0	0.45	0.45

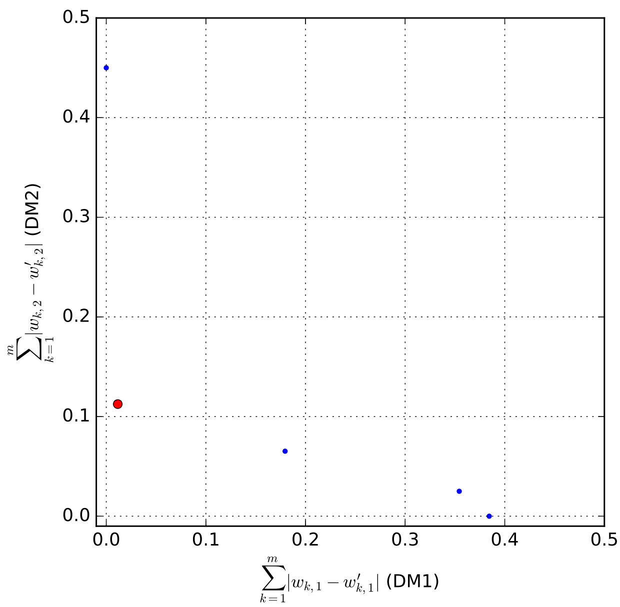


Figure 3: Best cities ranking subset - 2 decision makers consensus

In the next section, we will analyze the performance of the MILP compared to the WSI with two different case studies.

5. Case studies - Performance analysis

5.1. Experimental setup

For each case study, a sample of random alternatives chosen uniformly is considered for the MILP. The PROMETHEE II ranking is established as well as its weight stability intervals. Each extreme value outside of the WSI will lead to the change of the first-ranked alternative. These new first-ranked alternatives will be used in the MILP for the constraint (43) in order to compare how the weights are changed between both methods. This process will be repeated for several iterations with different random subsets of alternatives in order to gather data for different PROMETHEE instanciations.

For the case studies, we have chosen a sample size of 15 randomly-chosen alternatives from each set and 500 iterations have been performed.

5.2. *Best cities ranking*

The first case study is based on the best cities ranking published by the Economist Intelligence Unit [10]. This dataset is composed of 70 cities evaluated on 6 criteria: stability, healthcare, culture and environment, education, infrastructure, and spatial characteristics (with weights respectively equal to 18.75%, 15%, 18.75%, 7.5%, 15%, 25%).

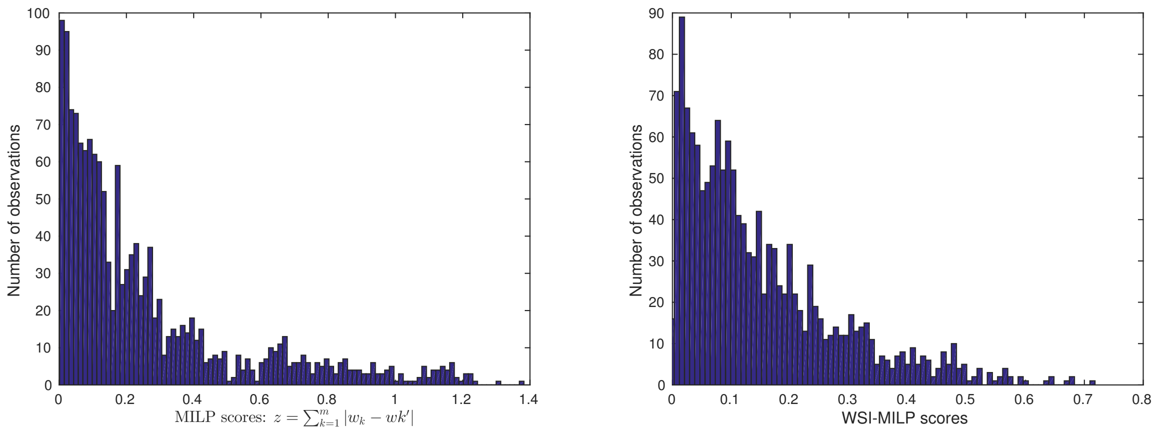
This study consists in an update of the existing EIU Liveability index to which a sixth criterion has been added to take into account spatial characteristics. This added factor carries a weight of 25% and seeks to account for spatial aspects such as urban form, the geographical situation of the city, cultural assets and pollution. As for the illustrative example in Section 4, we arbitrarily choose to make use of usual preference functions for all the criteria, for the same aforementioned reasons. Linear preference functions will be used for the next case study (EPI ranking).

The results obtained show that for all the simulated instanciations, it is always possible to find a weights modification with the MILP that is lower than the WSI. In addition, none of the simulations implies the modification of all the criteria simultaneously. In a large majority of cases only 2 criteria needed to be modified, as shown in Table 4. The distribution of MILP scores ($z = \sum_{k=1}^{m} |w_k - w'_k|$) is given in Figure 4a. The total number of observations is 1489, which is greater than the number of iterations (500). This is due to the fact that for each iteration, more than one alternative can be ranked first when using the WSI bounds. The median is equal to 0.1536 which is 11.11% of the maximum score (1.3822).

The distribution of the difference between the WSI and the MILP scores is given in Figure 4b. As we can observe, the values are all positive (MILP scores better than WSI) and the median is equal to 0.107 which is 14.9% of the maximum difference (0.7181). The computation time per iteration is 900 ms on average.

Table 4: Best cities ranking - Distribution of the number of modified criteria

Number of modified criteria	2	3	4	5	6
Proportion of cases	79.47%	14.47%	3.68%	2.37%	0%



(a) MILP scores

(b) Difference between the WSI and the MILP scores

Figure 4: Best cities ranking - Distributions

5.3. Environmental Performance Index

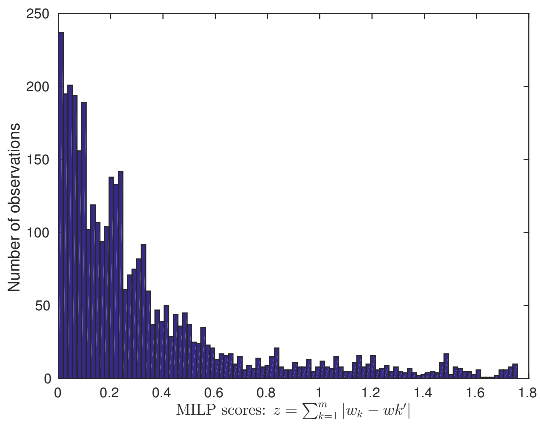
For the second case study, we use the Environmental Performance Index (EPI), a joint project between the Yale Center for Environmental Law & Policy (YCELP) and the Center for International Earth Science Information Network (CIESIN) at Columbia University, in collaboration with the World Economic Forum and support from the Samuel Family Foundation and the McCall MacBain Foundation [11]. The countries are evaluated on 9 criteria: air quality, health impacts, water sanitation, agriculture, biodiversity habitat, climate energy, fisheries, forests, water resources (with weights respectively equal to 16.67%, 16.67%, 16.67%, 2.5%, 12.5%, 12.5%, 5%, 5%, 12.5%). As for the previous case study, we can arbitrarily choose the preference functions and we decide, for all the criteria, to make use of linear functions, with the indifference thresholds set on the first quartile of all the evaluation differences and the preference thresholds set on the third quartile.

Similarly to the best cities ranking, the results for the EPI show that for all the simulated instanciations, it is always possible to find a weights modification with the MILP that is lower than the WSI. A large majority of the cases only requires to change 2 or 3 criteria, as shown in Table 5, and the need to modify all the 9 criteria simultaneously occurs only rarely (0.38% of the cases). The distribution of the MILP scores ($z = \sum_{k=1}^{m} |w_k - w'_k|$) is given in Figure 5a. The total number of observations is 3588. The median is equal to 0.2066 which is 11.76% of the maximum score (1.7565).

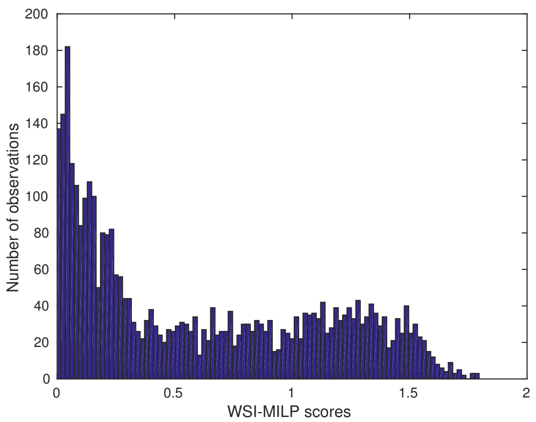
The distribution of the difference between the WSI and the MILP scores is given in Figure 5b. As we can observe, the values are all positive (MILP scores better than WSI) and the median is equal to 0.4626 which is 25.76% of the maximum difference (1.796). The computation time per iteration is 3.26 s on average. These values are larger than for the best cities ranking due to the higher number of criteria.

Table 5: EPI ranking - Distribution of the number of modified criteria

Number of modified criteria	2	3	4	5	6	7	8	9
Proportion of cases	50.58%	26.39%	10.61%	4.4%	2.2%	2.85%	2.59%	0.38%



(a) MILP scores



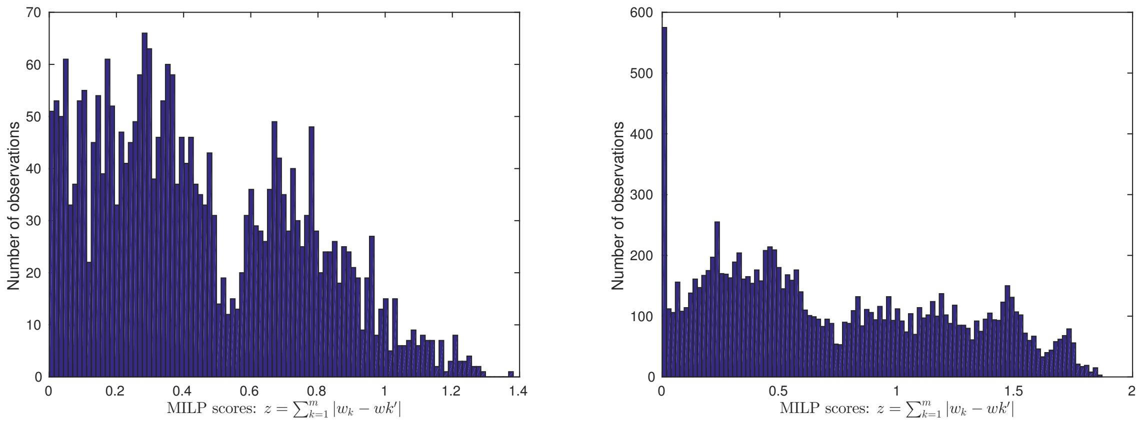
(b) Difference between the WSI and the MILP scores

Figure 5: EPI ranking - Distributions

5.4. Ranking any non-dominated alternative first

As stated in Section 4 only few alternatives can be ranked first when using the WSI. Since the problem we have formulated can be considered as an inverse optimization on the PROMETHEE II ranking, other (non-dominated) alternatives can also be considered for the stability analysis of the first-ranked. We have therefore tested if all the non-dominated alternatives could be ranked first. For the best cities ranking, it appears that less than 5% of these attempts were not possible. The distribution of the MILP scores for this case is shown in Figure 6a (2712 total observations). The median is equal to 0.3858 which is 27.93% of the maximum score (1.3813).

For the EPI ranking, 7.67% of the attempts of placing any non-dominated alternative first are not possible. The distribution of the MILP scores for this case is shown in Figure 6b (11426 total observations). The median is equal to 0.5983 which is 31.99% of the maximum score(1.8702).



(a) Best cities ranking

(b) EPI ranking

Figure 6: Distribution of MILP scores (any non-dominated alternative ranked first)

These results show that the stability of a ranking has also to take into account the other non-dominated alternatives, and not only the ones defined by the WSI, although the distance between the required weights and the initial set is significantly higher.

6. Future works and conclusion

In this contribution, we have proposed to analyzed the stability of a ranking, with respect to the criteria weights, using mixed integer linear programming. Compared to the traditional weights stability intervals, it appears that the MILP does not require to modify all the criteria. Indeed, with the two analyzed case studies, it appears that at most $m-1$ weights have to be changed while in most cases, only 2 or 3 criteria are to be altered.

In terms of distance from the initial set of weights, the results of the MILP also provide narrower intervals which may lead to the redefinition of the robustness of a ranking. Indeed, this means that a smaller change can already lead to the modification of the first-ranked.

As for the inverse optimization point of view, an interesting application could be to find a consensus when several decision makers are involved. Indeed, each party would express different preference models which will lead to different rankings. Using this inverse optimization approach could therefore be a way to analyze the minimum modification in the set of weights of a decision maker in order to get closer to the other parties, hence eventually finding a consensus, as proposed in [14].

In this work, we have only considered weights modification while other parameters are used for the preference modeling in PROMETHEE, such as the indifference and preference thresholds. However the problem does not seem to be linear anymore at first sight, so solving it would probably require other methodologies.

Finally, it is worth noting that the approach used in this study is not limited to PROMETHEE but could be extended to any multicriteria method that uses weighted sums (as long as it can be linearized).

References

[1] N. Doan, Multi-objective optimization and multi-criteria decision aid applied to the design of 3D-stacked integrated circuits, Ph.D. thesis, École polytechnique de Bruxelles (2015).

[2] J. Dyer, MAUT – multiattribute utility theory, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Verlag, Boston, Dordrecht, London, 2005, pp. 265–285.
URL http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-165-22-34954528-0,00.html

[3] T. Saaty, The Analytic Hierarchy and Analytic Network Processes for the Measurement of Intangible Criteria and for Decision-Making, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Verlag, Boston, Dordrecht, London, 2005, pp. 345–408.
URL http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-165-22-34954528-0,00.html

[4] J. Figueira, V. Mousseau, B. Roy, ELECTRE methods, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Verlag, Boston, Dordrecht, London, 2005, pp. 133–162.
URL http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-165-22-34954528-0,00.html

[5] C. Bana e Costa, J. De Corte, J. Vansnick, On the mathematical foundation of MACBETH, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer Verlag, Boston, Dordrecht, London, 2005, pp. 409–443.
URL http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-165-22-34954528-0,00.html

[6] J. Brans, P. Vincke, A preference ranking organization method, Management Science 31 (6) (1985) 647–656.

[7] M. Behzadian, R. Kazemzadeh, A. Albadvi, M. Aghdasi, PROMETHEE: A comprehensive literature review on methodologies and applications, European Journal of Operational Research 200 (1) (2010) 198–215.
URL http://EconPapers.repec.org/RePEc:eee:ejores:v:200:y:2010:i:1:p:198-215

[8] Q. Hayez, Y. De Smet, J. Bonney, D-Sight: A New Decision Making Software to Address Multi-Criteria Problems., IJDSST 4 (4) (2012) 1–23.
URL http://dblp.uni-trier.de/db/journals/ijdsst/ijdsst4.html#HayezSB12

[9] B. Mareschal, Weight stability intervals in multicriteria decision aid, European Journal of Operational Research 33 (1) (1988) 54 – 64. doi:http://dx.doi.org/10.1016/0377-2217(88)90254-8.
URL http://www.sciencedirect.com/science/article/pii/0377221788902548

[10] E. I. Unit, Best cities ranking and report: A special report from the economist intelligence unit, http://pages.eiu.com/rs/eiu2/images/EIU_BestCities.pdf (2012).
URL http://pages.eiu.com/rs/eiu2/images/EIU_BestCities.pdf

[11] A. Hsu, J. Emerson, M. Levy, A. de Sherbinin, L. Johnson, O. Malik, J. Schwartz, M. Jaiteh, The 2014 environmental performance index, http://www.epi.yale.edu/ (2014).
URL http://www.epi.yale.edu/

[12] J. Figueira, S. Greco, M. Ehrgott, Multiple criteria decision analysis: state of the art surveys, Vol. 78, Springer Verlag, 2005.
URL http://books.google.com/books?hl=en&lr=&id=YqmvlTiMNqYC&oi=fnd&pg=PR13&dq=Multiple+Criteria+Decision+Analysis:+State+of+the+Art+Surveys&ots=31ET2be92p&sig=NGeeSCP4_ixXDku8a3z_koMERhg

[13] B. Mareschal, J.-P. Brans, Geometrical representations for MCDA, European Journal of Operational Research 34 (1) (1988) 69–77.
URL http://ideas.repec.org/a/eee/ejores/v34y1988i1p69-77.html

[14] J. Roland, J. R. Figueira, Y. De Smet, Finding compromise solutions in project portfolio selection with multiple experts by inverse optimization, Computers & Operations Research 66 (2016) 12 – 19. doi:http://dx.doi.org/10.1016/j.cor.2015.07.006.
URL http://www.sciencedirect.com/science/article/pii/S0305054815001732