Unit: Multicriteria decision analysis
Learning goals

I. General classification of strategies in Multi-objective Optimization based on the time of decision maker interaction.
II. Ways to structure the decision making process
III. What are utility functions? How can we construct rankings?
IV. Pareto optima and related definitions
V. Interpretation and Visualization of Pareto fronts
Multicriteria Decision Analysis

Miettinen makes the following distinction between three general classes of multicriteria decision analysis approaches:

**A posteriori**: First compute set of non-dominated solutions using optimization, and afterwards make a decision by viewing solutions.

**A priori**: First design utility function and then find single optimum with respect to the utility function. The decision making is already finished before optimization.

**Progressive**: Interactive methods that combine a posteriori and a priori decision making and optimization in several feedback loops.

A priori multicriteria decision making
Questions in MCDA

**Given** A finite set of alternatives $x \in \mathcal{X}$, multiple objective functions $f_1(x), \ldots, f_m(x)$ (evaluation criteria).

**Wanted:** A (partial) ranking or a preferred choice of alternatives that is compatible with a decision maker's preferences.

The information on the criteria is typically not sufficient to establish full ranking (e.g. conflicting objectives),

Additional *preference information* is asked from the decision maker (DM): *Preference elicitation process*

For instance use questionnaires, or past actions and statements. Information: pairwise comparisons, weights, etc.
**Utility Function, Indifference Curves**

**Definition:** A utility function $U : \mathbb{R}^m \rightarrow \mathbb{R}$ is a mapping from objective function values to a single objective function value called utility.

**Definition:** An *indifference curve* is a set of points in the objective space, where the utility function has the same value.
Multi-Attribute Utility Theory (MAUT)

In MAUT several criteria (attributes) are aggregated to a single utility function value. Choose a functional form for the aggregation, e.g.:

**Additive Utility Function:**

$$U(f_1(x), \ldots, f_m(x)) = w_1f_1(x)\ldots + w_mf_m(x)$$

**Keeney and Raiffa Utility Function:**

$$U(f_1(x), \ldots, f_n(x)) = (1 + f_1(x))^{w_1} \cdot \ldots \cdot (1 + f_m(x))^{w_m}$$

Keeney and Raiffa utility has linear indifference curves in log-log plot.


Two ways to define parameters of utility function

(Robust) ordinal regression: The weights or parameters of the utility function are found based on statements of decision maker (DM) on pairwise comparisons or preferred choices (e.g., past actions). They should lead to a compatible and robust ranking of alternatives.

Guided specification of parameters, desirability functions The DM specifies the parameters utility function in a guided process, e.g., first focusing on single objectives and then on their relative importance.
### Ordinal regression (example)

<table>
<thead>
<tr>
<th></th>
<th>QTeaching max $f_1(x)$</th>
<th>QScience max $f_2(x)$</th>
<th>QCoffee max $f_3(x)$</th>
<th>QTechnology max $f_4(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delf TU</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Leid Univ.</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Mit Univ.</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Star Univ.</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

**Linear Rating:** $Score(x) = w_1 * f_1(x) + ... + w_4 * f_4(x)$, all weights $\geq 0$.

**Simple question:** Find weights for a utility function that ranks Star Univ. first?

**More difficult question:** The DM ranks $L \geq D$ and $S \geq M$. Is there a compatible utility function?
Ordinal regression (solution)

<table>
<thead>
<tr>
<th></th>
<th>QTeaching $f_1(x)$</th>
<th>QScience $f_2(x)$</th>
<th>QCoffee $f_3(x)$</th>
<th>QTechnology $f_4(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delf TU</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Leid Univ.</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Mit Univ.</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Star Univ.</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

Linear Rating: $Score(x) = w_1 * f_1(x) + \ldots + w_4 * f_4(x)$, all weights $\geq 0$.

$L \geq D$: $\iff 8w_1 + 8w_2 + 5w_3 + 7w_4 \geq 8w_1 + 6w_2 + 5w_3 + 8w_4$
$\iff 2w_2 - w_4 \geq 0$

$S \geq M$: $\iff 5w_1 + 5w_2 + 9w_3 + 5w_4 \geq 8w_1 + 7w_2 + 2w_3 + 9w_4$
$\iff -3w_1 - 2w_2 + 7w_3 - 4w_4 \geq 0$

One solution is $w_1 = 0, w_2 = 0.1, w_3 = 0.7, w_4 = 0.1$
Ordinal regression

For a set of pairs $(x_1^l, x_2^l) \in \mathcal{X} \times \mathcal{X}, \ l = 1, \ldots, q$ the DM specified $x_1^l$ better or equal to $x_2^l$. We want to find compatible $w_i, \ i = 1, \ldots, m$ such that:

$$\sum w_i f_i(x_1^l) \leq \sum w_i f_i(x_2^l), \ l = 1, \ldots, q$$

$$\sum w_i = 1, \ w_i \geq 0, \ i = 1, \ldots, m$$

A robust solution maximizes the margins to the constraint boundaries:

$$\max \delta$$

$$\sum w_i f_i(x_1^l) + \delta \leq \sum w_i f_i(x_2^l), \ l = 1, \ldots, q$$

$$\sum w_i = 1, \ w_i \geq 0, \ i = 1, \ldots, m$$

Robust ordinal regression typically uses piecewise linear utility functions, instead of linear ones.
Further Reading


Multi-Attribute Utility Theory (MAUT)

In order rank alternatives, the decision analyst can follow these steps:

1. Determine all criteria and constraints (informal)

2. Determine decision space $\mathcal{X}$ (set of possible decisions)

3. Determine quantitative measures for criteria

4. Measure criteria values $f_1(x),...,f_m(x)$ for $x \in \mathcal{X}$

5. Normalized scale, e.g. between 0 and 1

6. Weight objectives by importance $w_1,...,w_m$

7. Compute weighted utility function $U(f_1(x),...,f_m(x),w_1,...,w_m)$ for each solution; Rank solution in $\mathcal{X}$ based on $U$.

Client-Theory by Kahneman and Tversky


How good are people feeling when winning a lottery?

Difference between 0 and 1000 appears bigger than difference between 100000 and 101000.

Bernoulli (1738): ’The psychological response to the change of wealth is inversely proportional to the initial amount of wealth’ Degressive utility function, e.g., \( \log(f(x)) \) instead of linear one.

Kahnemann and Tversky: Loss is higher weighted than win by decision makers (prospect theory).

⇒ The initial wealth matters. (cannot be modeled by utility function)
Desirability functions (DFs)

Creating a nonlinear utility by using DFs:

Step 1: Define desirability functions for each objective function value transformation

Step 2: Weight objective functions

Step 3: Aggregate multiple weighted objective functions

\[ w_1 = 7 \]
\[ w_2 = 8 \]
\[ w_3 = 10 \]

\[ U(f(x)) = \alpha \frac{1}{m} \sum_{i=1}^{m} w_i v_i(f_i(x)) + \beta \min_{i \in \{1, \ldots, m\}} w_i v_i(f_i(x)), \]

common interest

(minority interest)

(Here: \( m = 3 \))

\[ s.t. \ v_i(f_i(x)) > 0, i = 1, \ldots, m \]

Desirability functions (DFs)

Harrington vs. Derringer Suich type of DFs

Standardized curves controlled by few parameters.

**Harrington type:** never equal 0 or 1.  
**Derringer-Suich type:** 0 if \( \leq T_i \), and 1 if \( \geq USL_i \)

Value pairs \((Y_i^{(1)}, d_i^{(1)}), (Y_i^{(2)}, d_i^{(2)})\)

\[
d_i(Y_i^{(j)}) = \exp(-\exp(-Y_i^{(j)}))
\]

\[
Y_i^{(j)} = b_{0i} + b_{1i}Y_i^{(j)}, \quad j = 1, 2.
\]

Desirability index:  
\[
D(Y_1, \ldots, Y_m) = d_1(Y_1)^{\alpha_1} \cdots d_m(Y_m)^{\alpha_m}
\]
Harrington vs. Derringer Suich type of DFs (2)

Harrington desirability functions never reach 0.

$T_i$ is the target value; in case of Harrington it is 0, $n_i$ and $r_i$ control curvature (smoothness).

\[
d_i(Y_i^{'}) = \exp(-|Y_i'|^{n_i}) \\
Y_i' = \frac{2Y_i - (USL_i + LSL_i)}{|USL_i - LSL_i|} \\
\]

Desirability index: $D(Y_1, \ldots, Y_m) = d_1(Y_1)^{\alpha_1} \cdots d_m(Y_m)^{\alpha_m}$
A posteriori multicriteria decision making
Multicriteria Problems: Car Example

Your objectives: Cost $\rightarrow$ min, Speed $\rightarrow$ max

Add constraint: Only red cars!
Decision vs. Objective Space, pre-images

- **Def.** The space of candidate solutions is called the decision space, for instance $X = \{Beetle, CV2, Ferrari\}$ in the figure.

- **Def.** The space of objective function values (vectors) is called the objective space, for instance $\mathbb{R}^2$ in the figure.

- **Def.** For a point in the objective space the corresponding point(s) in the decision space are called their pre-image(s). In the figure, Beetle and CV2 are pre-images of the green point.

- **Remark:** Two different points in the decision space (e.g. Beetle and CV2) can map to the same point in the objective space, but two points in the objective space have never the same preimage in the decision space.
Incomparable and indifferent solutions

- **Def.:** Given two solutions and some criterion functions, the solutions are said to be **incomparable**, if and only if
  1. the first solution is better than the second solution in one or more criterion function value and
  2. the second solution is better than the first alternative in one or more other criterion function values.

- **Def.:** Given two alternatives and some criterion functions, the alternatives are said to be **indifferent** with respect to each other, if and only if they share exactly the same criterion function values.

**Remark:** In both cases additional preference information is required to decide which solution is best.
Pareto dominance

- **Def.**: Given two alternative solutions and some objective functions, the first solution is said to Pareto dominate the second solution, if and only if
  1. the first solution is better or equal in all objective function values, and
  2. the first solution is better in at least one objective function value.

Francis Y. Edgeworth
Irish Economist
1845-1926

Vilfredo Pareto
Italian Economist
1848-1923
Given: A decision space $X$ comprising all (feasible) decision alternatives, a number of criterion functions $f_i : X \rightarrow \mathbb{R}$, $i = 1, \ldots, m$

Def.: A decision alternative $x$ in $X$ dominates a solution $x'$ in $X$, iff it is not worse in each objective function value, and better in at least one objective function value.

Def.: If a solution $x \in X$ is not dominated by any other solution in $X$, then it is called Edgeworth-Pareto optimal (or Pareto optimal) (in $X$).

Def.: The set of all Pareto optimal solutions in $X$ is called efficient set.

Def.: The set of all Pareto optimal function vectors of solutions in the efficient set is called the Pareto front.
Fundamental Concepts: Dominance diagram

- Dominating Subspace
- Dominated Subspace
- Space of incomparable solutions
- Reference Solution

What about the points on the dark red lines?
How would this diagram look for maximization of $f_1$ and minimization of $f_2$?
Construction of Pareto front in 2-D

Geometrical construction for separating dominated and non-dominated points in 2-D:

1. Indicate for each point the dominated subspace by shading
2. The covered subspace consists of dominated points within the set, that is points that are dominated by at least one other point
3. The outer corner points on the lower left boundary form the Pareto front of the point set.
For $f=(f_1, ..., f_m)$ the **image set** $f(X)$ is defined as the set of all $(f_1(x), ..., f_m(x))$ for $x \in X$.

The non-dominated solutions in $f(X)$ are located at the lower left boundary and form the Pareto front.

Note, that for unbounded or non-closed sets $f(X)$ the Pareto front does not always exist.

If it exists, the Pareto front of a $m$-objective problem has at most $m-1$ dimensions.

The set of all preimages of points in the Pareto front is the efficient set.
Interpreting and visualizing Pareto fronts
Thinking about trade-off, knee points

Moving from $x_2$ to $x_1$ is an unbalanced tradeoff.

Solutions (knee point region)

Moving from $x_1$ to $x_2$ is a balanced trade-off.

$x_1$ and $x_4$ are ideal solutions

Region of good compromise

Pareto front

Image under $f$
Innovization – Design principles from multiobjective optimization

- By looking at changing designs across the Pareto front, designers can study design principle
- How does the design change when moving across the Pareto front?
- Innovization: Finding design principles by multicriteria optimization

Deb et al.: "Innovization: Innovating design principle through optimization" GECCO. ACM, 2006

Construction of Pareto front for 3-D point set

- A single in the objective space point dominates a 3-D cuboid
  - the point is the upper corner
  - the lower corner is \((-\infty,-\infty,-\infty)^T\)
- Dominated cuboids can be drawn in a perspectivic plot.
- The boundary between dominated and non-dominated space is called attainment surface
- The points that belong to the Pareto front are located at its outer corners.
Figure: Approximation to 3-D Pareto Front
Here, minimization is the goal.
Example Pareto front: embedded systems design

Figure: 4-D Pareto Front visualization. All functions to be maximized. VisuWei tool. http://natcomp.liacs.nl
Parallel Coordinates Diagrams (PCDs) are a common way to visualize solutions with many (typically > 3) attributes (multivariate data)

1. Each criterion is represented by a vertical axis.
2. A solution is represented by means of a polyline connecting points (criteria values) of that solution.

Normalize the range of criteria, that is minimum, maximum of criterion in data set determine scale of its axis.

Choose sign of the criteria can be chosen such that all criteria are to be minimized.

Example: Notebook comparison with three notebooks
N-Dimensions: Parallel Coordinates Diagram

Example: Comparing notebooks

Find pairs where A dominates B?
How to determine indifference and incomparability?
Use definition of Pareto dominance!
Brushing is an interactive tool that can be used to explore sensitivity to constraints.
(1) Indicate ranges by gray area using, for instance, a computer mouse.
(2) Highlighted solutions (polylines) are solutions that fall into the specified ranges.

In the freeware tool XMDV the Parallel Coordinates diagram allows brushing, see example (gray area indicates selected ranges).

Example (left): Filtering cars from Europe and in certain ranges of the continuous output variables.
Star plots are a similar idea to visualize multivariate data.

1. Axis originate from the same point and are spread in an equi-angular way.
2. The scaling of the axis is from minimum to maximum.
3. Solutions are represented by closed polygons.

Remarks:
- Shapes are easier perceived and remembered (so called Star-glyphs).
- Brushing, selection and grouping tools more difficult
Scatterplot Matrix - Example

- The scatter plot matrix consists of scatter plots for all values of $f_i, f_j$ with $i < j$ and $i \in \{1, \ldots, m\}$, $j=\{1, \ldots, m\}$
- Only the upper diagonal matrix needs to be depicted
- Points on the diagonal are not interesting, but often the diagonal is used to plot total frequency of values
- Scatterplots serves to investigate the correlation between two objectives functions

<table>
<thead>
<tr>
<th>X</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$f_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>x3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Multivariate Data Visualization (MDV)

- MDV tools
  - XMDV
  - Ggobi
- Spreadsheet tools
  - e.g. MS Excel, Open Office
- Scientific visualization and programming tools
  - MATLAB/Octave
  - SciLab
  - Python/Matplotlib
  - R: Parallel Coordinates

https://www.safaribooksonline.com/blog/2014/03/31/mastering-parallel-coordinate-charts-r/
Scatter plot matrix

- Criteria scatter plot matrix shows all pairwise scatter plots.
- (Anti)correlation between criteria is revealed in this plot:
  - Positive slope $\rightarrow$ Criteria are positively correlated plot
  - Negative slope $\rightarrow$ Criteria are anticorrelated (conflicting)
Interactive decision making tools work with questions to the decision maker, typically s/he is asked to do pairwise comparisons“

Preference elicitation: Deducting a utility function from comparisons and questionnaire data

- **Prometheus**: Construction of a utility function from questionnaires
- **ELECTRE**: Partial orders; resolve inconsistencies in ranking
- **AHP**: Pairwise comparison, hierarchy of criteria => ranking
- **NIMBUS**: Interactive industrial optimization and decision making tool developed by the Finnish group of Miettinen
- **Robust Ordinal Regression**: Constructs possibility space of utility function that are consistent with pairwise comparisons
- **NEMO**: Combines robust ordinal regression with heuristic optimization
Summary: Take home messages

1. In multiobjective optimization a priori, a posteriori, and progressive methods are distinguished, depending on when the DM interacts.
2. Multicriteria decision analysis structures decision process
3. Utility functions capture user preferences: Linear weighting, Keeney Raiffa, Derringer-Suich type and Harrington Desirability functions.
4. In multicriteria optimization and decision analysis, two solutions can be incomparable, indifferent, or one solution dominates the other.
5. Pareto dominance and (Edgeworth-)Pareto optimality characterizes Pareto optima in multiobjective optimization.
6. Pareto fronts can be used to reason about trade-offs and also to find new design principles. They reveal the nature of the conflict(s).
7. Pareto fronts can be interpreted as trade-off curves (2-D) or as a trade-off surfaces in 3-D. For >= 4-D: Parallel coordinates, Radar plots, Scatter plots.