Many objective optimization and scalarization

Many-objective optimization

Fleming P.J., Purshouse R.C., Lygoe R.J. (2005) Many-Objectiv
Optimization: An Engineering Design Perspective. In: Coello
Coello C.A., et al. EMO 2005. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-540-31880-4_2

- In more than 3 dimensions it is a convention to replace the term **Multi-objective Optimization** by the term **Many-objective Optimization**
- **Why is this regarded as a separate field?**
  - Other means of visualizing the Pareto front
  - More solutions tend to get non-dominated. Why?
  - Gedankenexperiment (left)
  - Articulation of preferences becomes more important

Gedankenexperiment

Let S denote a search space. And $f_1: S \to \mathbb{R}$ ... $f_n: S \to \mathbb{R}$denote n objective functions, and S_E denote the efficient set.

What happens to the efficient set if we add another objective $f_{n+1}: S \to \mathbb{R}$ (constraint)?

Please state
your answer
mentimeter!
We will discuss
Answer after lunchbreak

https://www.menti.com/svqfj4a3ih

Discussion

set if we add another objective $f_{n+1}: S \rightarrow \mathbb{R}$?

In case that we have all get indifferent solutions, one might dominate the others in the new objective and the set decreases.

If A dominates B and B is better in the new solution then A and B become incomparable, and B might enter the efficient set: the set increases. (more common case!)

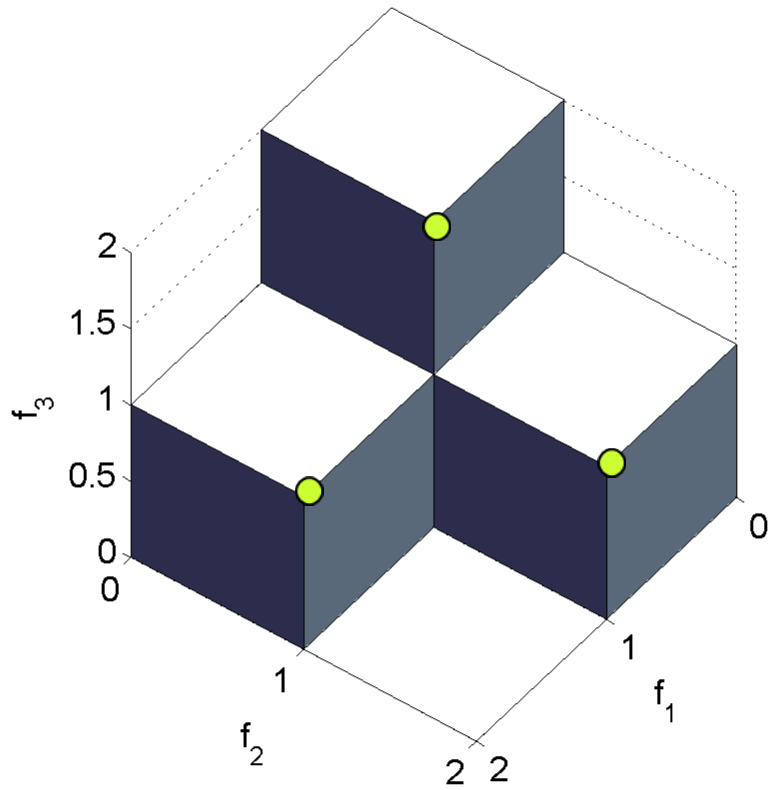
set if we add another constraint $g_{n+1}: S \rightarrow \mathbb{R}$?

If the constraint discards a solution that before dominated many other solutions the set might increase

If the constraint discards a non-dominated solution that does not dominate another solution, the set decreases.

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.



Universiteit Leiden

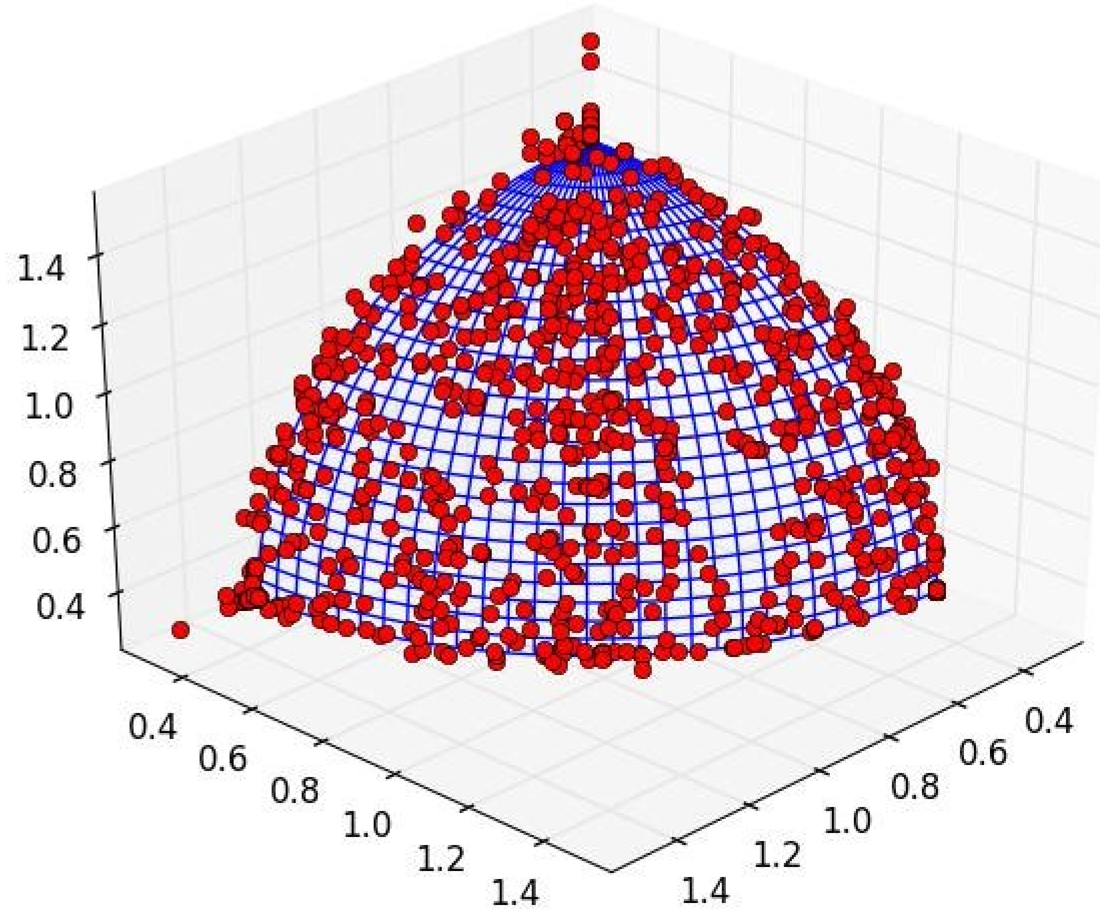
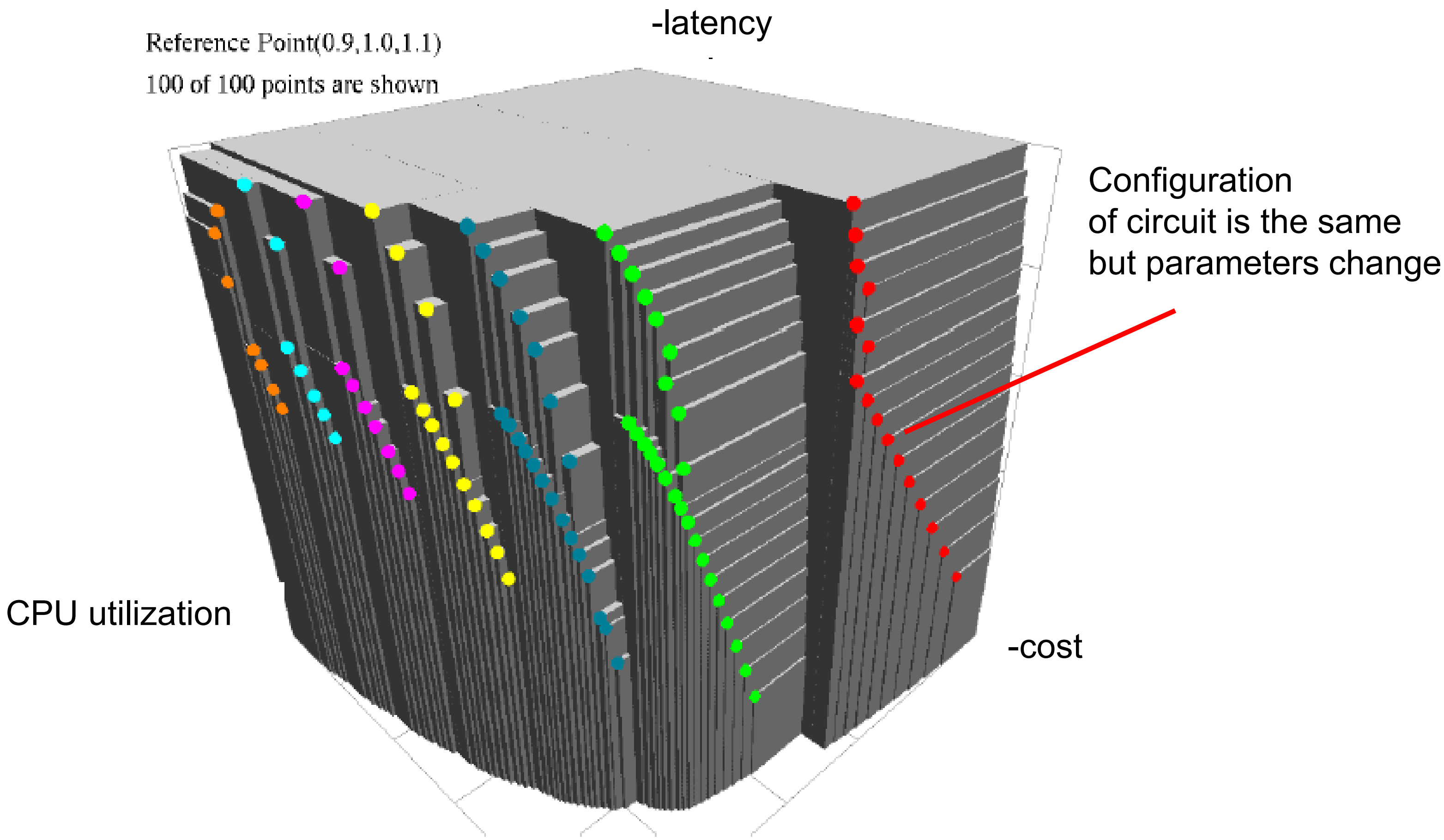


Figure: Approximation to 3-D Pareto Front
Here, minimization is the goal.

Example Pareto front: embedded systems design



Li, R., Etemaadi, R., Emmerich, M. T., & Chaudron, M. R. (2011, June). An evolutionary multiobjective optimization approach to component-based software architecture design. In *Evolutionary Computation (CEC), 2011 IEEE Congress on* (pp. 432-439). IEEE.

Universiteit Leiden

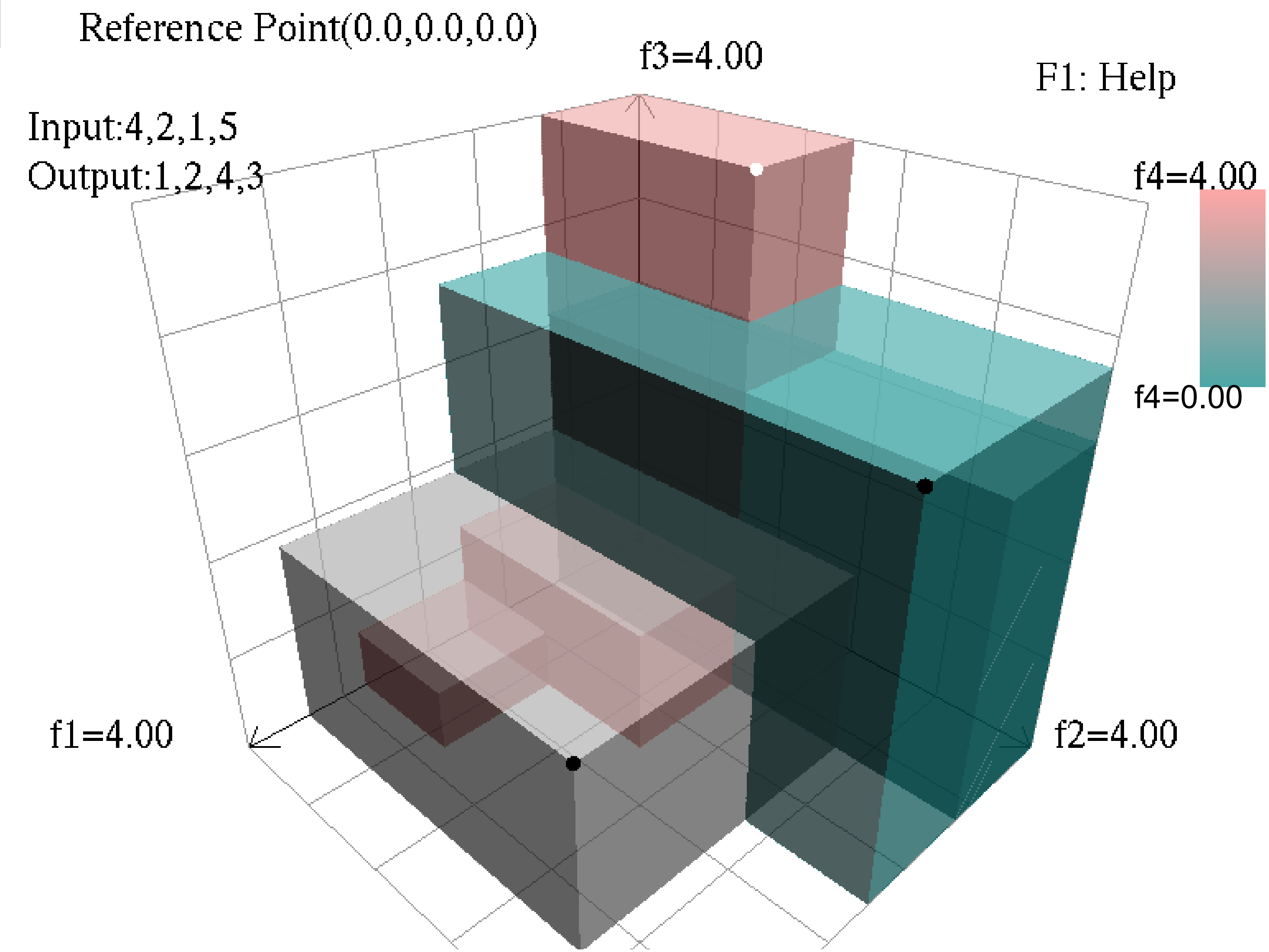
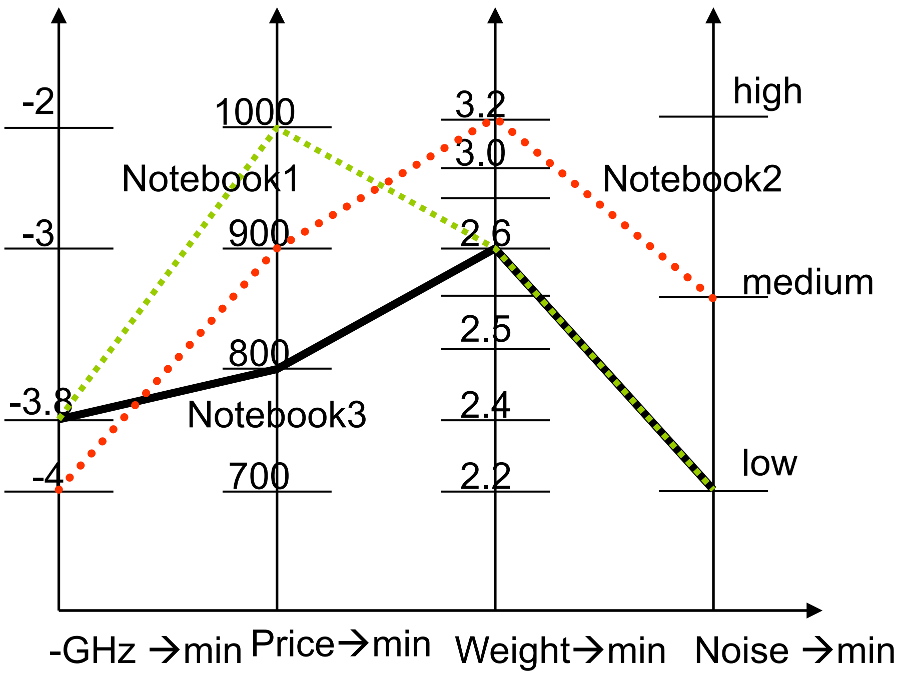


Figure: 4-D Pareto Front visualization.
All functions to be maximized. VisuWei tool.
http://natcomp.liacs.nl

N-Dimensions: Parallel Coordinates Diagram



Example: Notebook comparison with three notebooks

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.

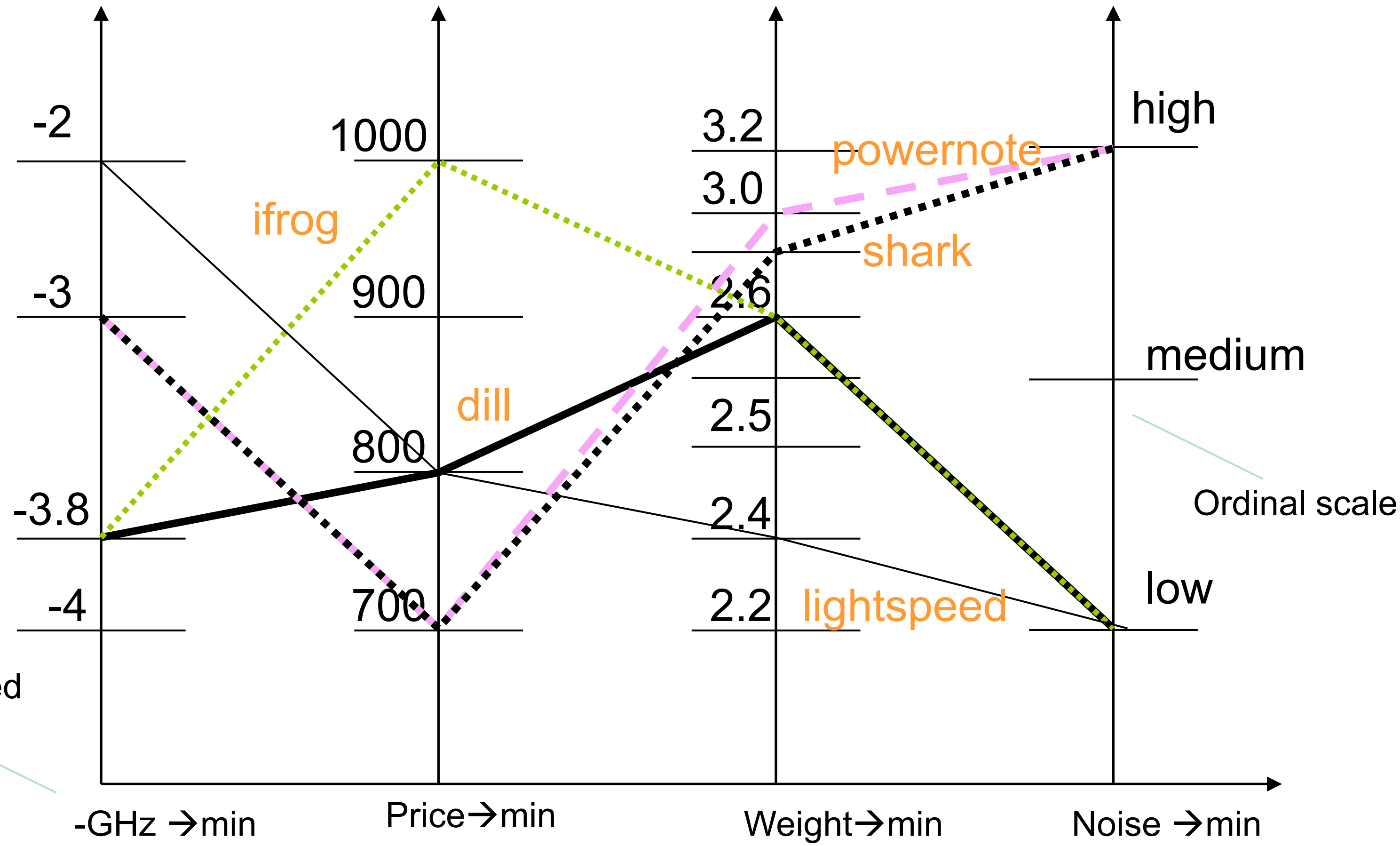
(2) A solution is represented by means of a **polyline** connecting points (criteria values) of that solutions.

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.

N-Dimensions: Parallel Coordinates Diagram

Example:
Comparing
notebooks



Normalize:
All to be minimized

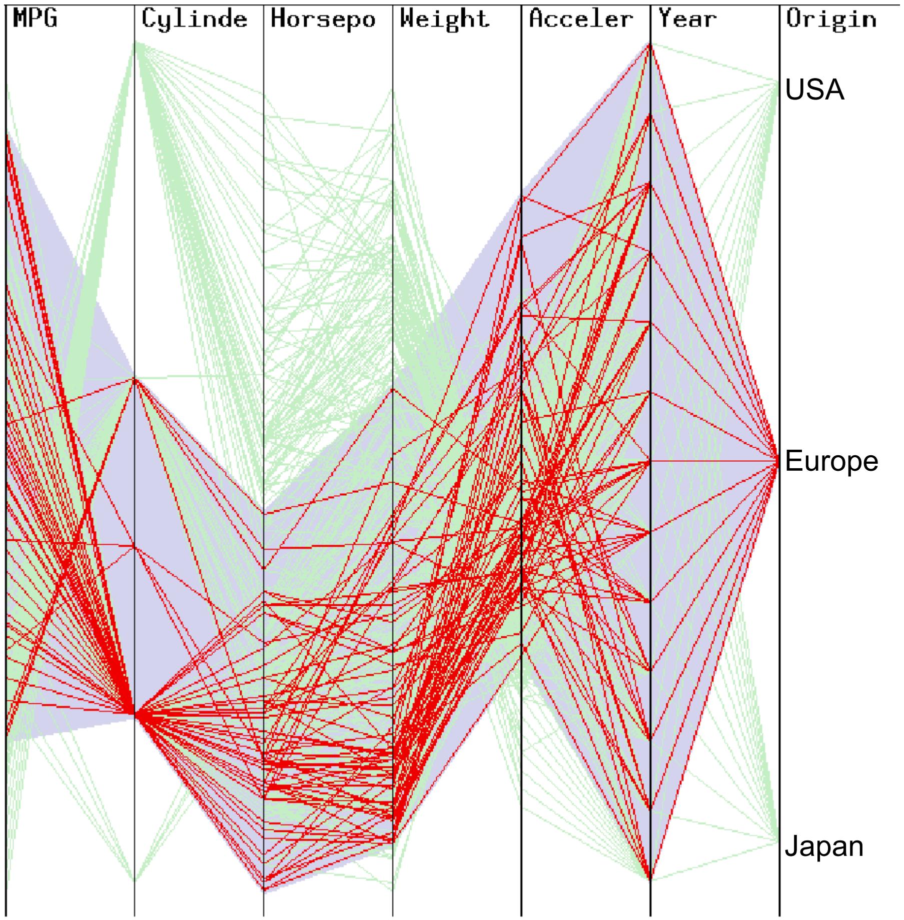
Find pairs where A dominates B?

How to determine indifference and incomparability?

Use definition of Pareto dominance!

Universiteit Leiden

Interactive Parallel Coordinates - Brushing tool



Brushing tool of XMDV Parallel Coordinates
Example: Car data.

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.

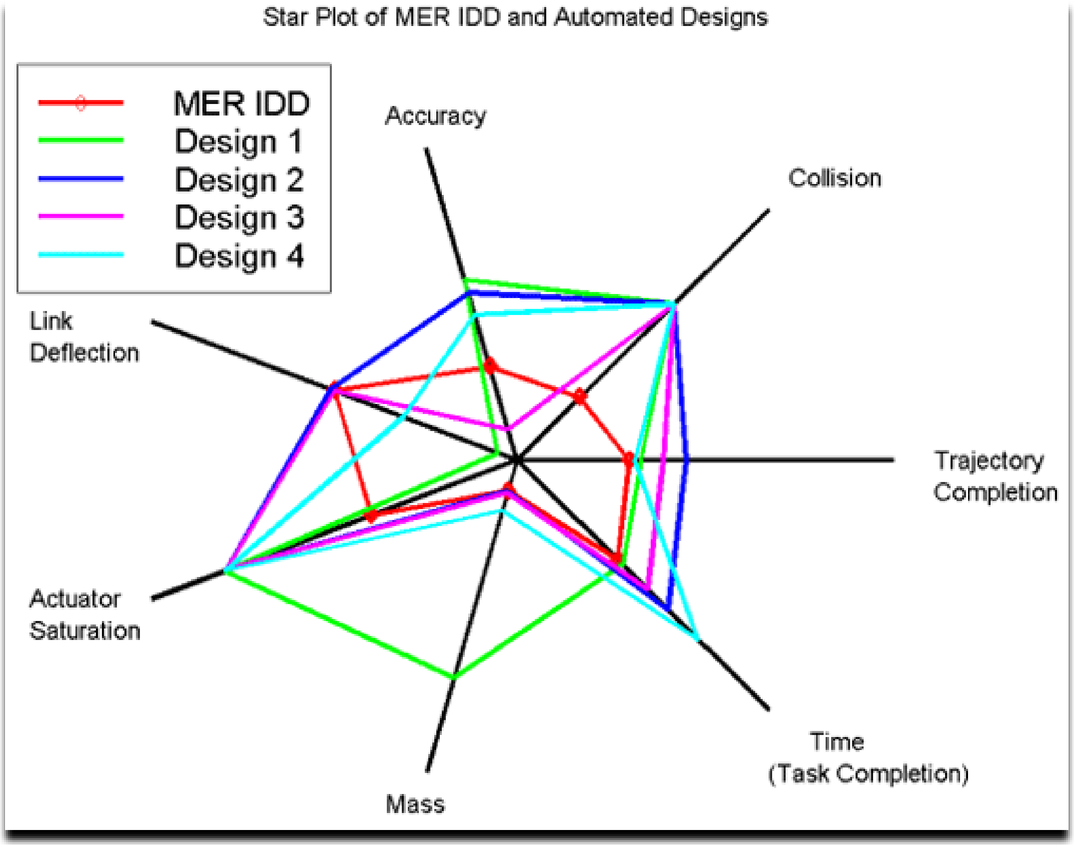
Star Plots or Radar Plots

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.



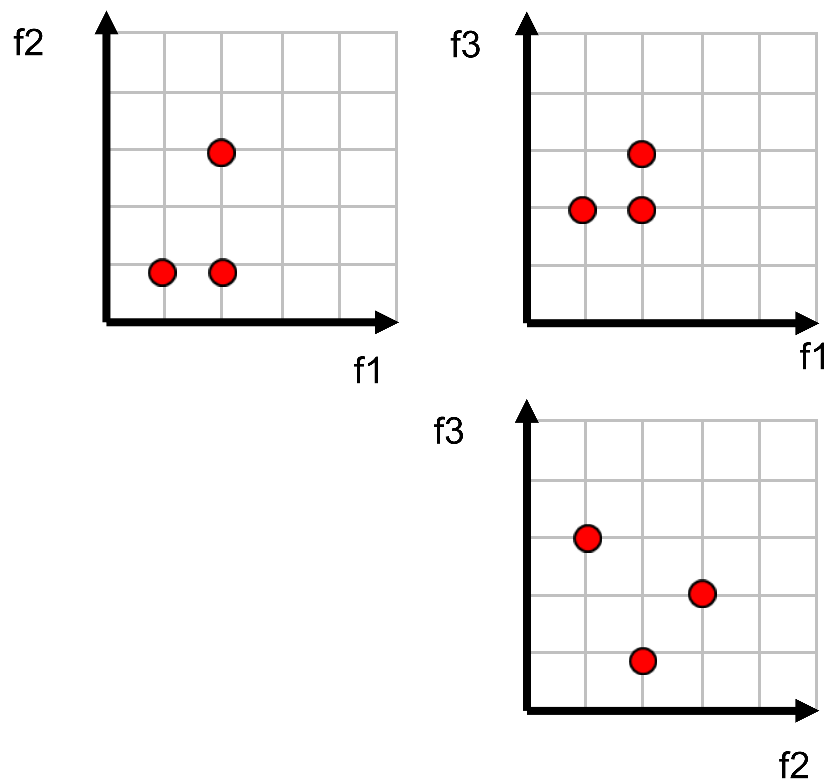
Remarks:

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

Scatterplot Matrix - Example

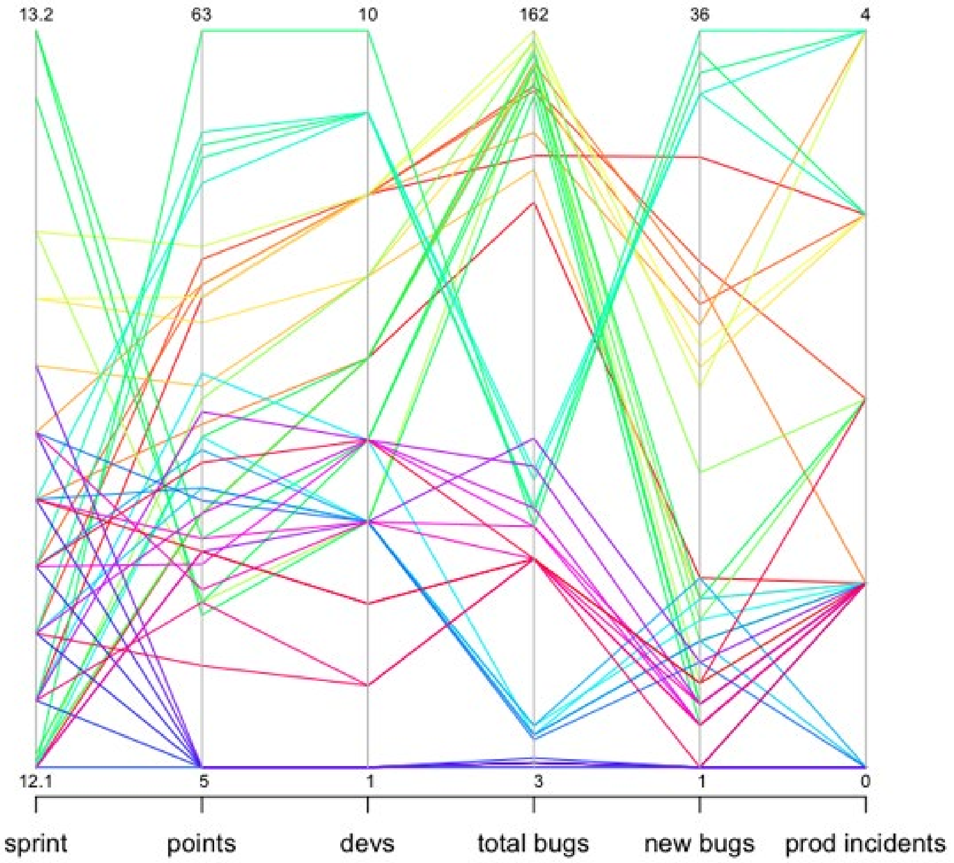
- The scatter plot matrix consists of scatter plots for all values of f_i, f_j with $i < j$ and $i \in \{1, \ldots, m\}$, $j = \{1, \ldots, m\}$
- Only the upper diagonal matrix needs to be depicted
- Points on the diagonal are not interesting, but often the diagonal is used to plot total frequency of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

X	f_1	f_2	f_3
x1	2	3	2
x2	1	1	2
x3	2	1	3



Multivariate Data Visualization (MDV)

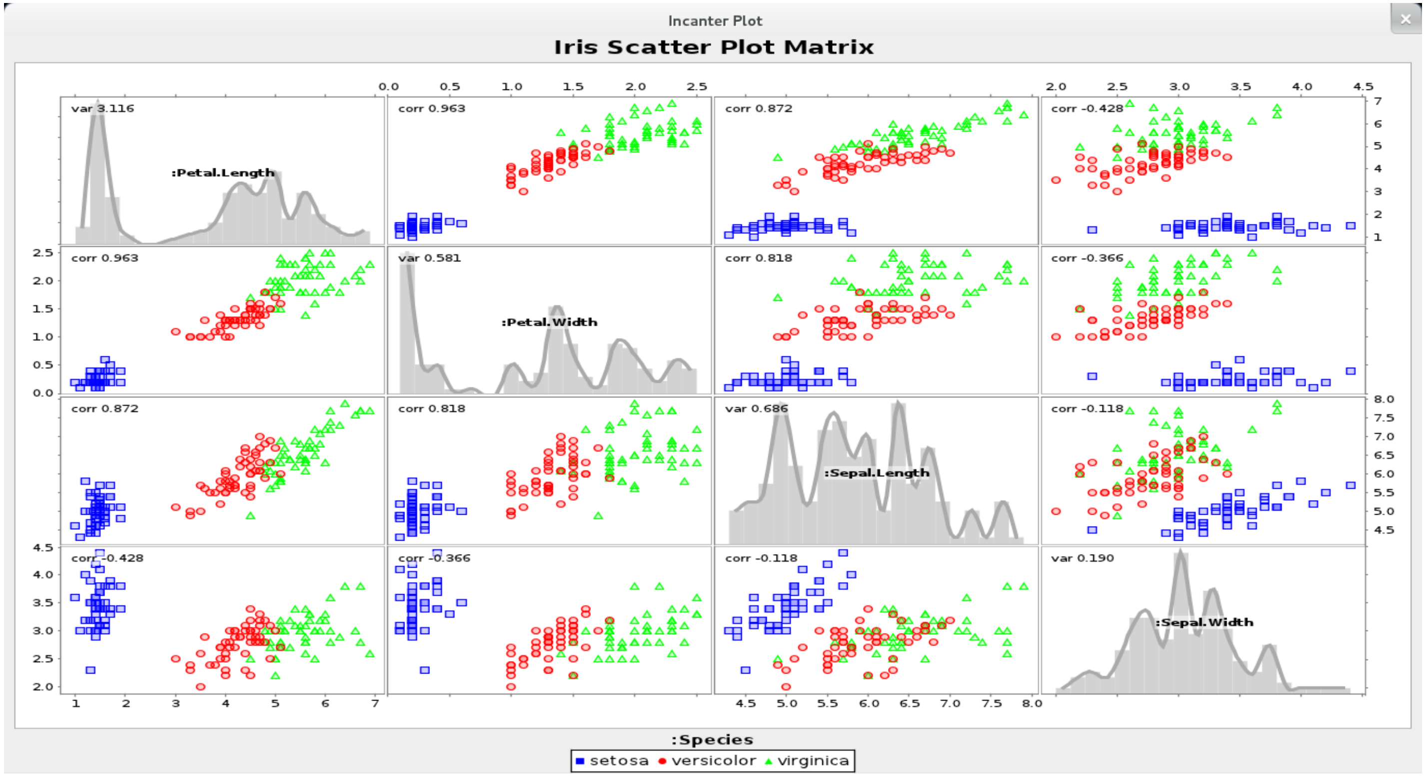
- MDV tools
  - XMDV
  - Ggobi
- Spreadsheet tools
  - e.g.MS Excel, Open Office
- Scientific visualization and programming tools
  - MATLAB/Octave
  - SciLab
  - Python/Matplotlib
  - R: Parallel Coordinates



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 → Criteria are positively correlated plot
- Negative slope → Criteria are anticorrelated (conflicting)



Universiteit Leiden

*Progressive and Interactive MCDA Methods

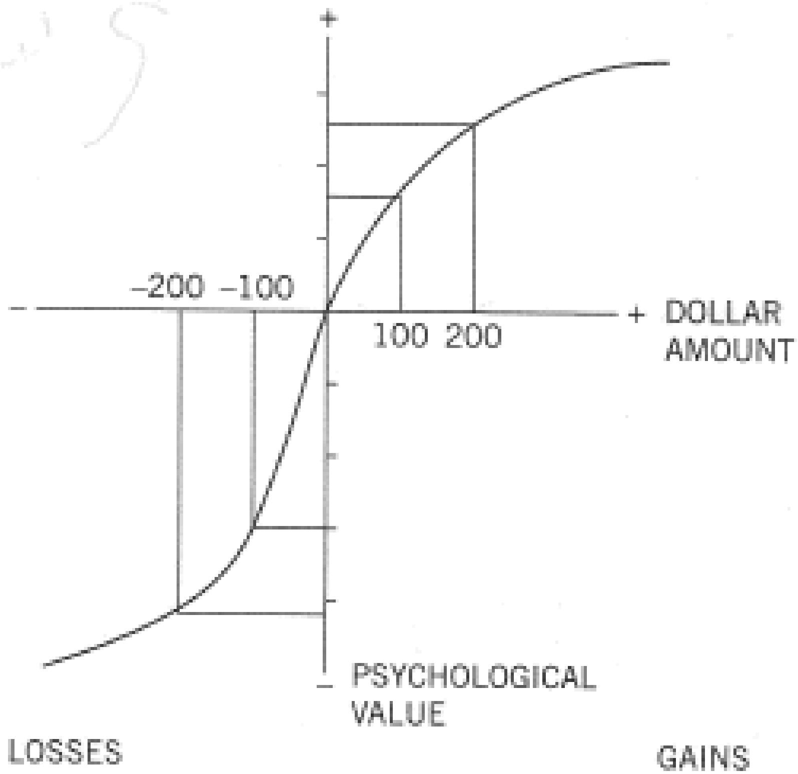
Interactive decision making tools work with questions to the decision maker, typically s/he is asked to do pairwise comparisons“

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

Universiteit Leiden

Client-Theory by Kahneman and Tversky



D. Kahneman: ‘Thinking fast and slow’
Penguin press. 2014.

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).

$\Rightarrow$ The initial wealth matters. (cannot be modeled by utility function)

Nautilus Navigator

- Start from a solution near Nadir point (all objectives have bad values)
- Then a preferred improvement direction is chosen (how much which objective is improved)
- The navigator improves objectives in small steps, until
  - It reaches the Pareto front
    OR
  - The user changes the preferred direction or reference point
- Other options: backtracking, resampling

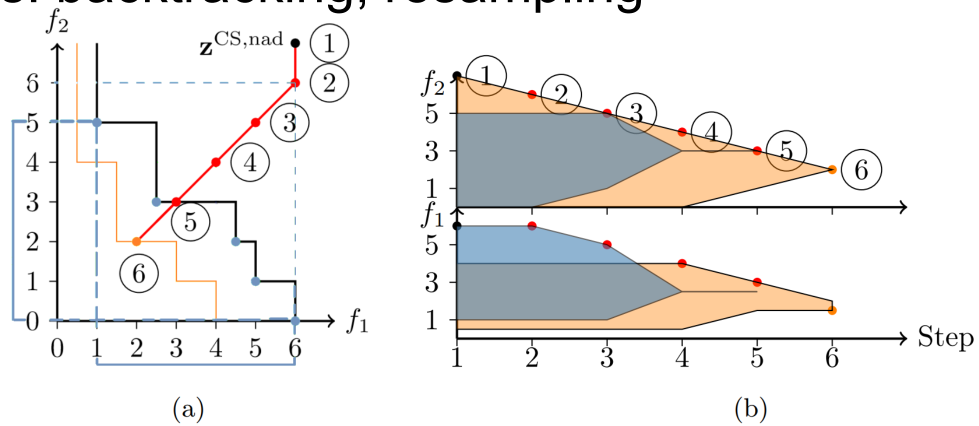
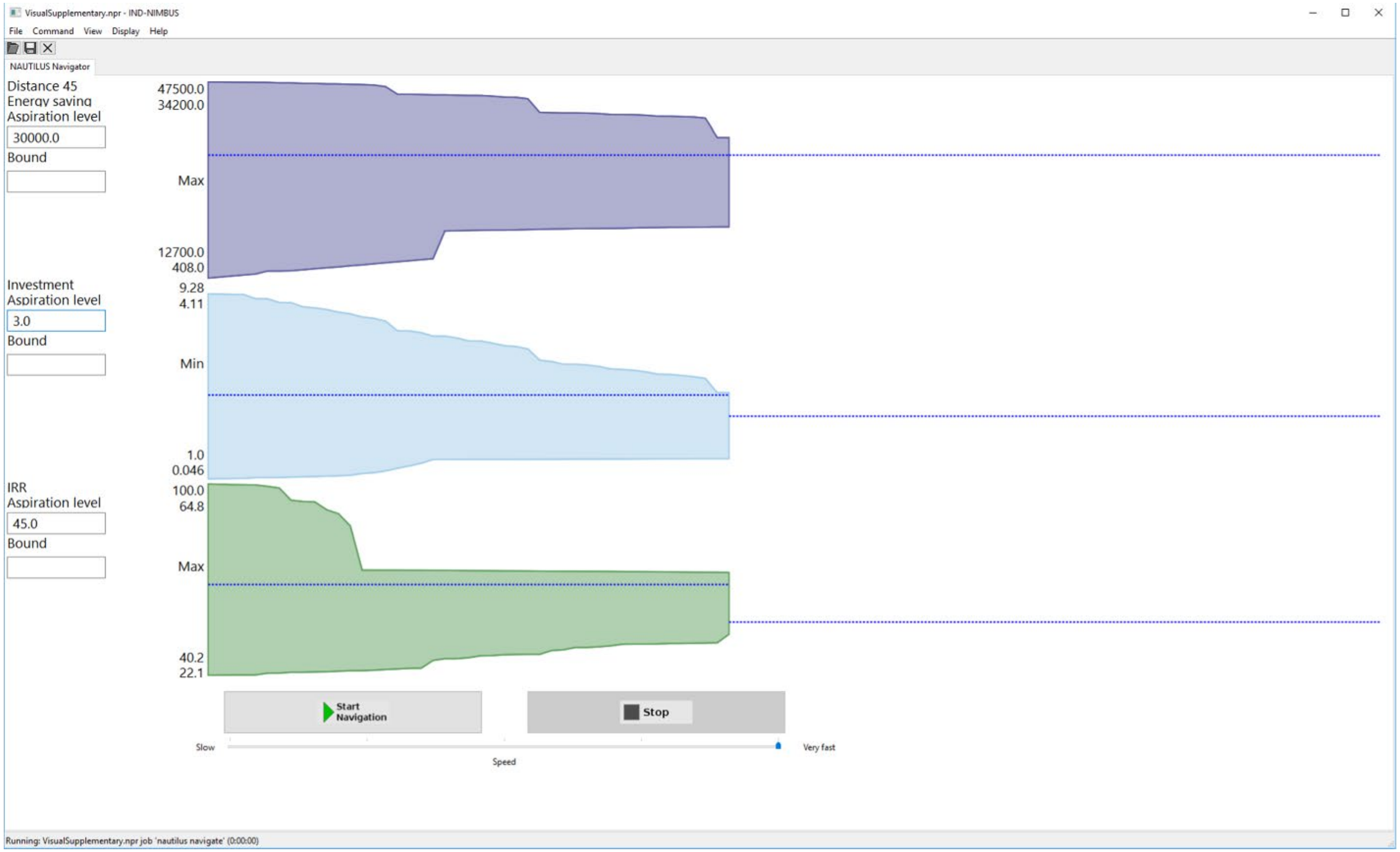


Figure 5: Path of O-NAUTILUS visualized as a 2-D scatter plot (a) and as reachable ranges plot (b).

Universiteit Leiden

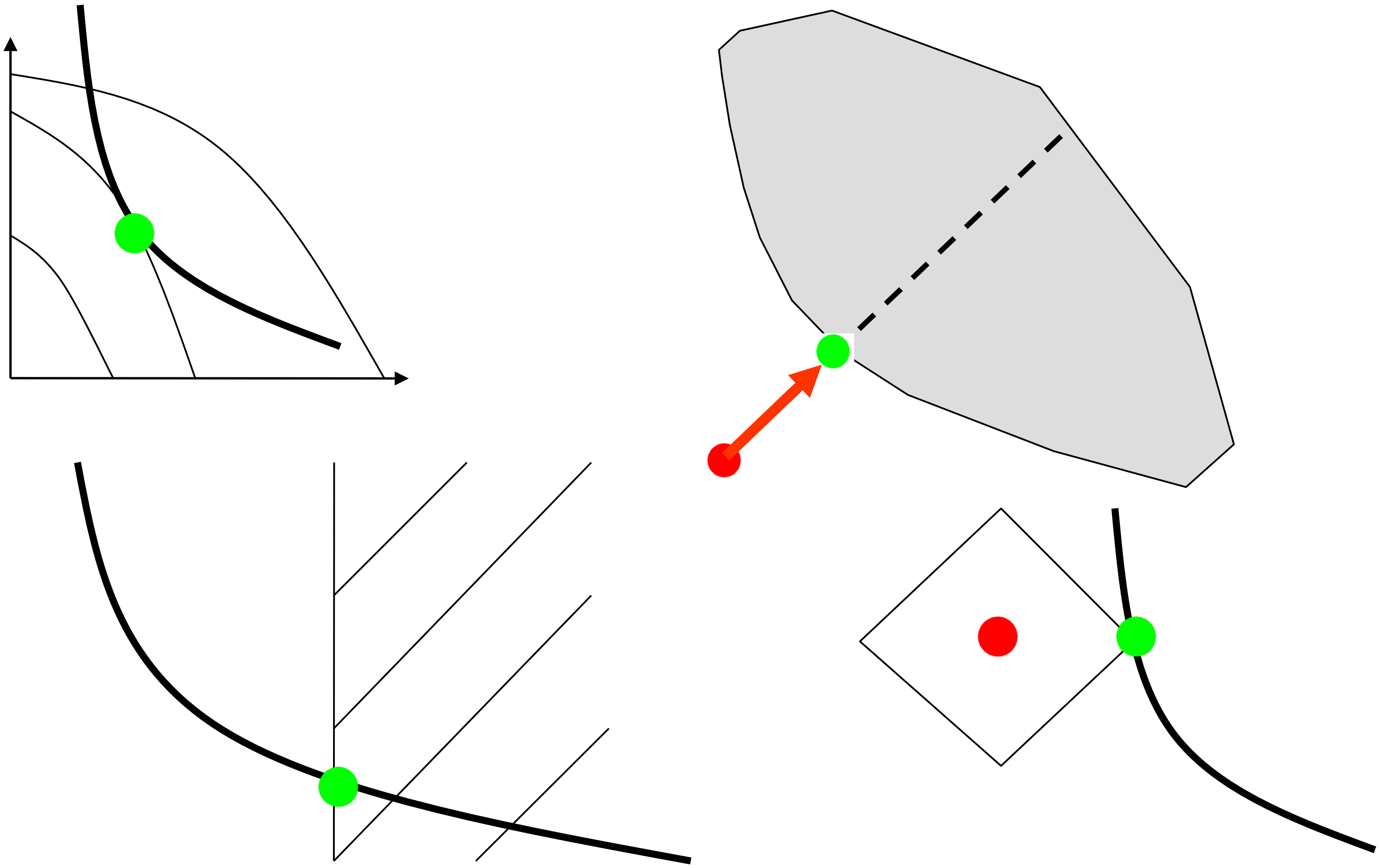
Nautilus Navigator



https://www.youtube.com/watch?v=gjvIG8PiPBo

Universiteit Leiden

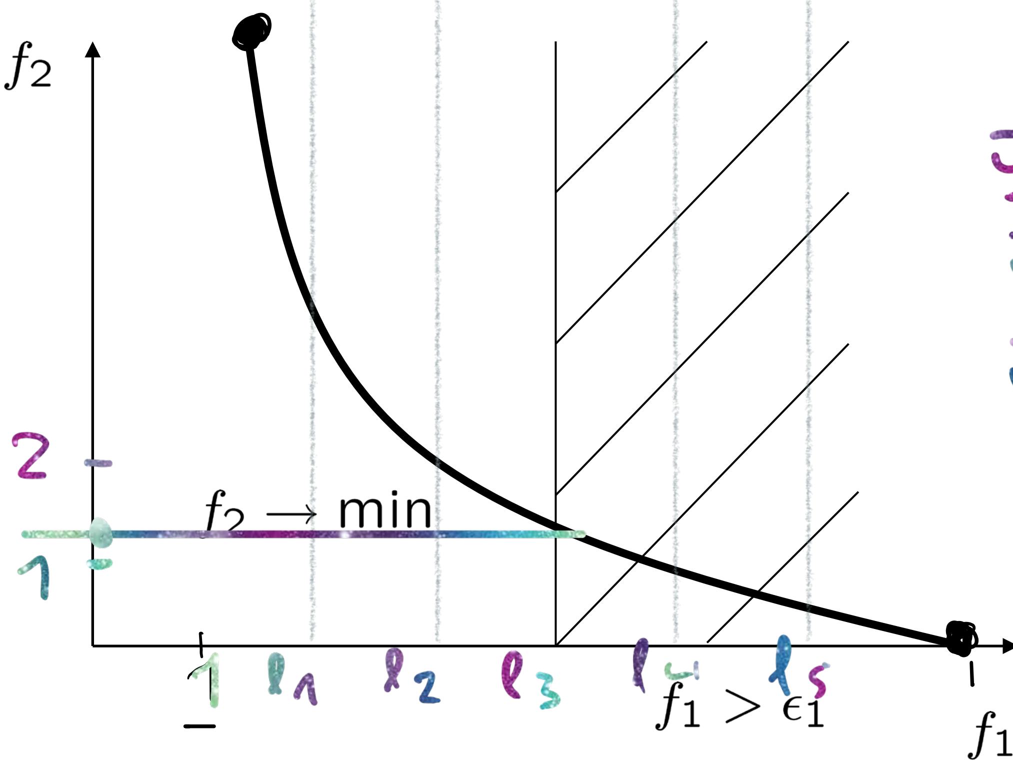
Scalarization methods for finding the Pareto front



Universiteit Leiden

ε-Constraint method

$$f_m(\mathbf{x}) \rightarrow \min, \text{ s.t. } f_i(\mathbf{x}) \leq \epsilon_i^{(j)}, i = 1, \ldots, m, \quad j = 1, \ldots, n_{levels}$$



-2

	$f_1 \leq \varepsilon_i^{(j)}$	$f_2 \rightarrow \min$
$\vec{y}^{(1)}$	l_1	
$\vec{y}^{(2)}$	l_2	
$\vec{y}^{(3)}$	l_3	1.2
$\vec{y}^{(4)}$	l_4	
$y^{(5)}$	l_5	

With the dimension the number of ϵ combinations grows exponentially.

But it is also possible to use the epsilon-Constraint method interactively: change Constraints manually and see what happens to the objective function(s)

Universiteit Leiden

Learning goals

- What are different ways to solve multiobjective optimization problems by formulating them as single objective optimization problems (with constraints)?
- Can we use linear weighting functions to find all Pareto optimal points?
- Which points (on the Pareto front) do we find for different scalarization functions?
- How and when can we use single point methods to find all points on a Pareto front?

Universiteit Leiden

Weighted sum scalarization

$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} w_i f_i(\mathbf{x}) \quad w_i > 0, i = 1, \dots, m \qquad \sum_{i=1}^{m} w_i = 1$

Regardless of the choice of weights, the weighted sum minimization will always result in an efficient solution $\mathbf{x}^*$

Proof: $f_i(\mathbf{x}) \leq f_i(\mathbf{x}')$ for $i = 1, \dots, m$ and $f_j(\mathbf{x}) < f_j(\mathbf{x}')$ for some $j \in \{1, \dots, m\} \Rightarrow \sum w_i f_i(\mathbf{x}) < \sum w_i f_i(\mathbf{x}')$

Not all Pareto optimal solutions can be obtained with the weighted sum scalarization

(1) Efficient solutions obtained with the weighted sum approach are Pareto optimal in the Geoffrion sense.

(2) Solutions that belong to concave parts of the Pareto front cannot be obtained

Proper efficiency

Definition: Domination in the Geoffrion sense:
A solution $\mathbf{x}^* \in S$ is called a proper Pareto optimal solution iff:

(a) it is efficient

(b) there exists a number $M > 0$ such that $\forall i = 1, \ldots, m$ and $\forall x \in \mathcal{X}$ satisfying $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$, there exists an index j such that $f_j(\mathbf{x}^*) < f_j(\mathbf{x})$ and:

$$\frac{f_i(\mathbf{x}^*) - f_i(\mathbf{x})}{f_j(\mathbf{x}) - f_j(\mathbf{x}^*)} \leq M$$

Definition: *Proper efficient solutions* are optimal due to Geoffrion's domination criterion. They have a bounded tradeoff considering their objectives.

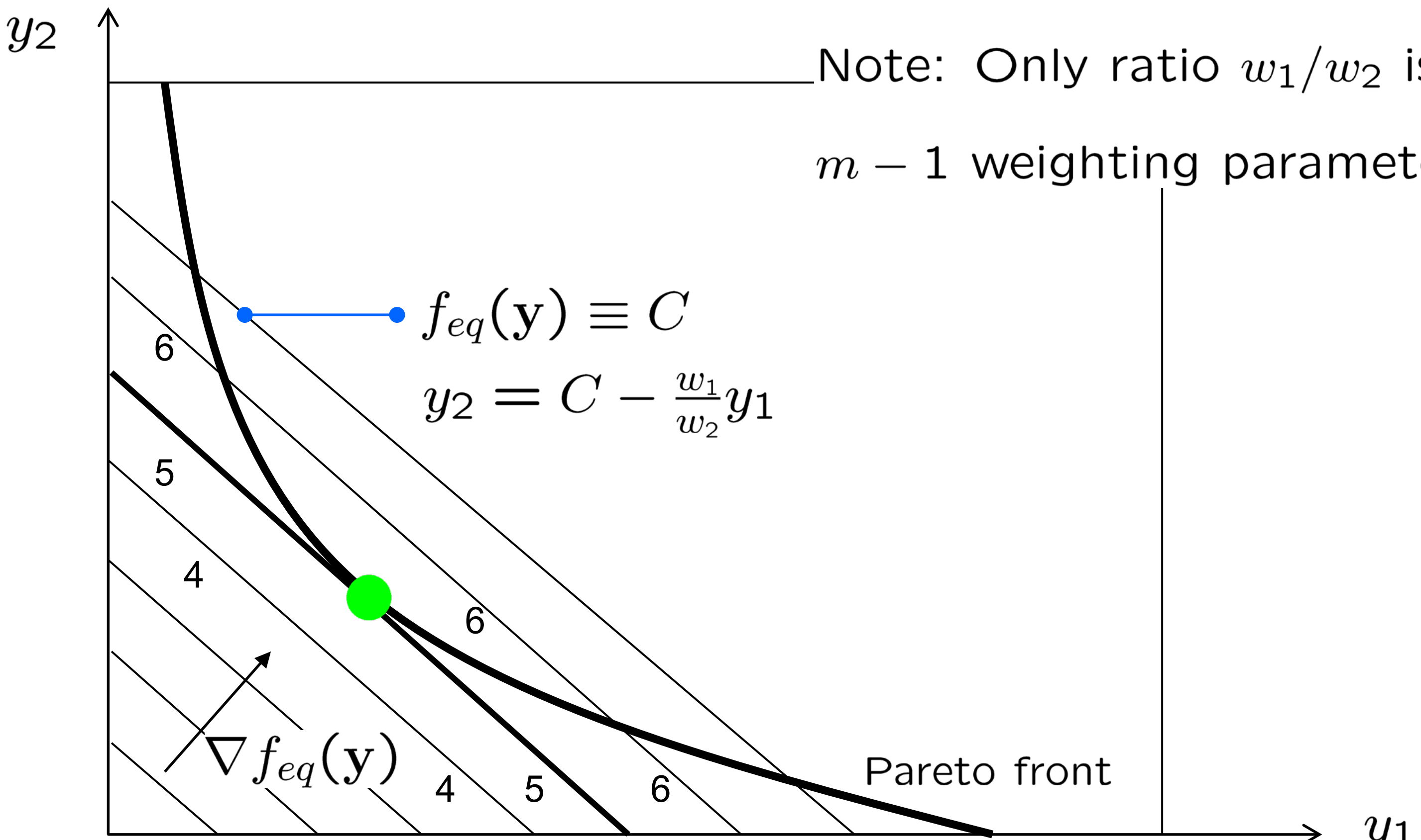
Convex Pareto front

Linear scalarization $f_{eq}(\mathbf{y}) = w_1 y_1 + w_2 y_2, w_1 > 0, w_2 > 0$

● : $\binom{y_1}{y_2} \in \mathcal{Y}$ with minimal value for $w_1 y_1(\mathbf{x}) + w_2 y_2(\mathbf{x})$

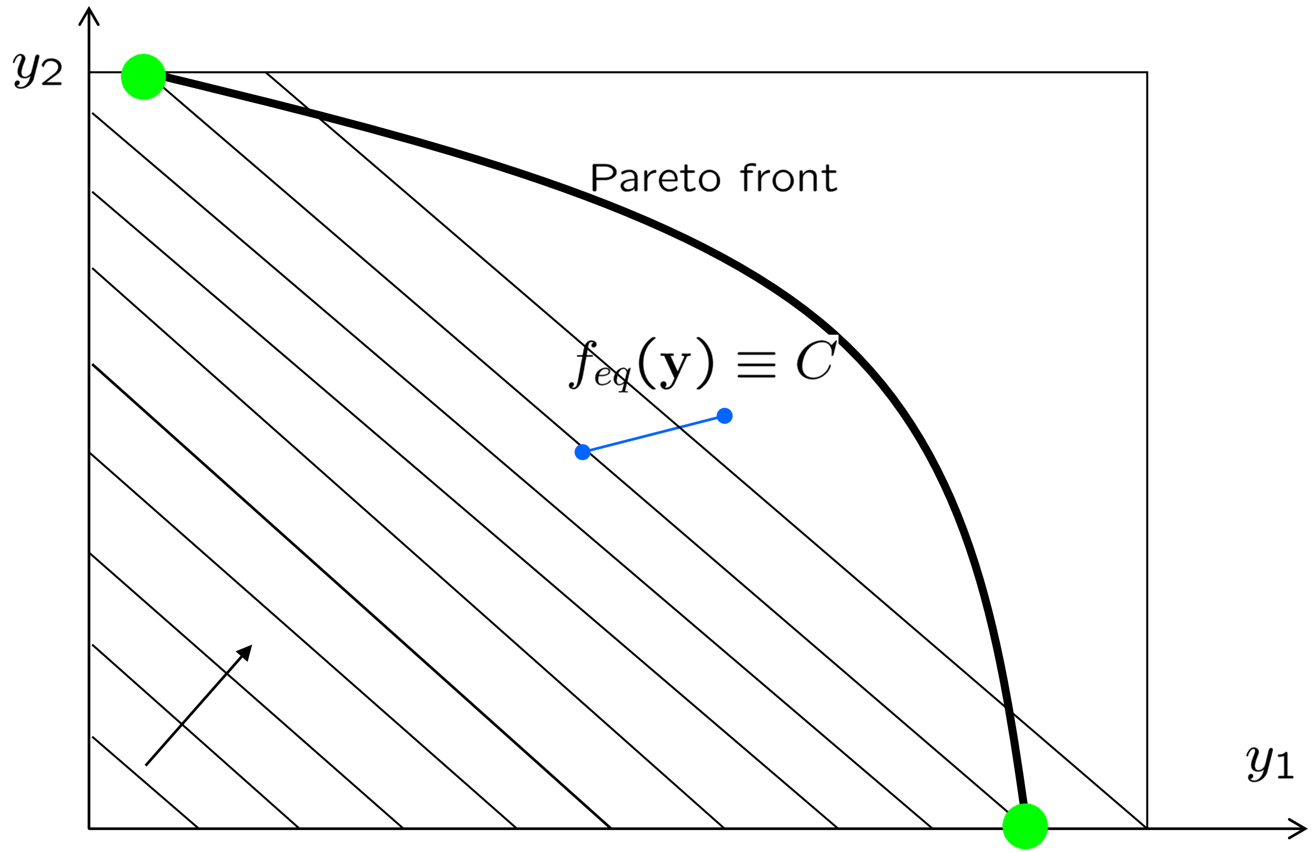
Note: Only ratio w_1/w_2 is important!

$m - 1$ weighting parameters suffice!



Universiteit Leiden

Concave Pareto front



Only extremal points can be obtained in case of concave Pareto fronts

Example: Schaffer problem

$$f_1(x) = x^2,\ f_2(x) = (x-2)^2 \qquad x \in \mathcal{X} = [0,2]$$

$$f_{eq} = w_1 x^2 + w_2 (x-2)^2 \rightarrow \text{min}$$

Ansatz: Find all x with $\frac{\partial f_{eq}}{\partial x} = 0,\ \frac{\partial^2 f_{eq}}{\partial x^2} > 0$

$$\frac{\partial f_{eq}}{\partial x} = 2x(w_1 + w_2) - 4w_2 = 0 \text{ (a)},\ \frac{\partial^2 f_{eq}}{\partial x^2} = 2(w_1 + w_2) > 0$$

(b) is always fulfilled, since $w_i > 0$

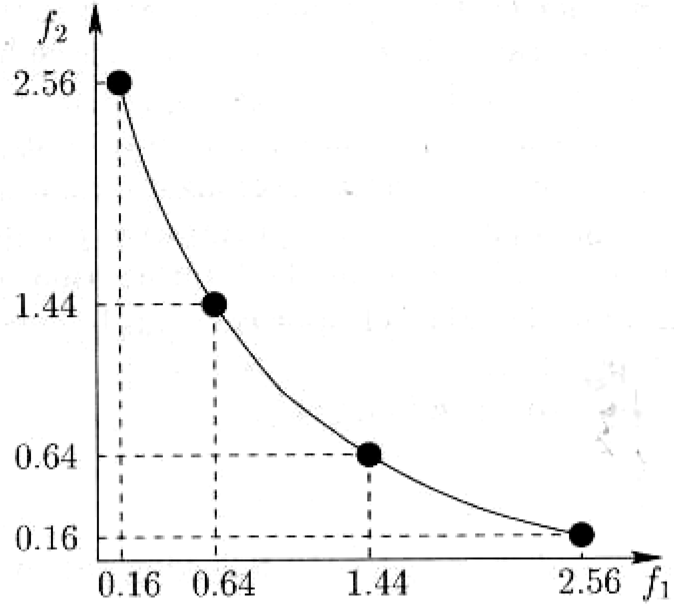
(a) $\Leftrightarrow x^* = 2w_2/(w_1 + w_2) \underbrace{=}_{w_1+w_2=1} 2w_2$

Universiteit Leiden

Example: Schaffer problem

Table 2.1. Recapitulatory table.

w_1	0.2	0.4	0.6	0.8
w_2	0.8	0.6	0.4	0.2
x^*	1.6	1.2	0.8	0.4
$f_1(x^*)$	2.56	1.44	0.64	0.16
$f_2(x^*)$	0.16	0.64	1.44	2.56

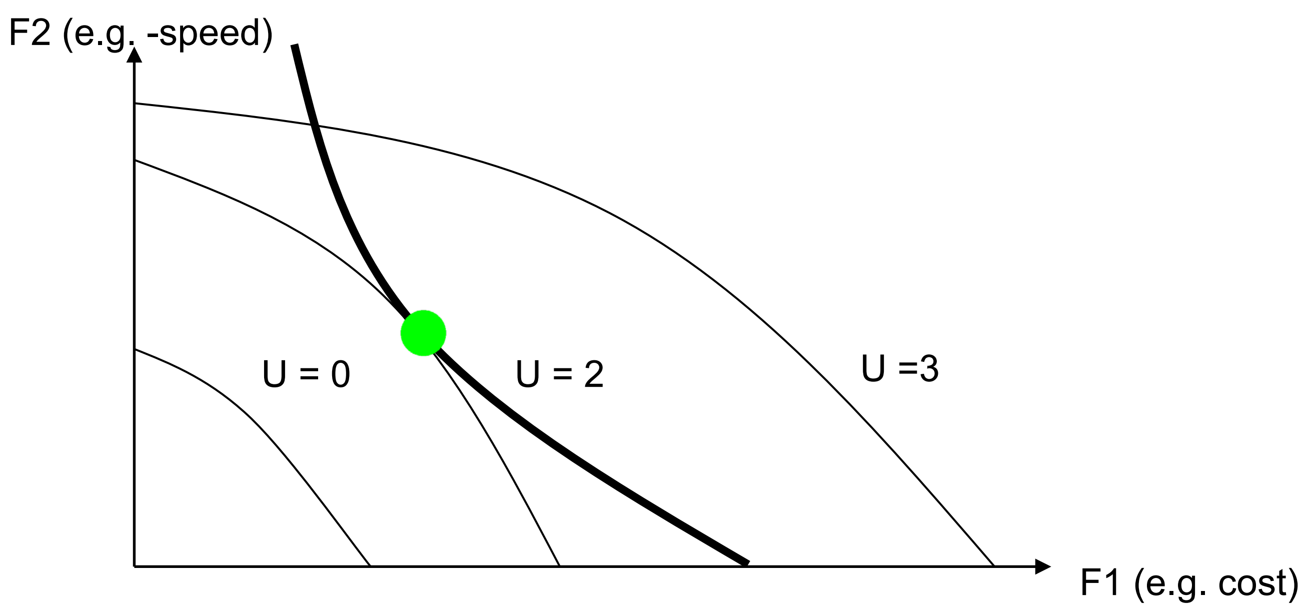


Source: Siarry et al. Multiobjective Optimization, Springer, Berlin

Universiteit Leiden

Utility functions

Once a (proper) utility function is given the tangential points of the iso-utility curves with the Pareto front are the obtained non-dominated points.

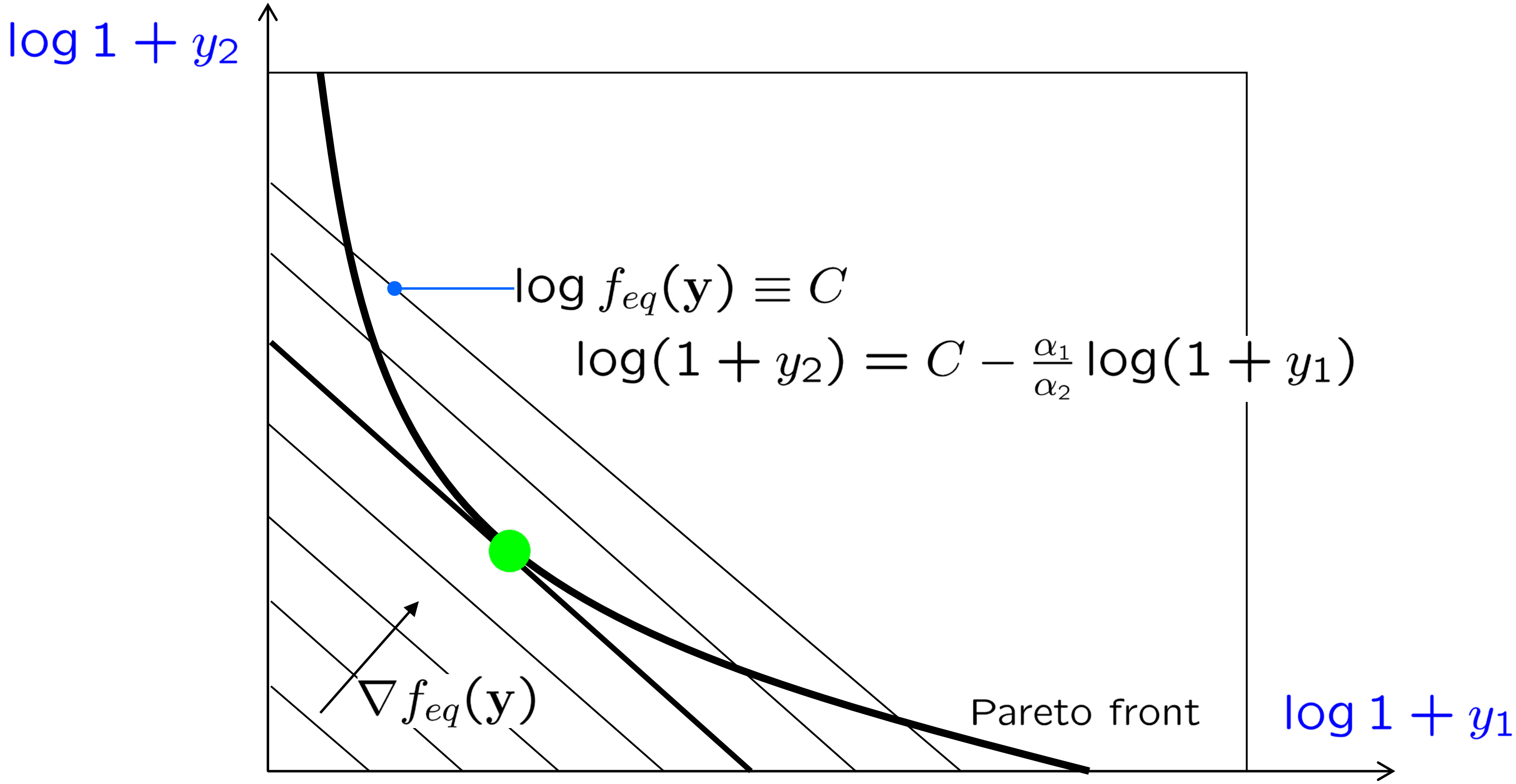


Keeney and Raiffa: Decisions with Multiple Objectives: Preferences and Value Tradeoff, Cambridge Univ. Press, 1993

Cobbs Douglas utility functions

Cobbs Douglas Utility: $f_{eq}(\mathbf{x}) = \prod_{\mathrm{i}=1}^{\mathrm{m}} (1 + \mathbf{f}_{\mathrm{i}}(\mathbf{x}))^{\alpha_{\mathrm{i}}}, \alpha_{\mathrm{i}} > \mathbf{0}$

Note: $\log f_{eq}(\mathbf{x}) = \sum_{\mathrm{i}=1}^{\mathrm{m}} \alpha_{\mathrm{i}} \log \mathbf{f}_{\mathrm{i}}(\mathbf{x}), \alpha_{\mathrm{i}} > \mathbf{0}$ can be used for drawing iso utility lines!



Universiteit Leiden

Distance to a reference point (DRP) method

- These methods aim for minimizing the distance to an ideal point
- The ideal point has multiple components and these are the objective function values to be minimized
- Examples:
  - In a machine learning problem the false positive rate fp and false negative fn rate should be simultaneously minimized. The ideal point is $fn=(0,0)^T$
  - In a control problem the pressure should be kept close to p* and the temperature close to T*. The ideal point is $(T,p)^T=(p^*, T^*)^T$.
  - In an building optimization problem the fuel consumption EC should be ideally 0 and the annual operation cost AC and investment cost IC, too. The ideal point is $(EC,AC,IC)^T=(0,0,0)^T$

Minkowsky distance functions

General distance to reference point $\mathbf{f}^* \in \mathbb{R}^m$

$$f_{eq}(\mathbf{x}) = \left(\sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|^p\right)^{1/p}$$

Example $p = 1$:

$$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|$$

Example $p = 2$:

$$f_{eq}(\mathbf{x}) = \left(\sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|^2\right)^{1/2}$$

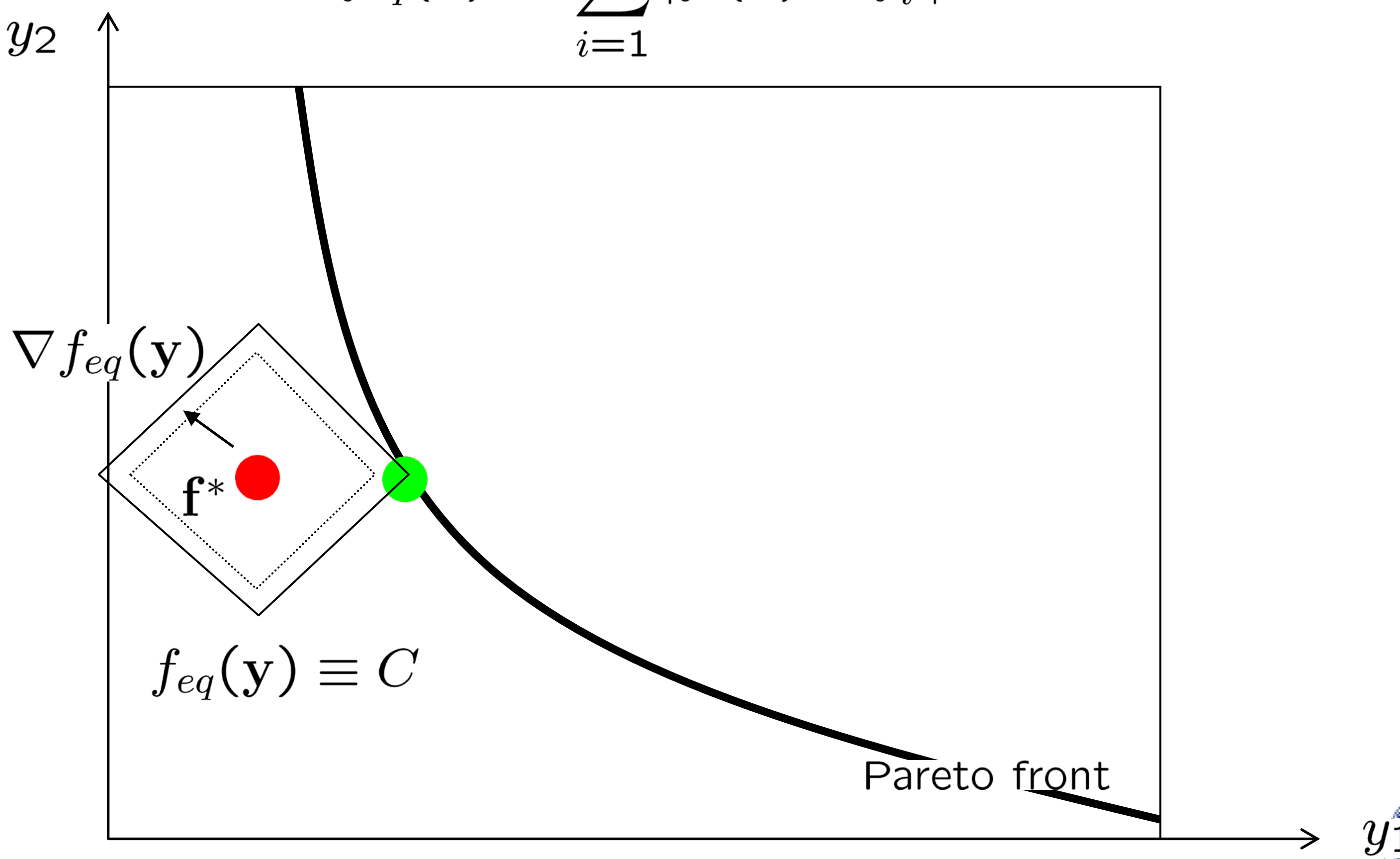
Example $p = \infty$: Tschebyscheff Distance

$$f_{eq}(\mathbf{x}) = \max_{i=1,\ldots,m} |f_i(\mathbf{x}) - f_i^*|$$

View of DRP as a utility function:

Example $p = 1$:

$$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|$$

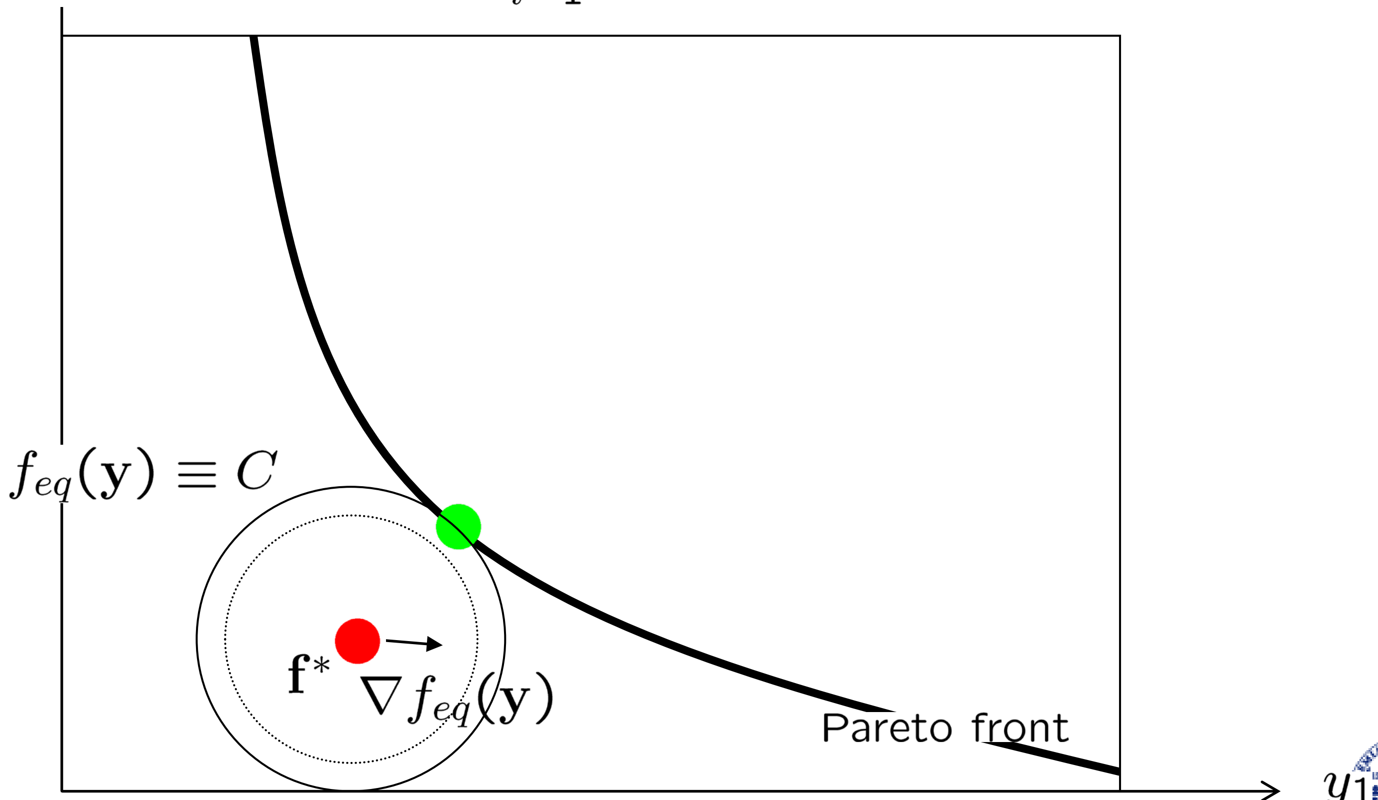


Universiteit Leiden

View of DRP as a utility function:

Example $p = 1$:

$$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|^2$$

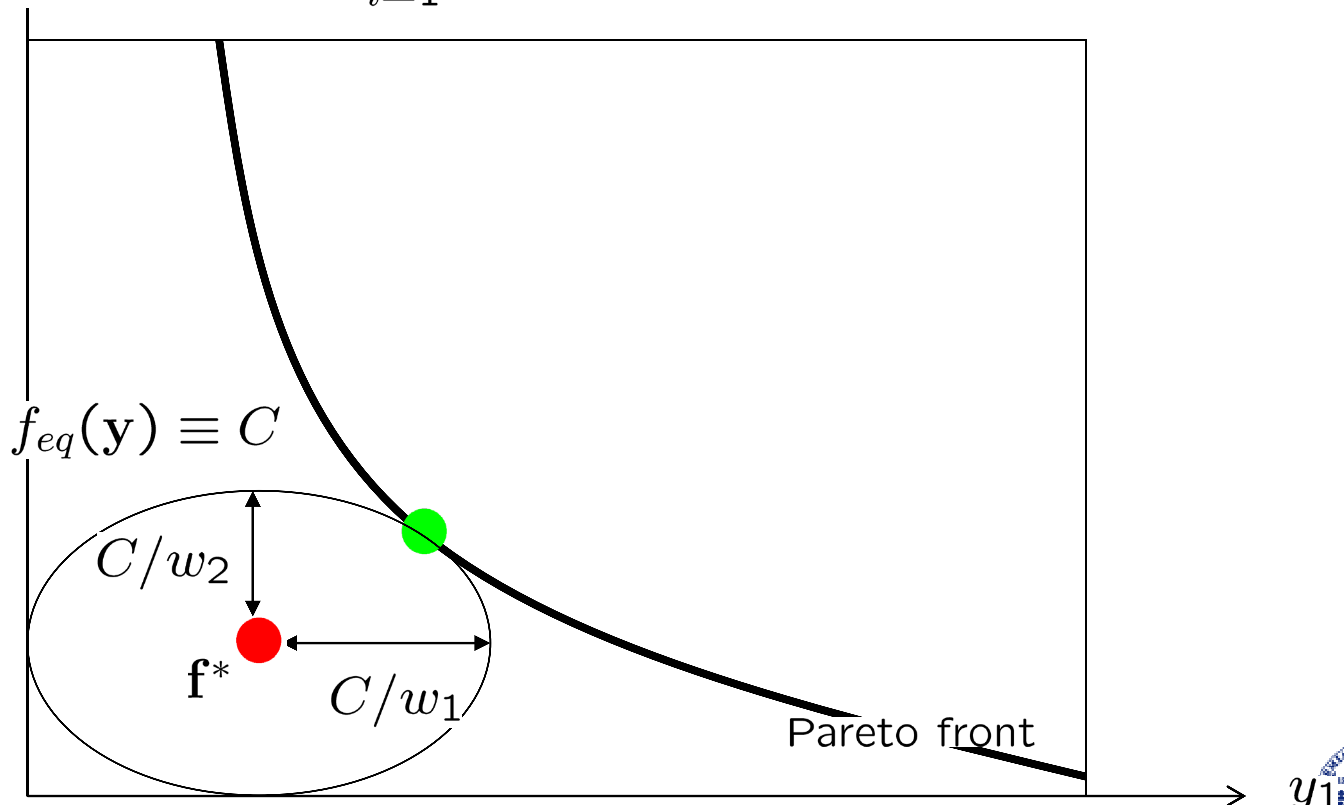


Universiteit Leiden

View of DRP as a utility function: Weighted euclidian distance function

Example $p = 1$:

$$f_{eq}(\mathbf{x}) = (\sum_{i=1}^{m} w_i |f_i(\mathbf{x}) - f_i^*|^2)^{1/2}$$

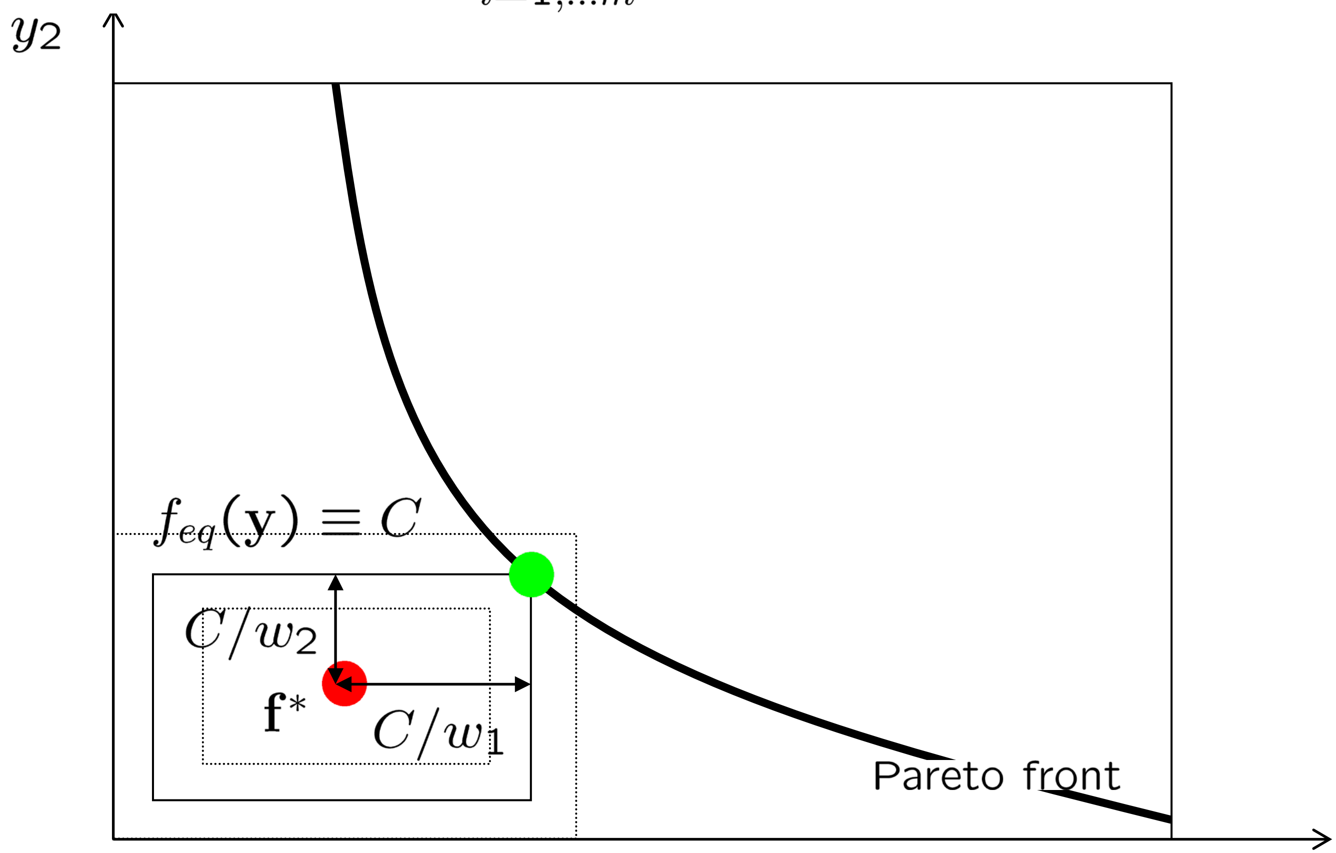


Universiteit Leiden

Tschebychev DRP:

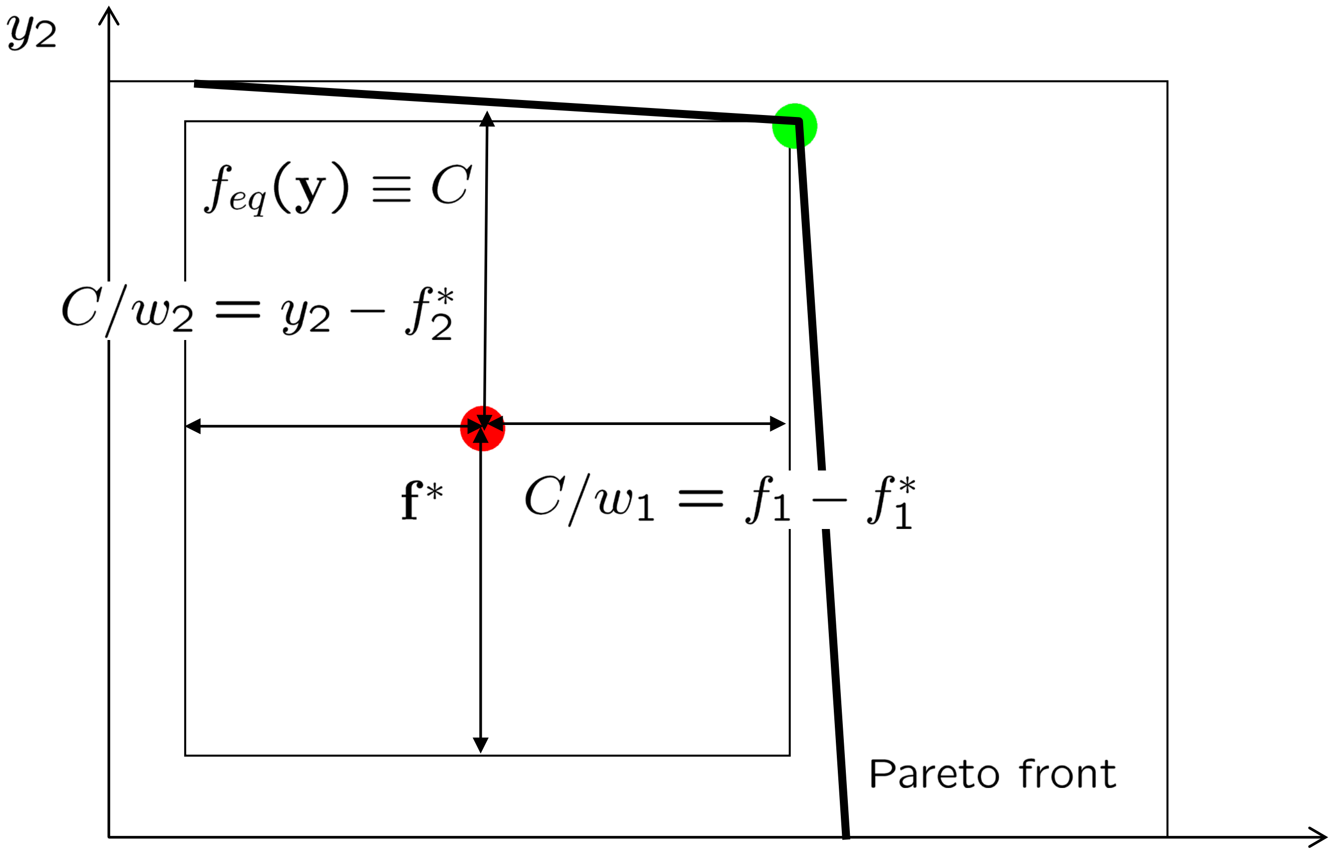
Example $p = 1$:

$$f_{eq}(\mathbf{x}) = \max_{i=1,\ldots m} w_i |f_i(\mathbf{x}) - f_i^*|$$



Tschebychev DRP:

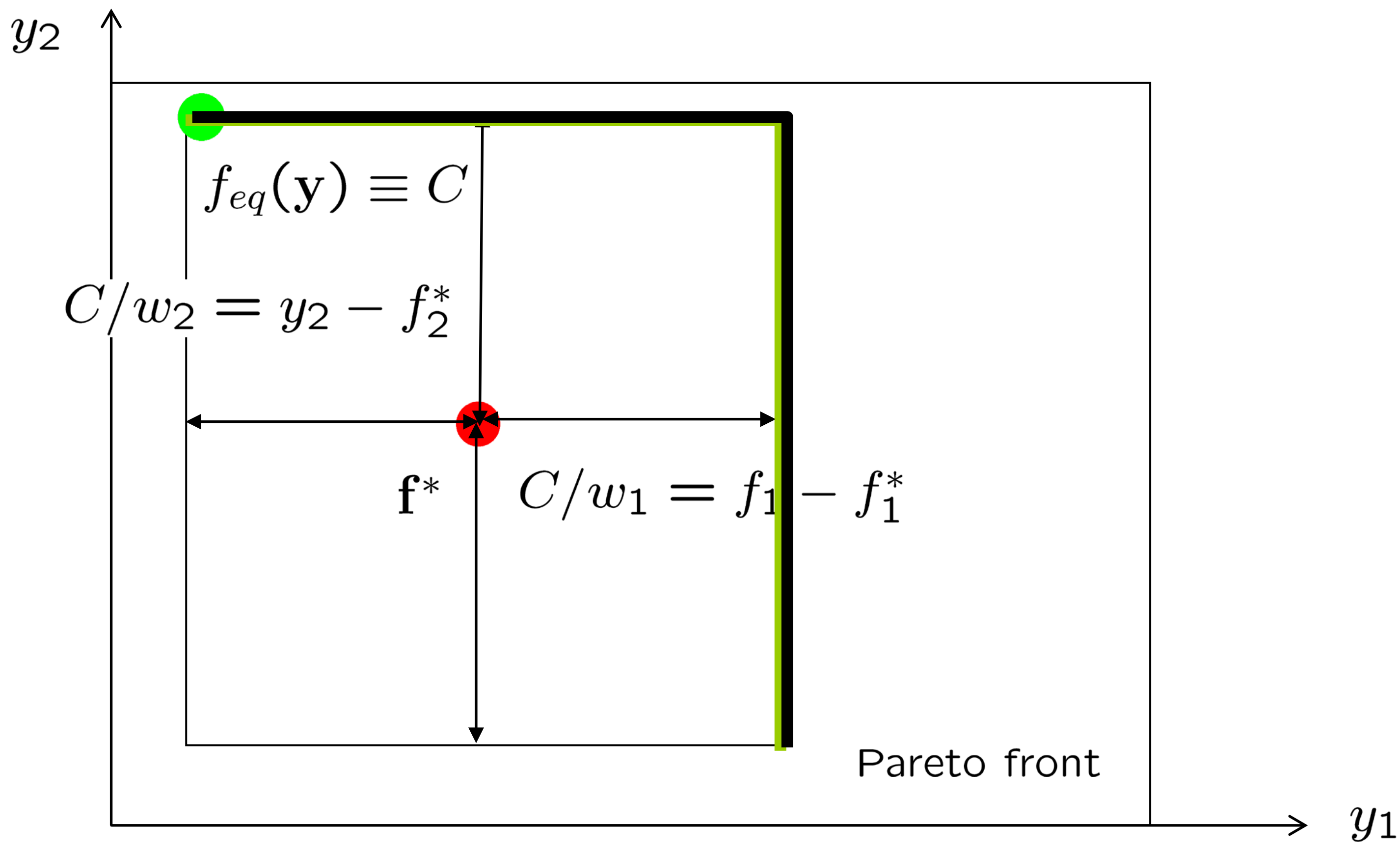
Let $\mathbf{f}^* \preceq \mathbf{f}^I$ (reference point is dominated by ideal point). For every properly efficient point $\mathbf{y} \in \mathcal{Y}_N$ we can find a combination of weights, such that the minimization of Tschebytscheff utility f_{eq} leads to $\mathbf{x}$ with $\mathbf{f}(\mathbf{x}) = \mathbf{y}$.



Universiteit Leiden

Tschebychev DRP:

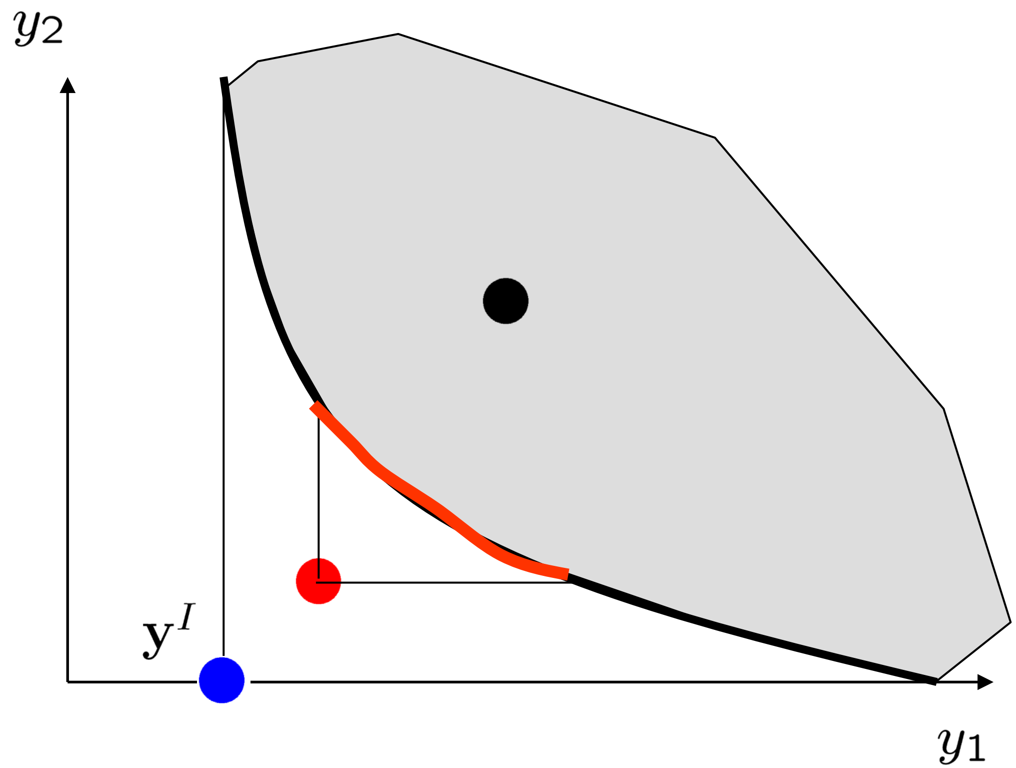
Weakly efficient solutions may be among the results!



Universiteit Leiden

Choice of reference point

Reference point should be dominated by ideal point $\mathbf{y}^I$!



Pessimistic choice of $\mathbf{f}^* \Rightarrow$ not all points on Pareto surface can be obtained, or even dominated point result in the minimization of f_{eq}.

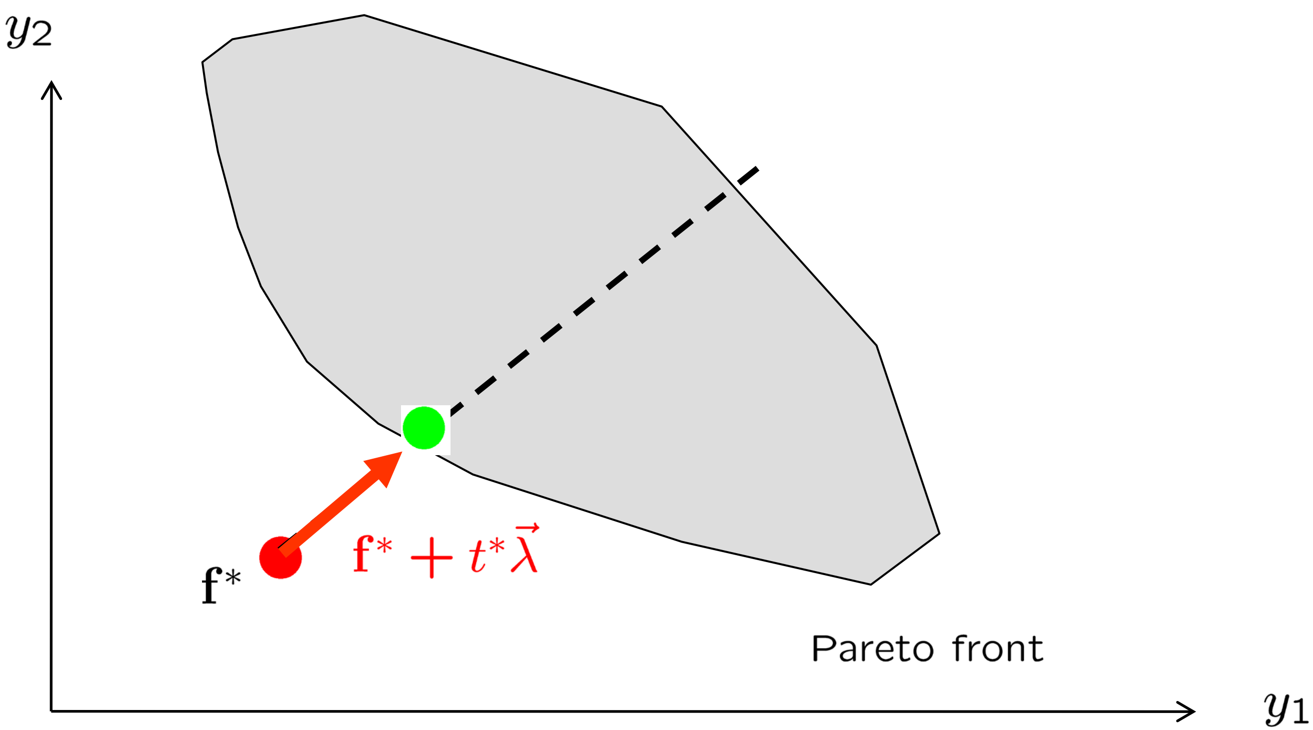
Universiteit Leiden

Goal programming

(1) Choose reference point!

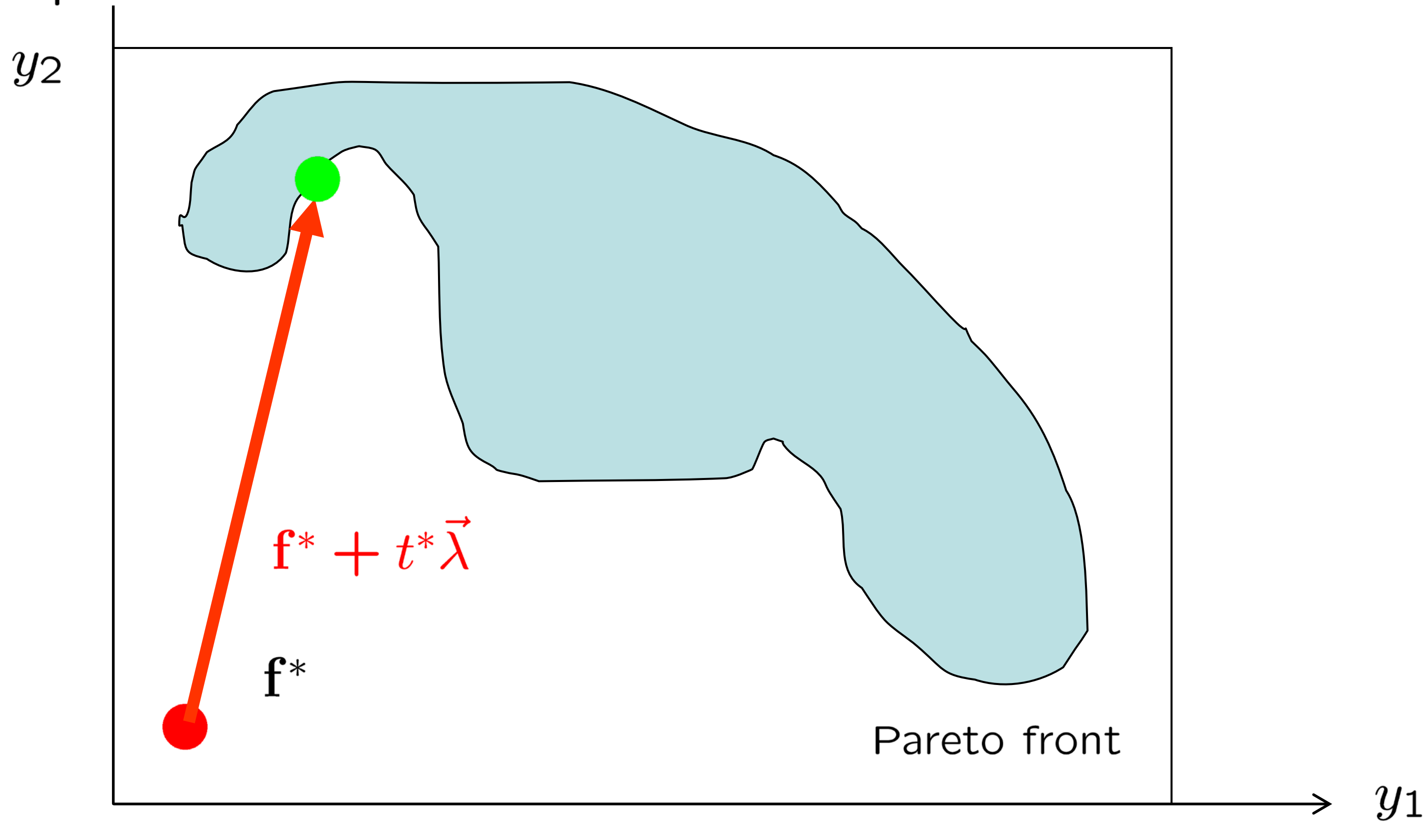
(2) Choose positive direction $\lambda \in \mathbf{R}^m_>$!

(3) Find minimal t s.t. $\mathbf{f}^* + t\vec{\lambda} \in \mathcal{Y}$



Goal programming

If $\lambda_i > 0$ goal programming can obtain all properly efficient points!



Goal programming might result in dominated solutions!

This is also possible if reference point dominates ideal point!

Summary:Scalarization methods

Scalarization methods can obtain Pareto optimal solutions.

Except Tschebyscheff scalarization the methods cannot find all proper efficient solutions, , in particular concave parts are easily overseen!

The goal attainment method may even find non-efficient points

The weights w_i and ϵ-constants have different meaning, the understanding of which is essential to the understanding of the respective method.

Finding tangential points of the $f_{eq} \equiv C$ isolines gives us a practical means for geometrically determining the solution of the monocriterial functions.

Universiteit Leiden