Short Unit: Single point methods for finding the Pareto front

Learning goals

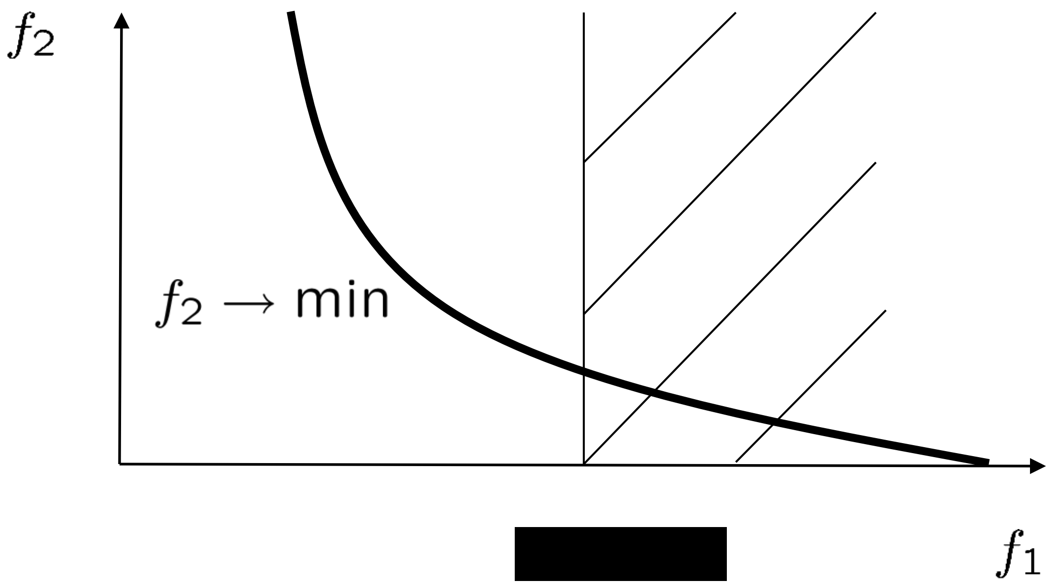
- What are different ways to solve multiobjective optimization problems by formulating them as single objective optimization problems (with constraints)?
- Can we use linear weighting functions to find all Pareto optimal points?
- Which points (on the Pareto front) do we find for different scalarization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

Single point methods

Strategy 1: Sum up function values by means of an utility function

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

Strategy 2: Add constraints to the problem (e.g. transform objectives into constraints)



Weighted sum scalarization

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

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

Proof: We show that a dominated solution can always be improved: $f_i(\mathbf{x}) \leq f_i(\mathbf{x}')$ for $i = 1, \ldots, m$ and $f_j(\mathbf{x}) < f_j(\mathbf{x}')$ for some $j \in \{1, \ldots, m\} \Rightarrow \sum w_i f_i(\mathbf{x}) < \sum w_i f_i(\mathbf{x}')$

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.

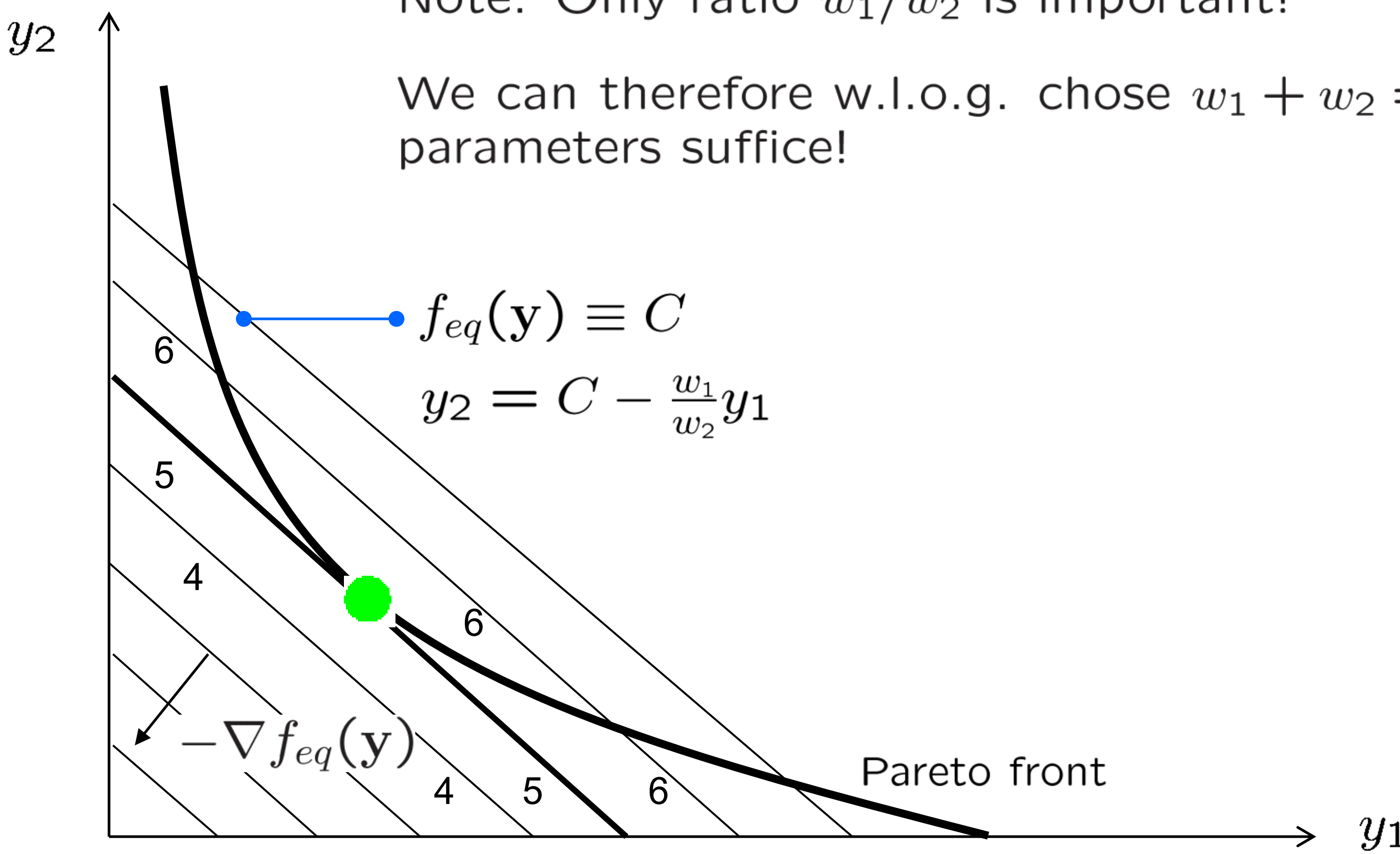
Convex Pareto front

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

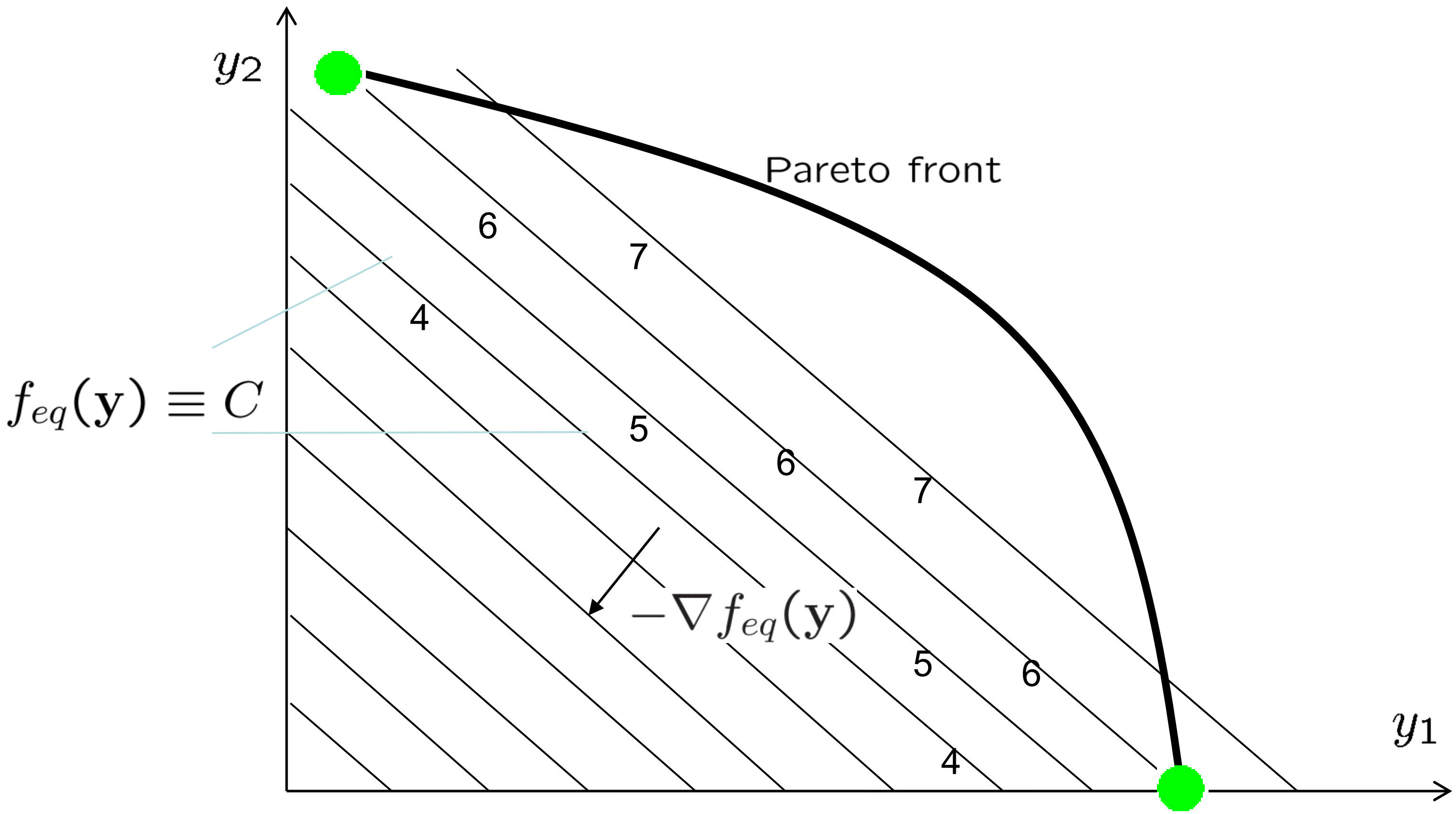
● : $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathcal{Y}$ with minimal value for $w_1 y_1(\mathbf{x}) + w_2 y_2(\mathbf{x})$

Note: Only ratio w_1/w_2 is important!

We can therefore w.l.o.g. chose $w_1 + w_2 = 1$ weighting parameters suffice!



Concave Pareto front



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

Example: Schaffer problem

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

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

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

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

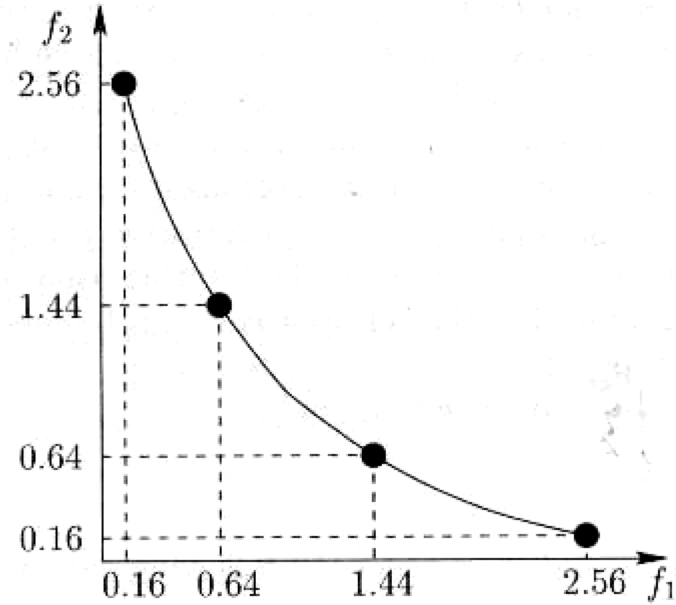
(b) is always fulfilled, since $w_i > 0$

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

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

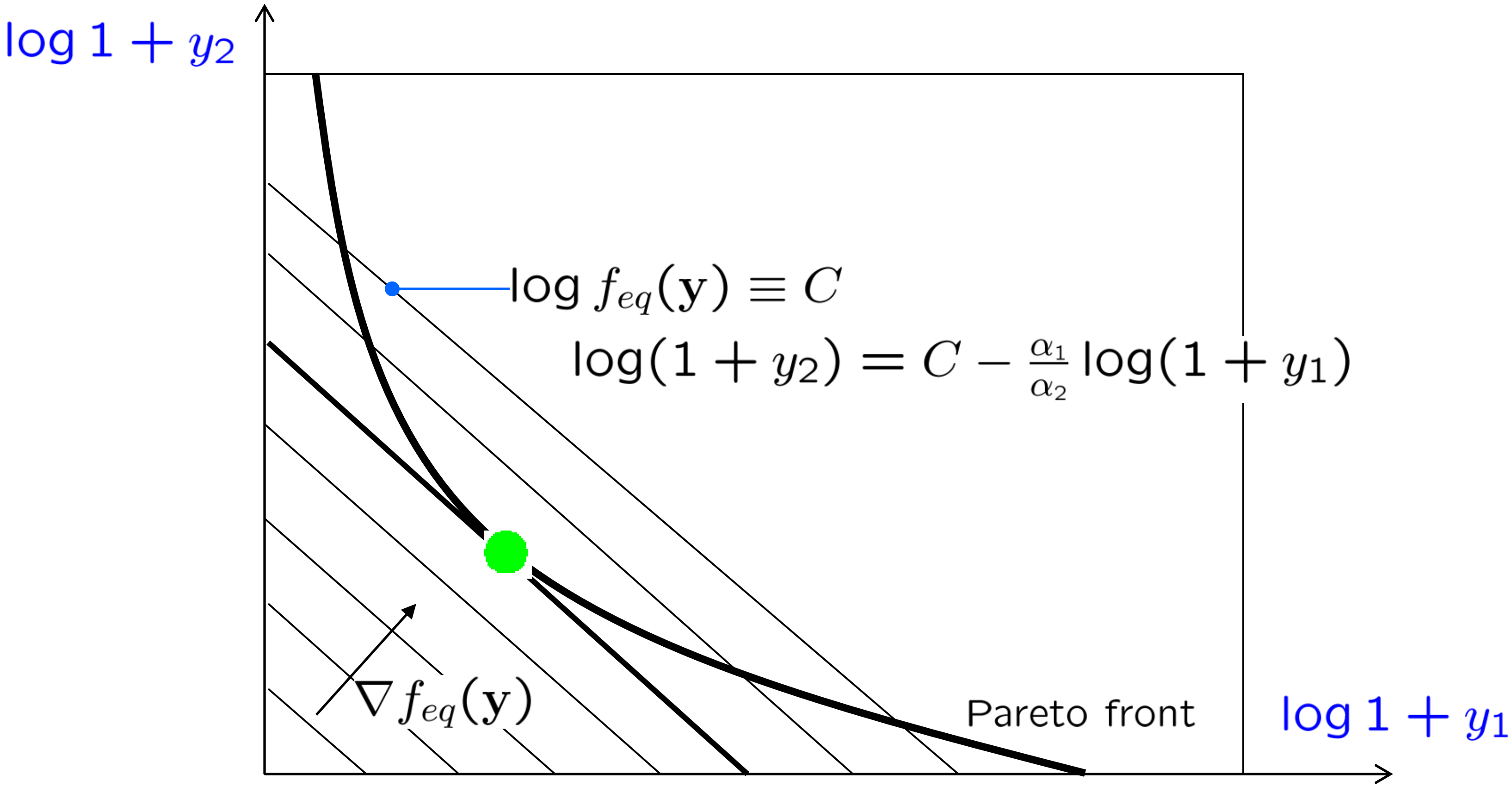


Source: Siarry et al. Multiobjective Optimization, Springer, Berlin

Cobbs Douglas utility functions

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

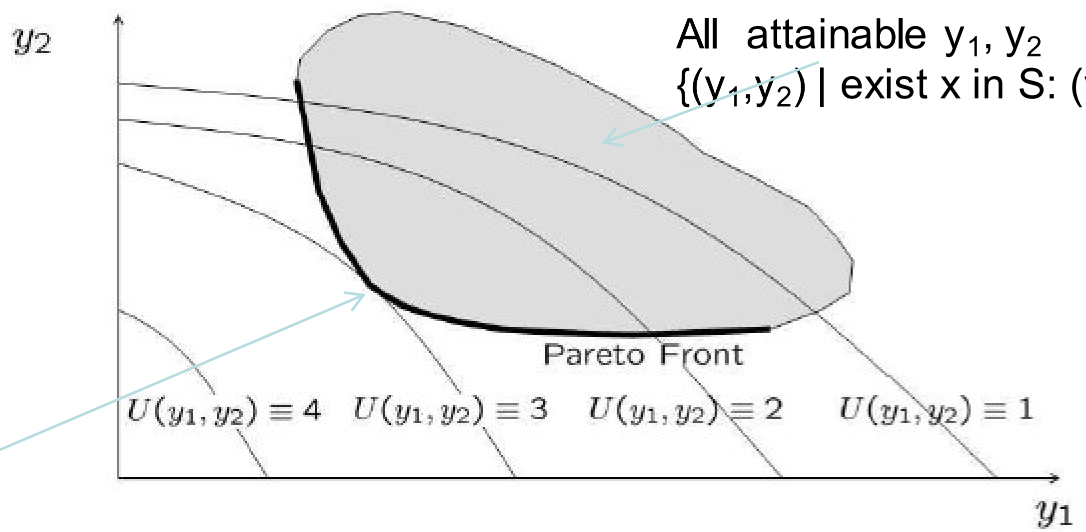
Note: $\log f_{eq}(\mathbf{x}) = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{m}} \alpha_\mathbf{i} \log(\mathbf{1}+\mathbf{f_i}(\mathbf{x})), \alpha_\mathbf{i} > \mathbf{0}$ can be used for drawing iso utility lines!



General utility functions

The point where the level curve (or indifference curve) **with the best** level that has an intersection with the **Pareto front** is the point that will be obtained when **optimizing** the utility function.

All attainable y_1, y_2
$\{(y_1,y_2) \mid \text{exist } x \text{ in } S\text{: } (y_1,y_2) = (f_1(x),f_2(x))$



Optimizer of $U(f_1(x),f_2(x))$ over all x in S

Depending on the utility function class, some points on **PF might** not be accessible.

Distance to a reference point (DRP) method

- These methods aim for minimizing the distance to a reference point in the objective space
- 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 fpr and false negative fnr rate should be simultaneously minimized. The ideal point is $(fnr, fpr)^T=(0,0)^T$
  - 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 $p = 2$:

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

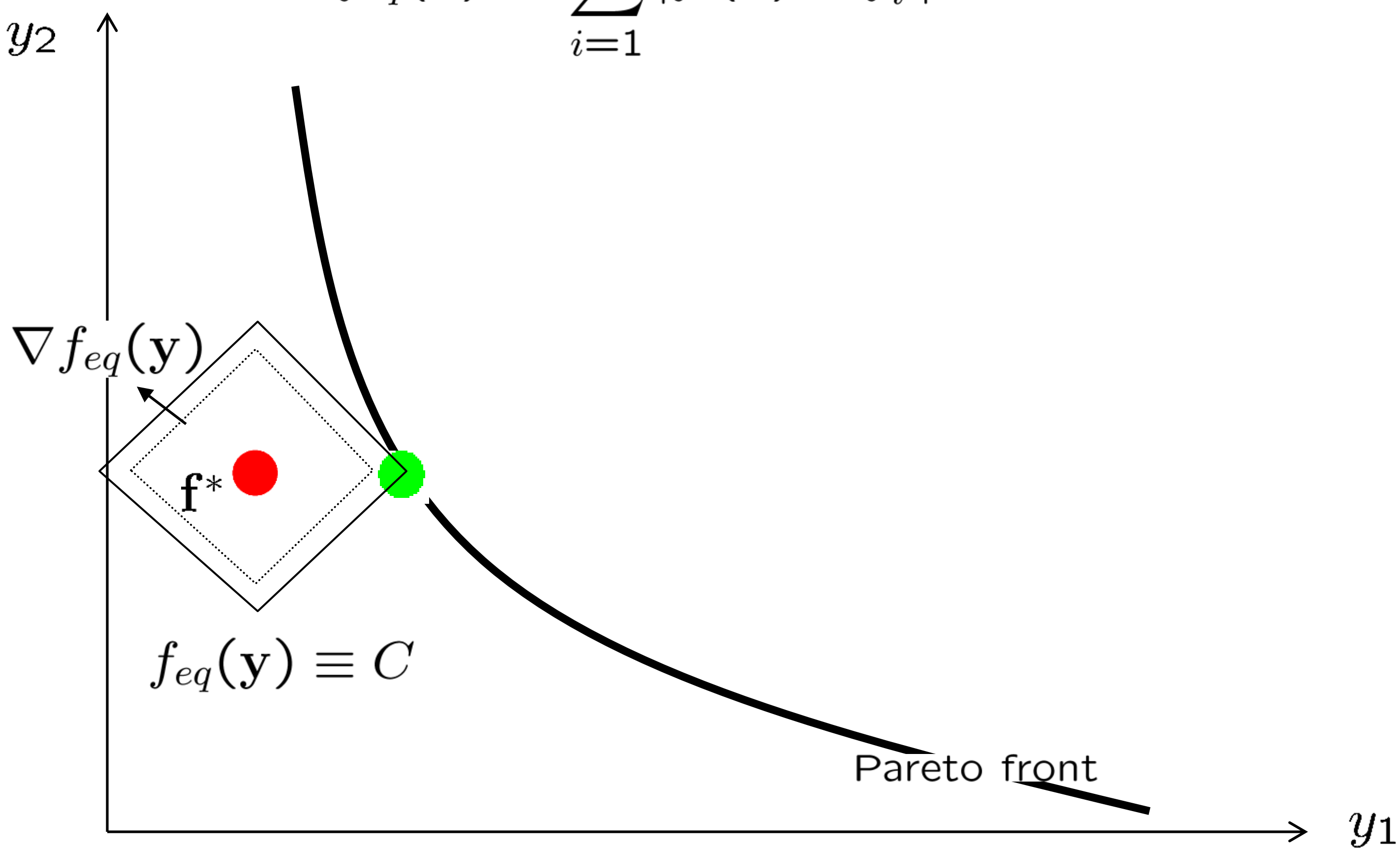
Example $p = \infty$: Tschebyscheff Distance

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

View of DRP as a utility function:

Example $p = 1$:

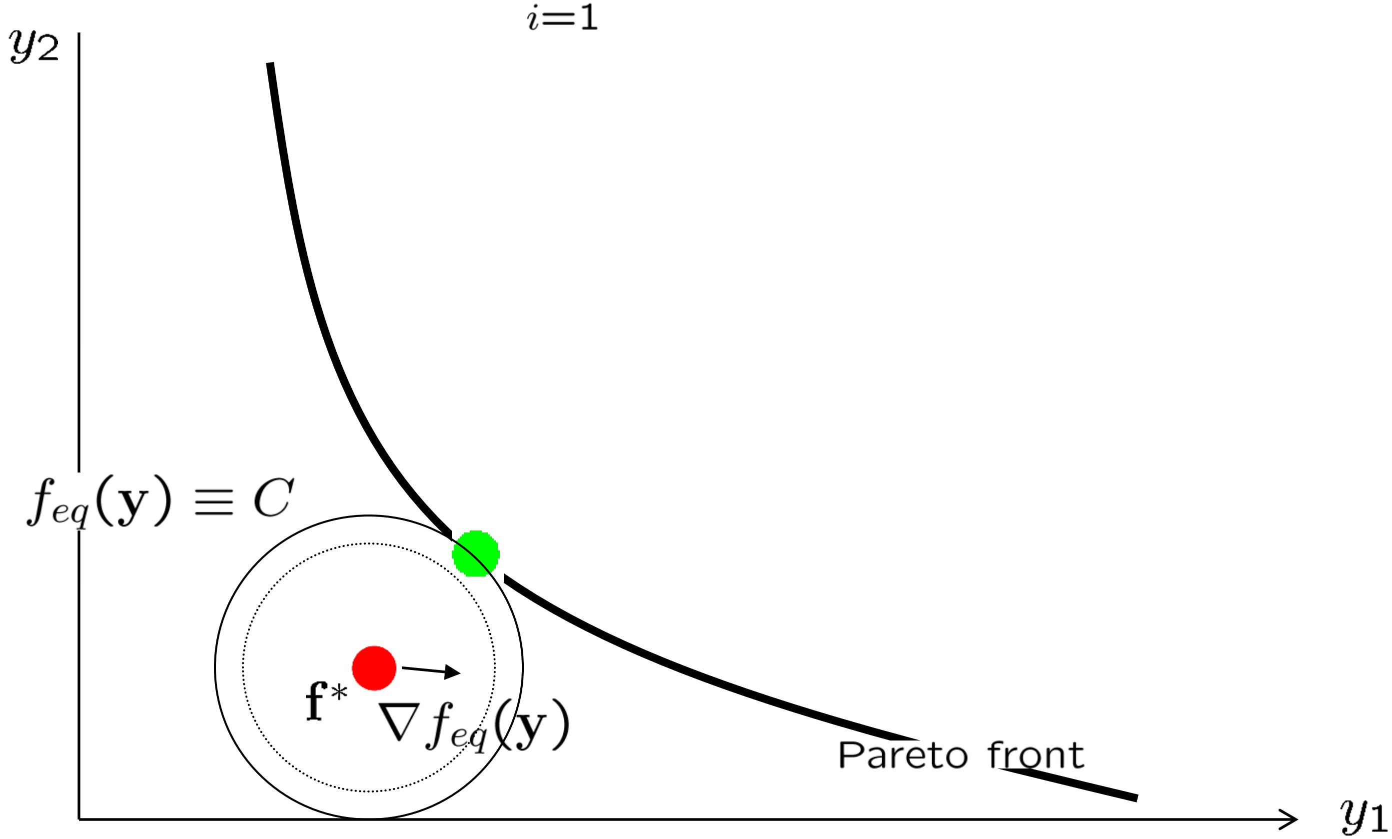
$$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|$$



View of DRP as a utility function:

Example $p = 1$:

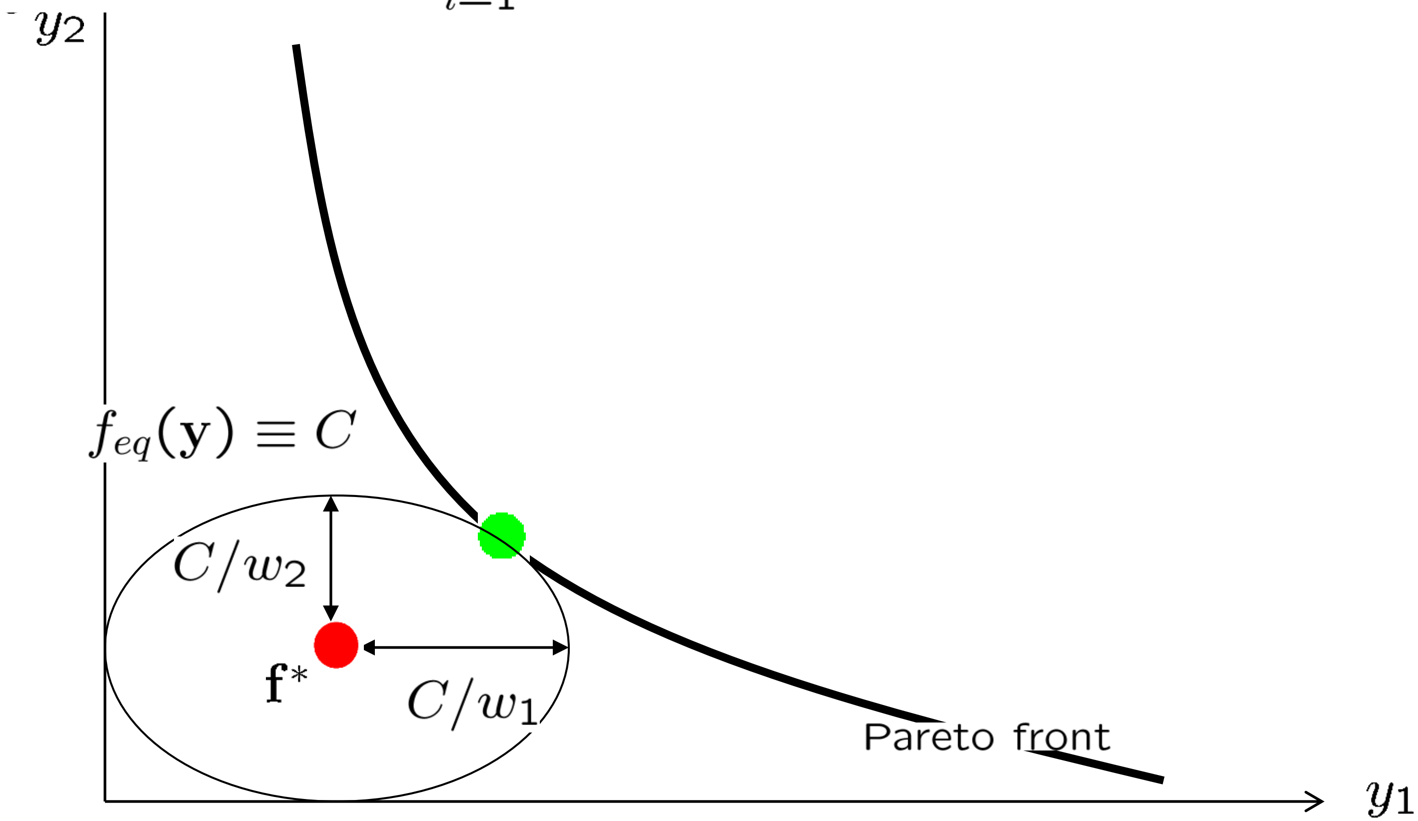
$$f_{eq}(\mathbf{x}) = \sum_{i=1}^{m} |f_i(\mathbf{x}) - f_i^*|^2$$



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

Example $p = 1$:

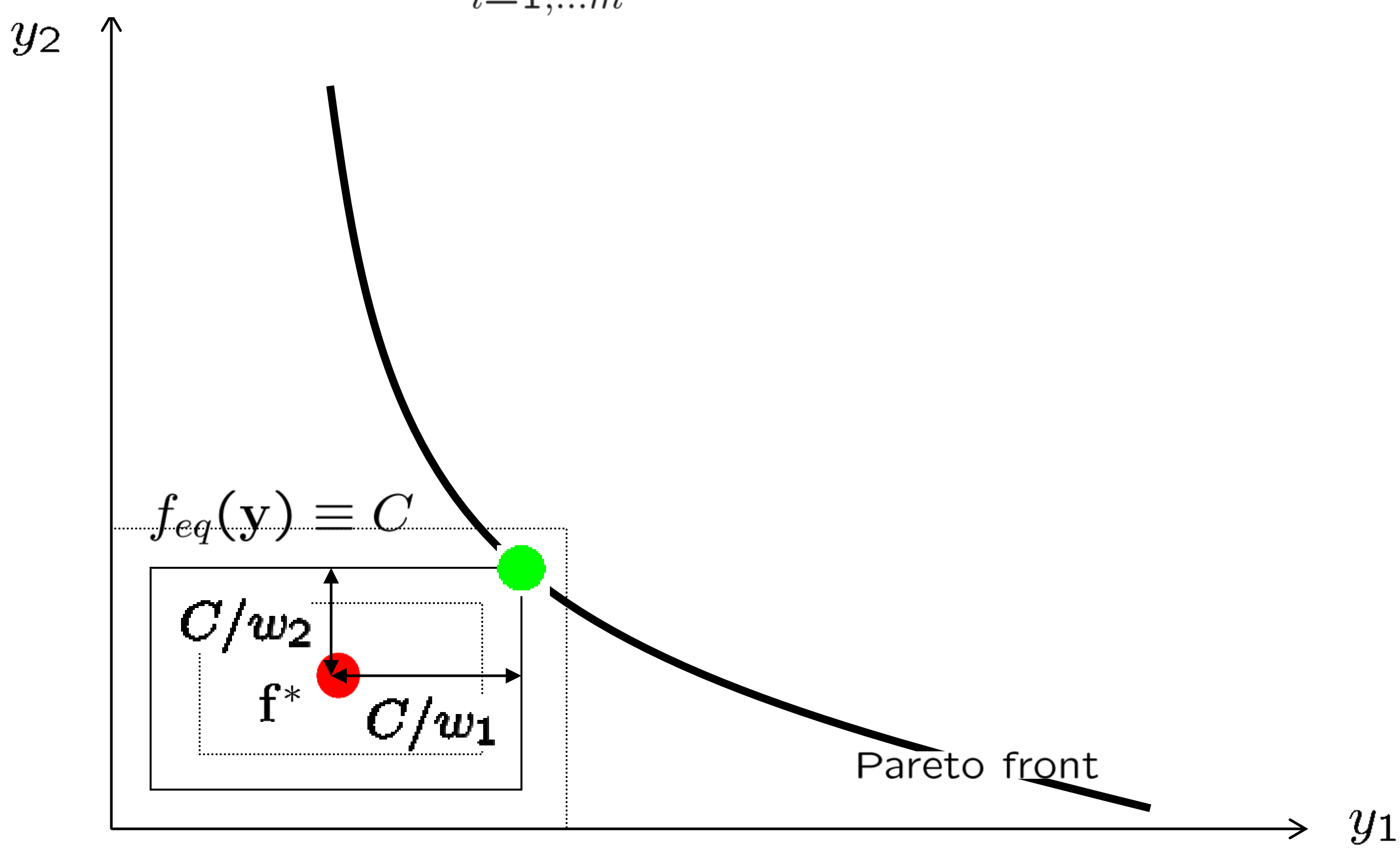
$$f_{eq}(\mathbf{x}) = (\sum_{i=1}^{m} w_i |f_i(\mathbf{x}) - f_i^*|^2)^{1/2}$$



Tschebychev DRP:

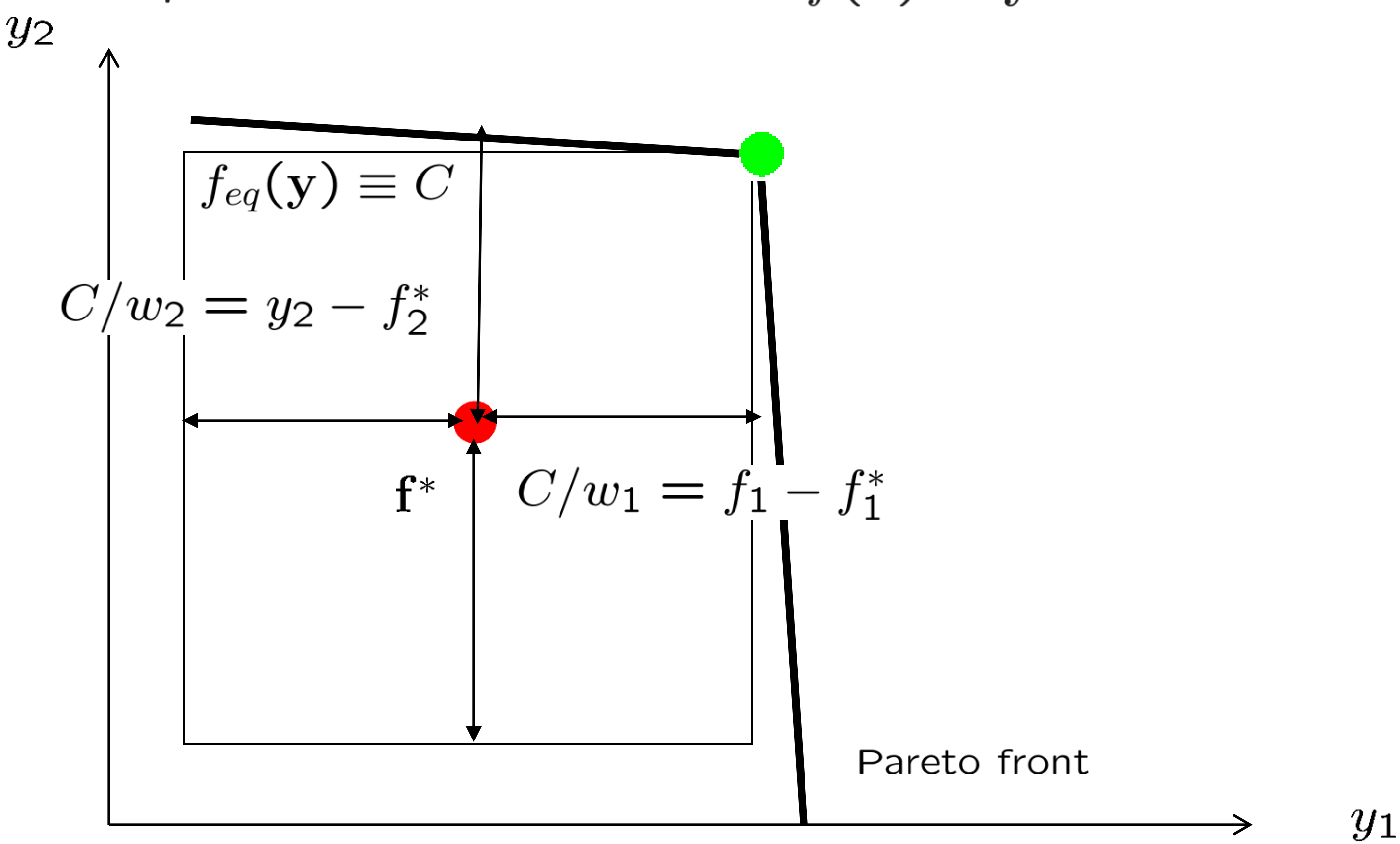
Example $p = \infty$:

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



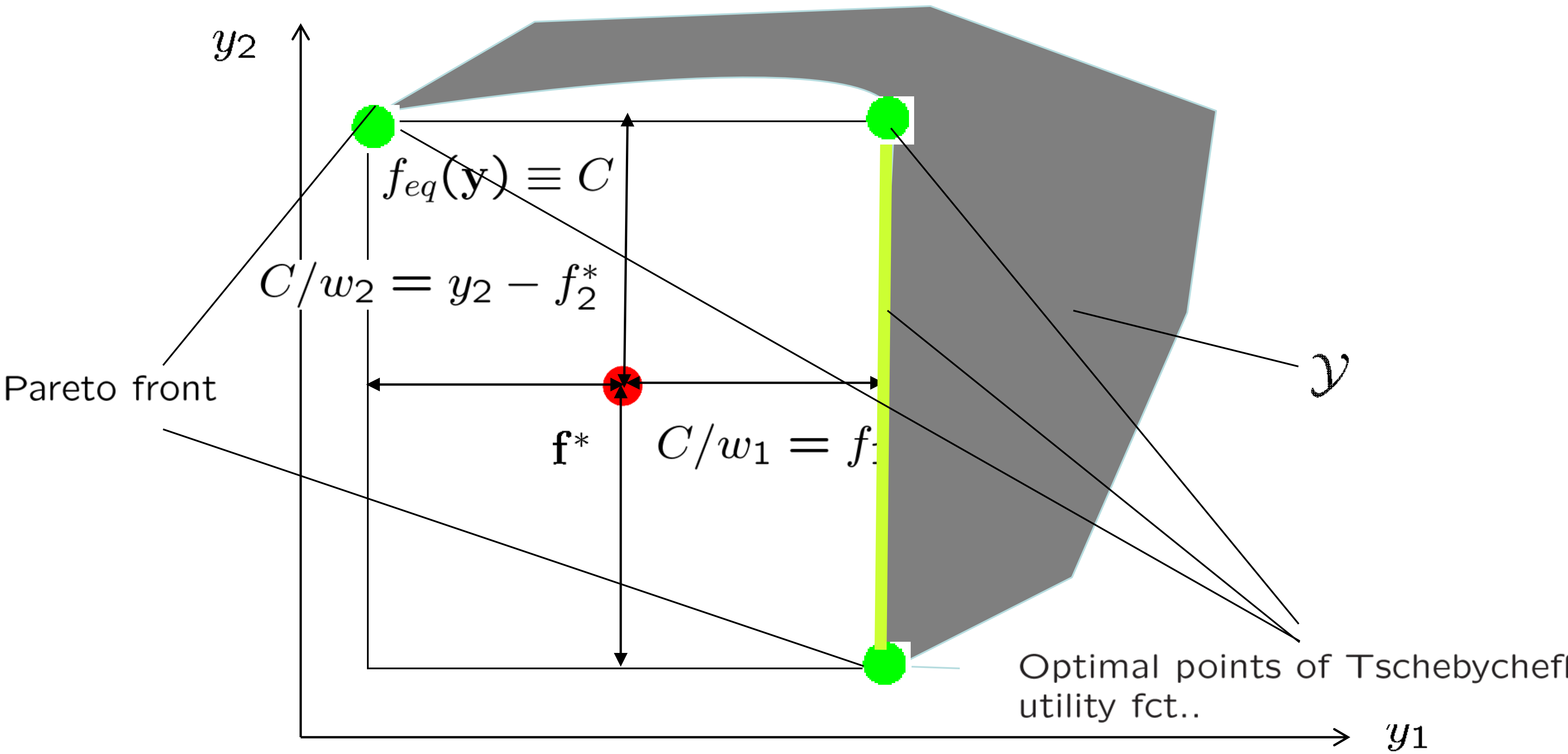
Tschebychev DRP:

Theorem: Given an reference point that is strictly dominated by the ideal point, i.e. $\mathbf{f}^* \preceq \mathbf{f}^I$.
For every properly efficient point $\mathbf{y} \in \mathcal{Y}_N$ we can find a combination of weights, such that the minimization of Tschebychev utility $f_{eq}(\mathbf{x}) = \max_{i=1,\ldots m} w_i |f_i(\mathbf{x}) - f_i^*|$ obtains an optimum in some $\mathbf{x}$ with $f(\mathbf{x}) = \mathbf{y}$.



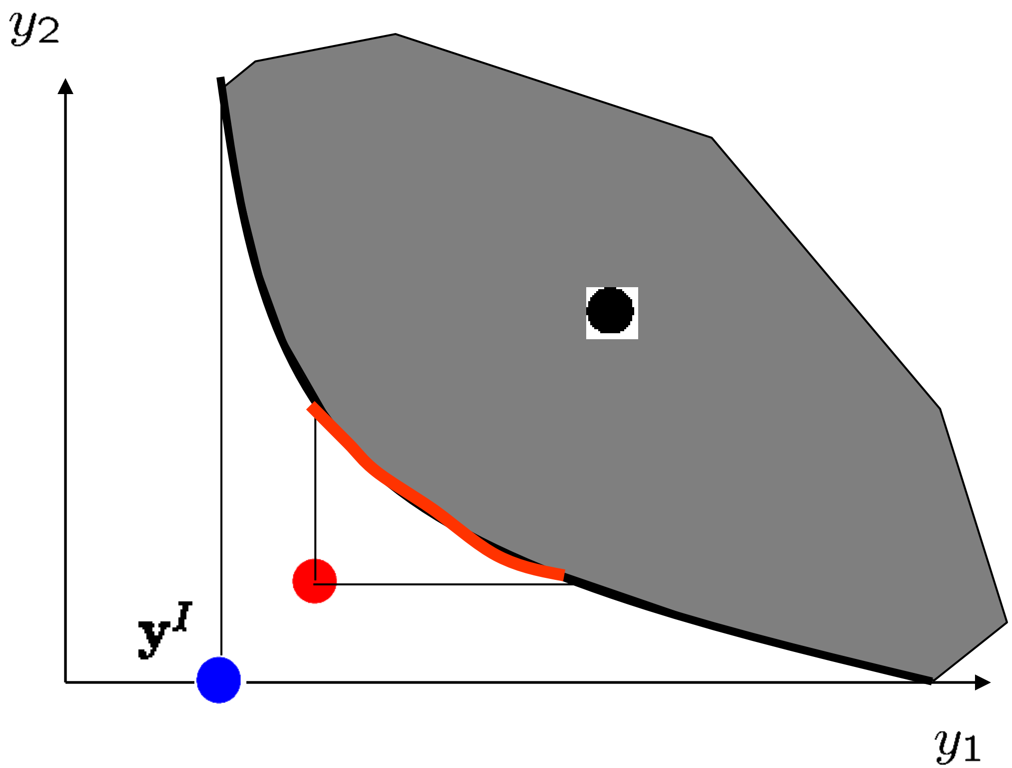
Tschebychev DRP:

However, for some Pareto fronts the optimization of the Tschebycheff function might yield weakly dominated solutions.



Choice of reference point

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



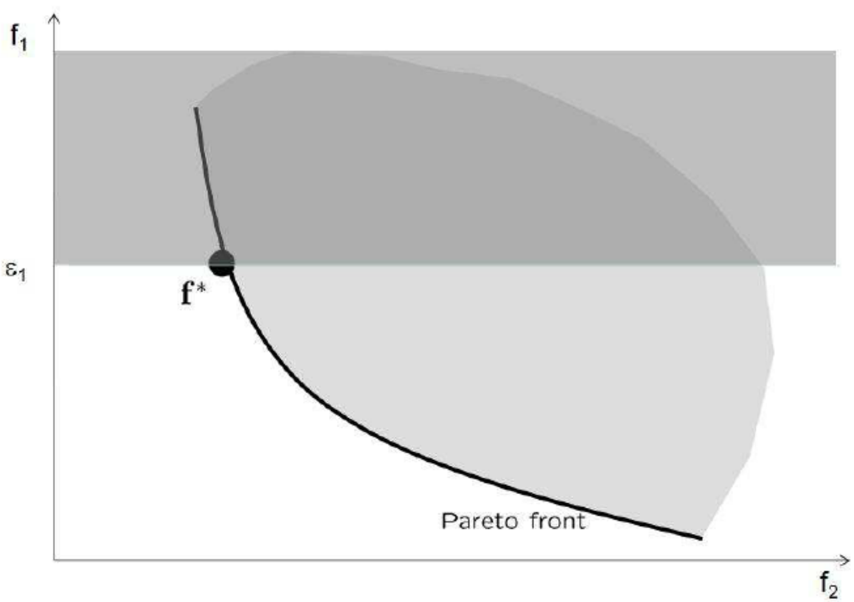
Pessimistic choice of $\mathbf{f}^*$ $\Rightarrow$ not all points on Pareto surface can be obtained, or even dominated point result in the minimization of f_{eq}.

Epsilon- Constraint method

The ϵ-constraint method maximizes f_1 while fixing the other objectives $f_2, \ldots, f_m$ to a constant ϵ.

$$f_1(\mathbf{x}) \to \min, f_2(\mathbf{x}) \leq \epsilon_1, \ldots, f_m(\mathbf{x}) \leq \epsilon_{m-1}$$

A grid is used to sample different ϵ_i values. Selection of grid determines resolution of Pareto front approximation.



In some cases,
ϵ can be eliminated
and f_2 can be expressed
in terms of f_1.
See example

Goal programming

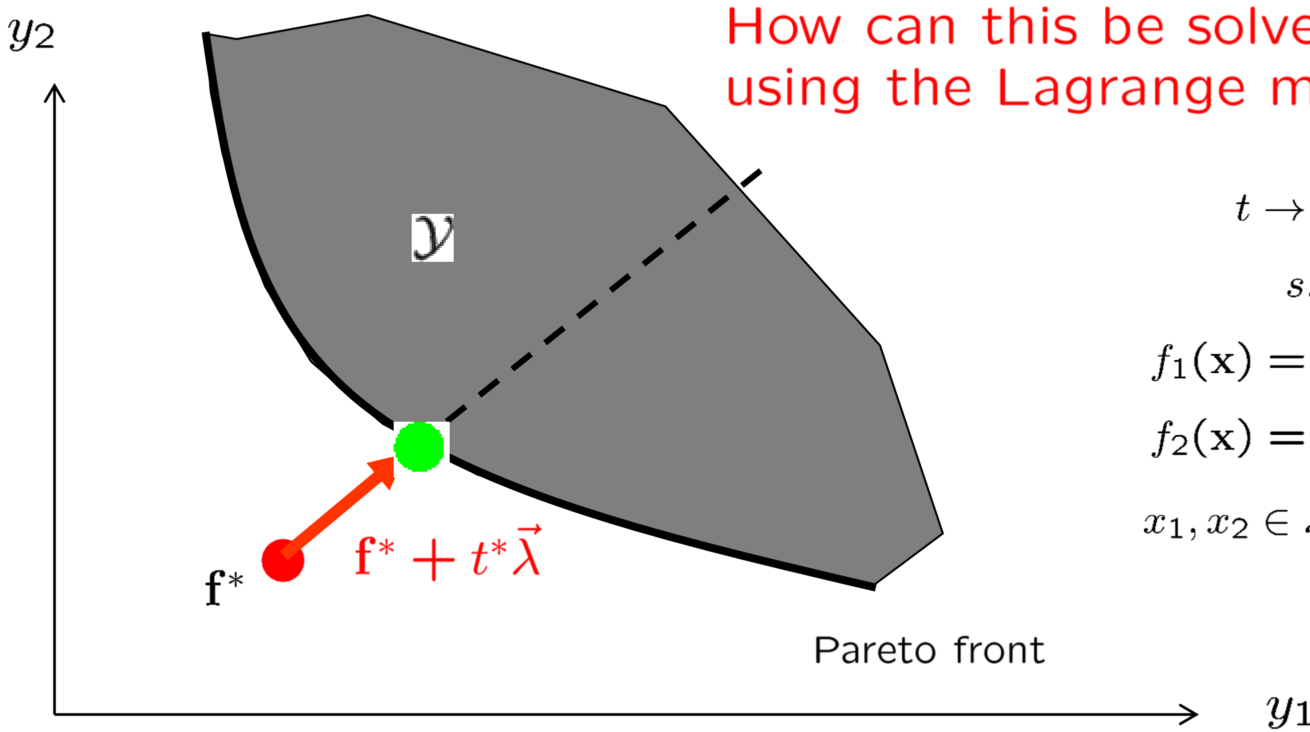
Algorithm:

Step 1: Choose reference point dominated or equal to ideal point

Step 2: Choose positive direction $\lambda \in \mathbf{R}^m_>$

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

How can this be solved by using the Lagrange multiplier rule?



$$t \to \min$$

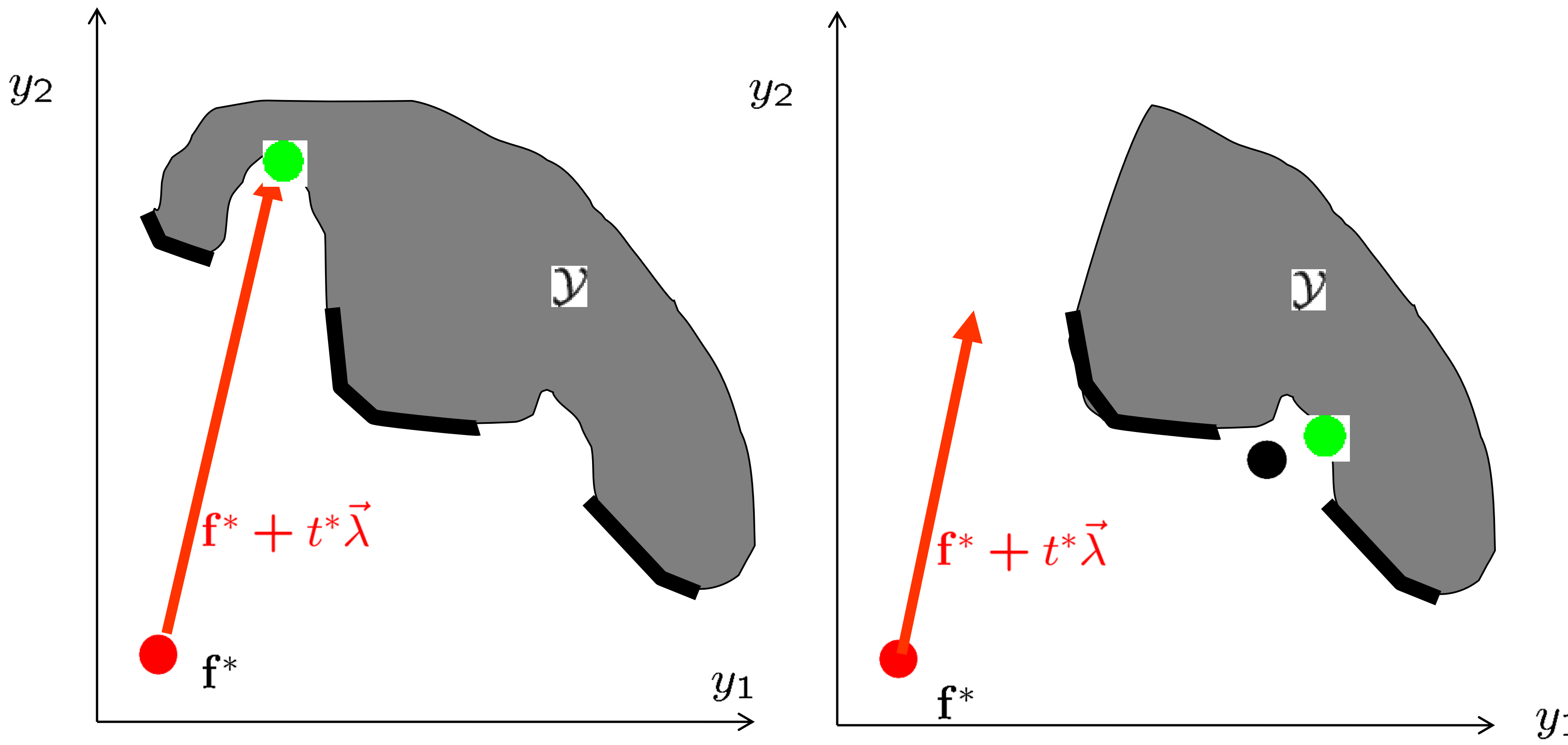
$$s.t.$$

$$f_1(\mathbf{x}) = f_1^* + t\lambda_1$$

$$f_2(\mathbf{x}) = f_2^* + t\lambda_2$$

$$x_1, x_2 \in \mathcal{X}, t \in \mathbb{R}_0^+$$

Goal programming, critics



Goal programming might result in dominated solutions! even, if reference point dominates ideal point.

Directions need to be chosen such that they meet the Pareto front..

Isolated points might be overlooked (right figure)

Take home messages: Single point methods

Scalarization methods can obtain Pareto optimal solutions.

[redacted]

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.