Table of contents

Four lectures course: 4x2 hrs

Lecture 1:

- Introduction to MOO framework and taxonomy
- Classical scalarization methods

Lecture 2:

- Meta-heuristic methods
- Simulated annealing
- Swarm Particle
- Genetic algorithms

Lecture 3:

- Evaluation metrics
- Python packages

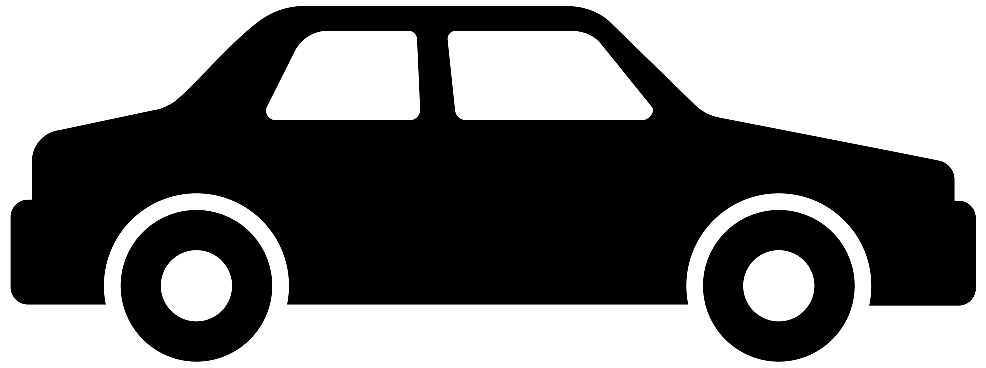
Lecture 4:

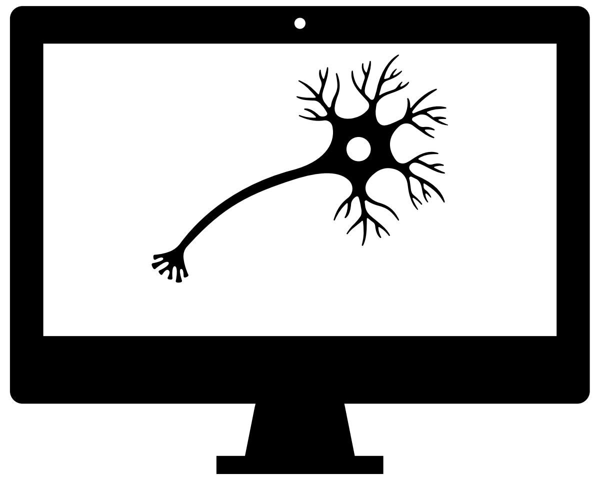
- Symbolic regression
- Multi-objective symbolic regression
- Case studies

Multi-Objective Optimization (MOO)

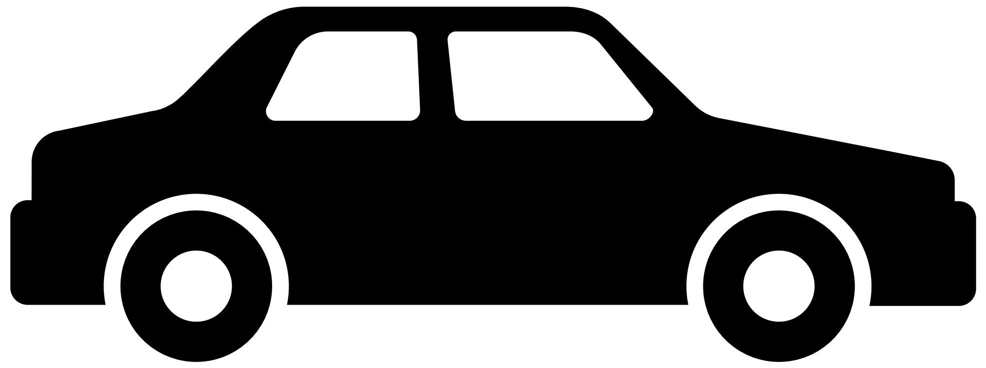
SCAN ME

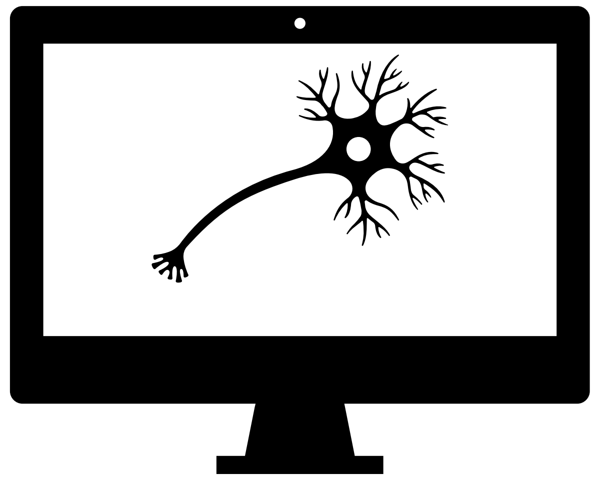
MOO in almost every real-world application

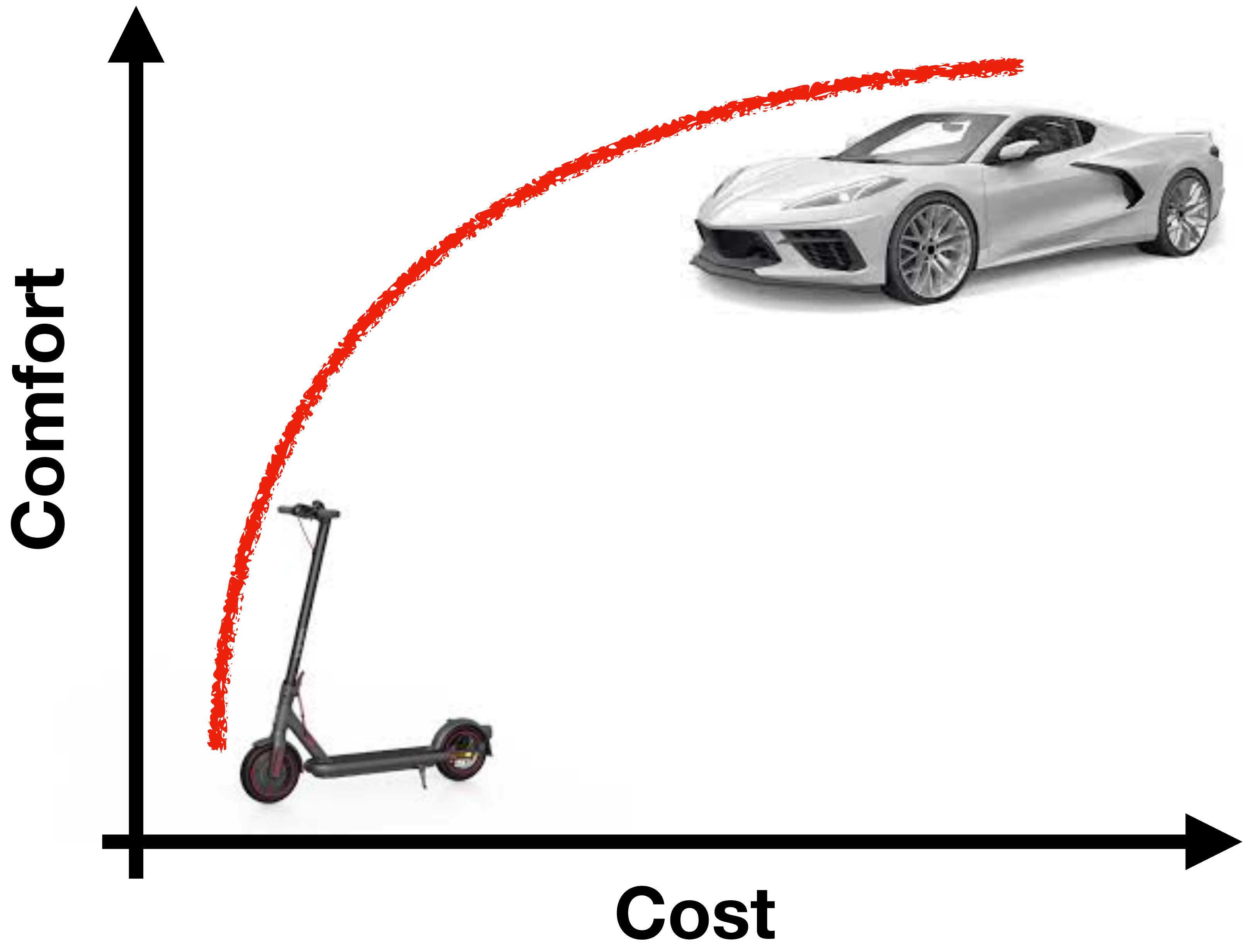




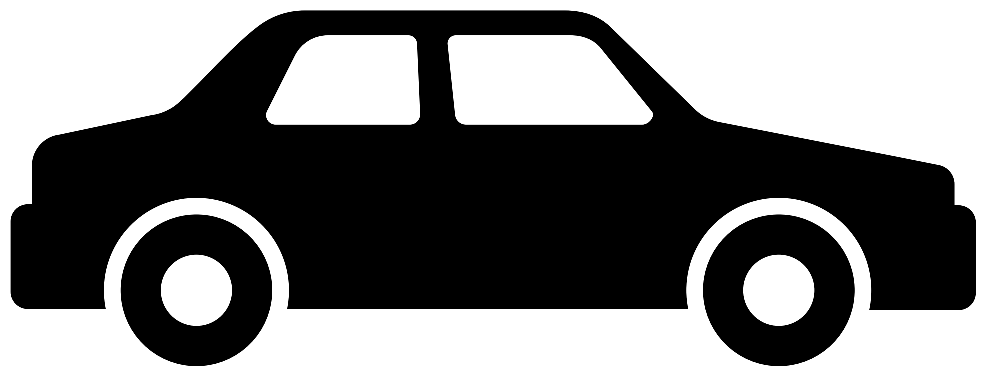
MOO in almost every real-world application

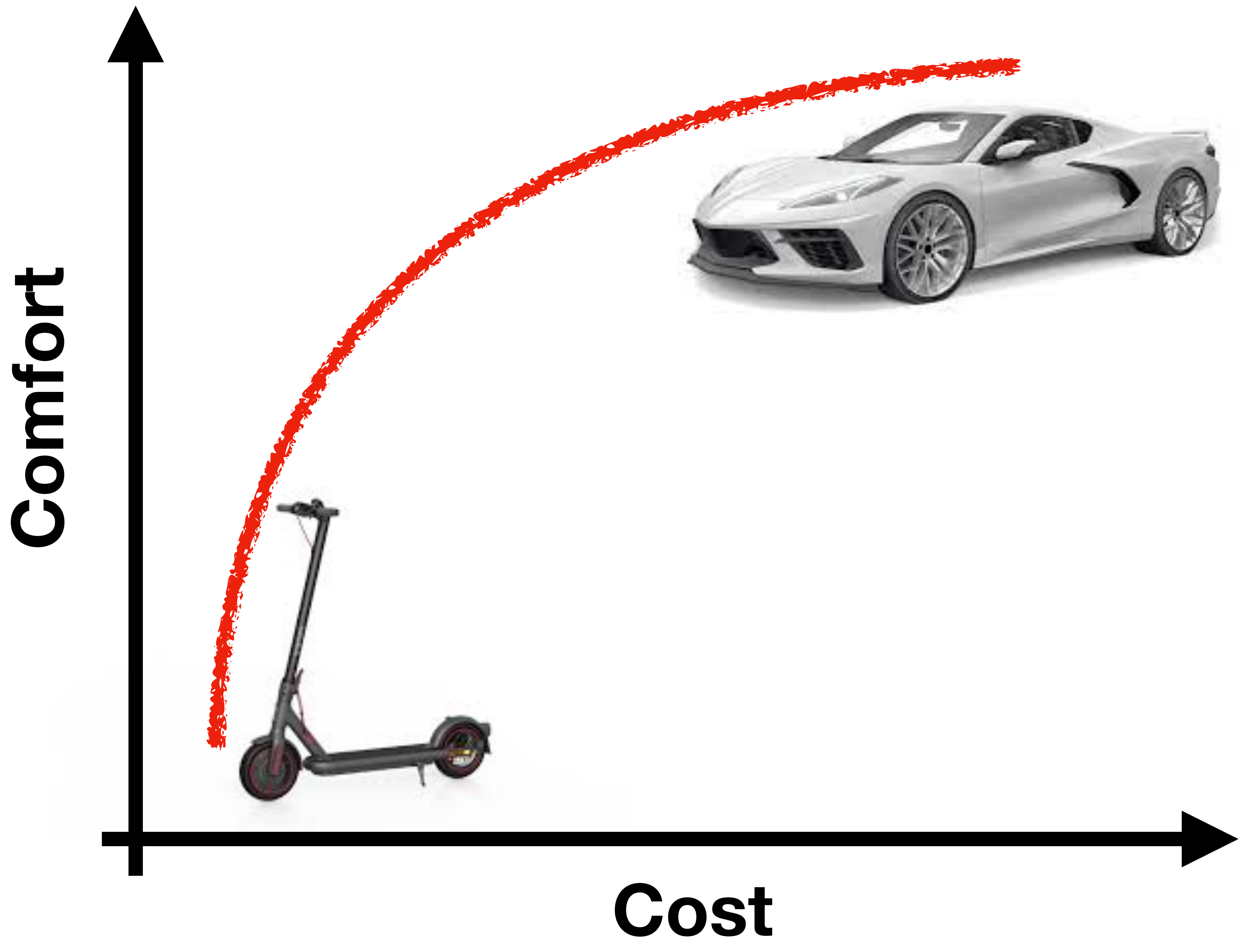


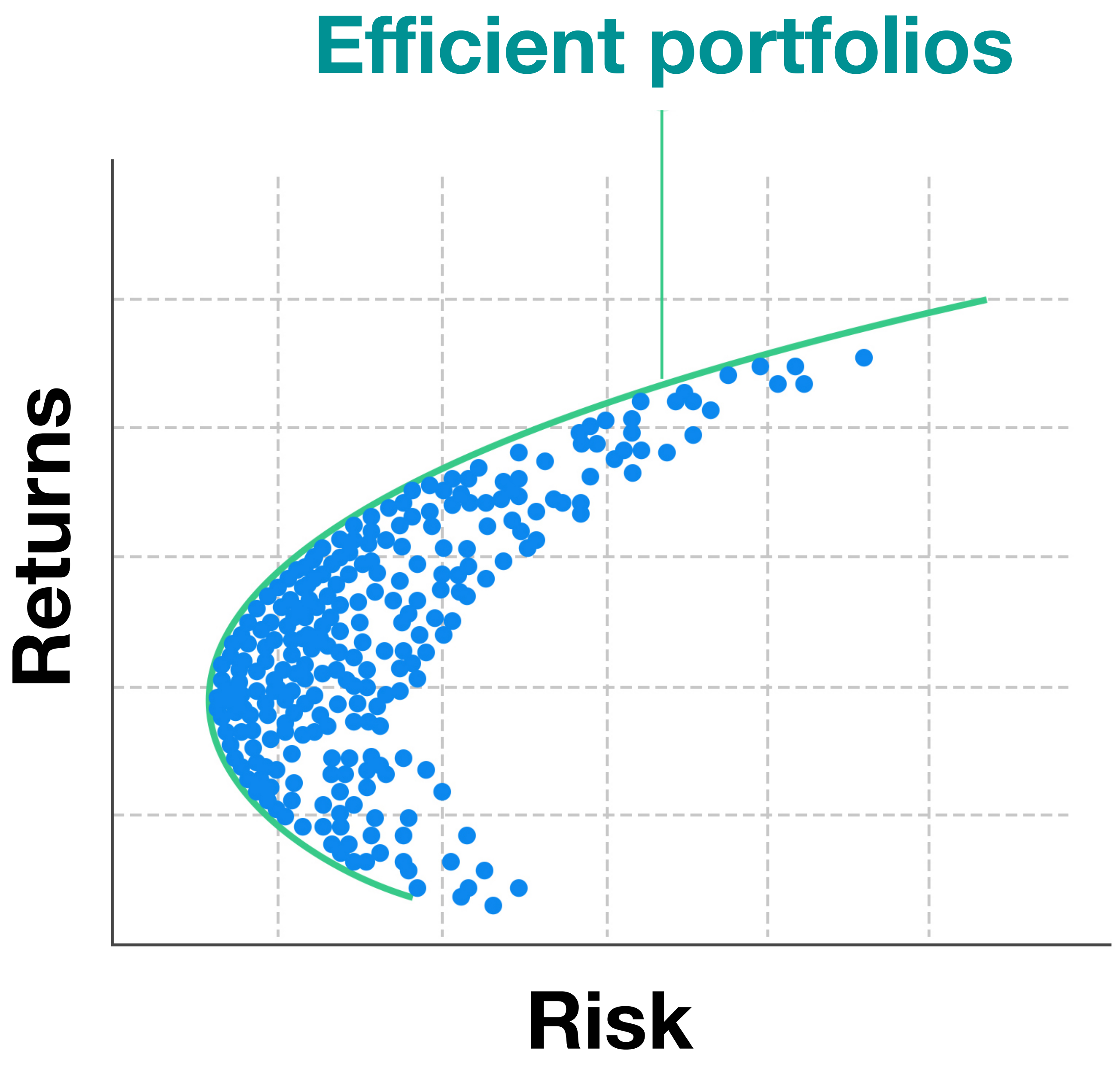


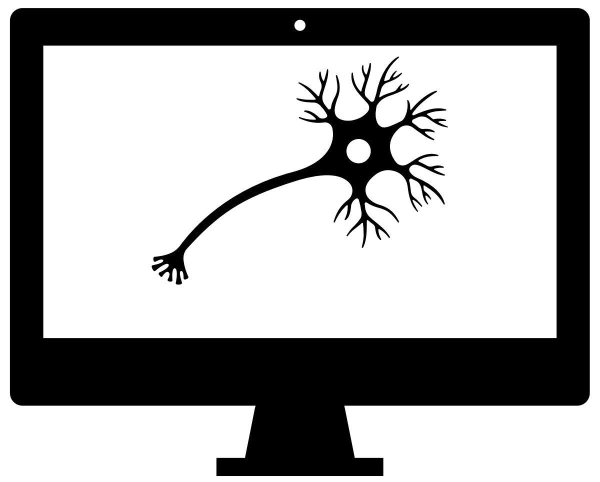


MOO in almost every real-world application

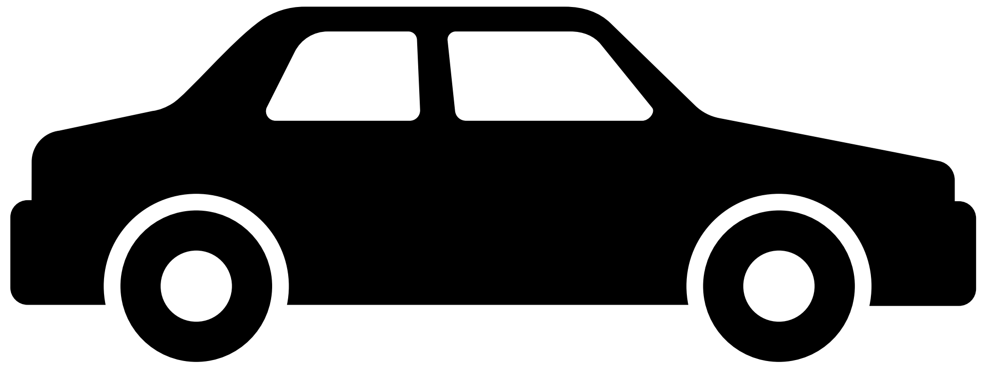


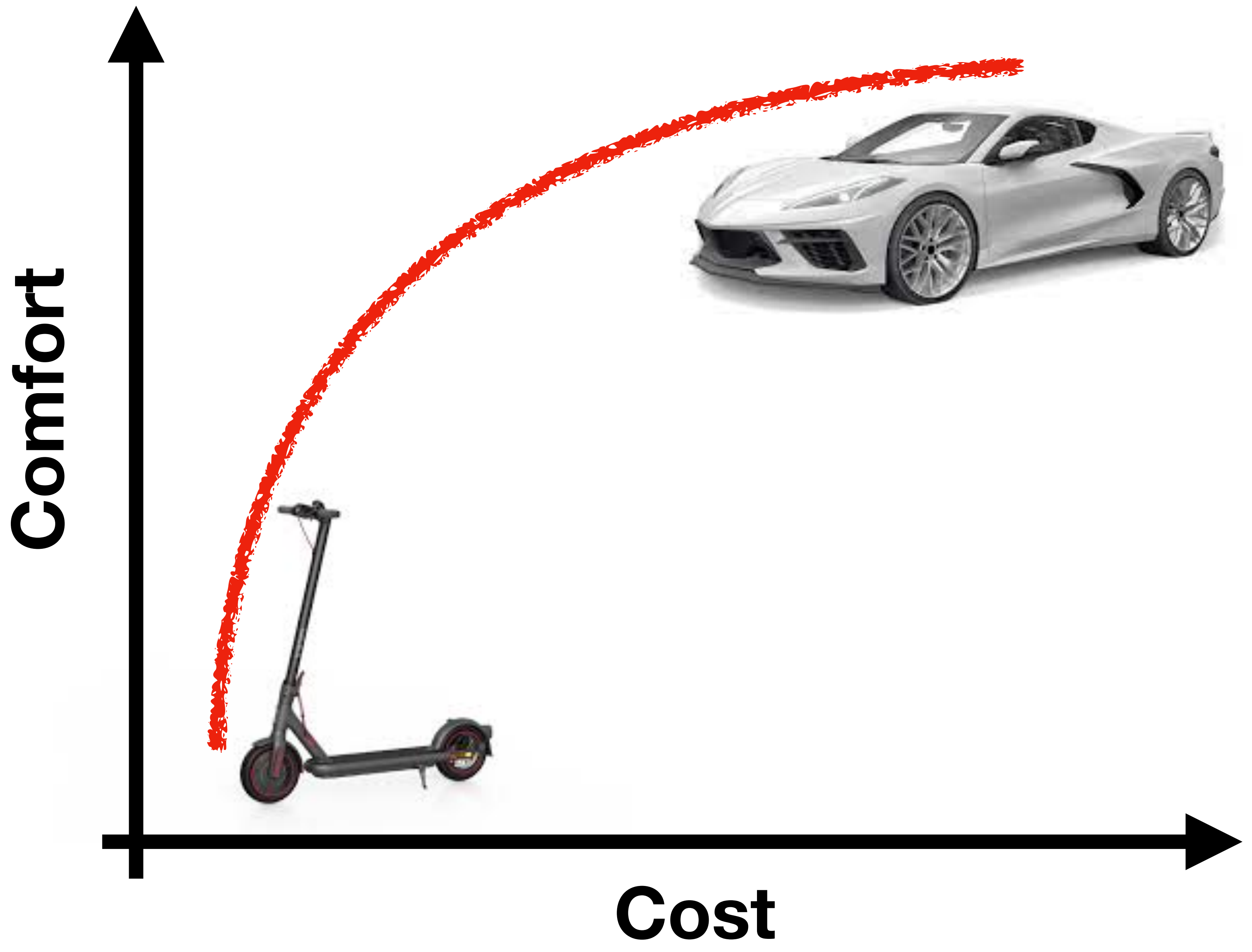


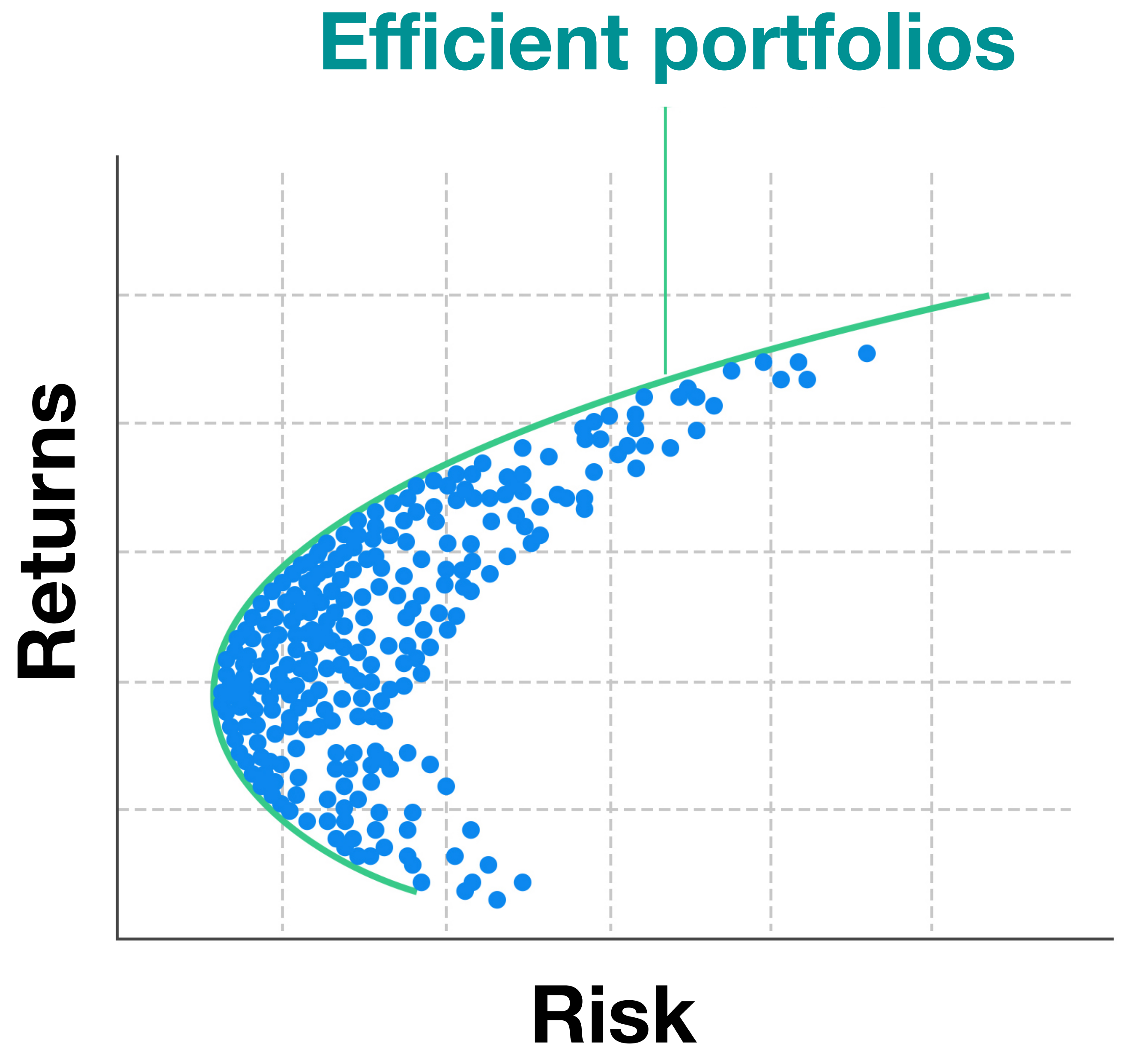


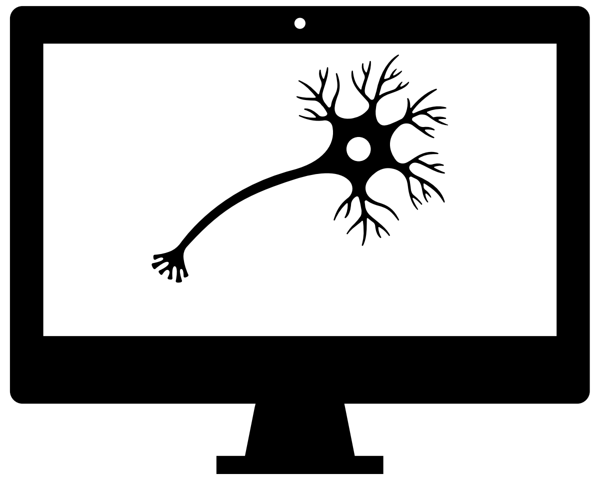


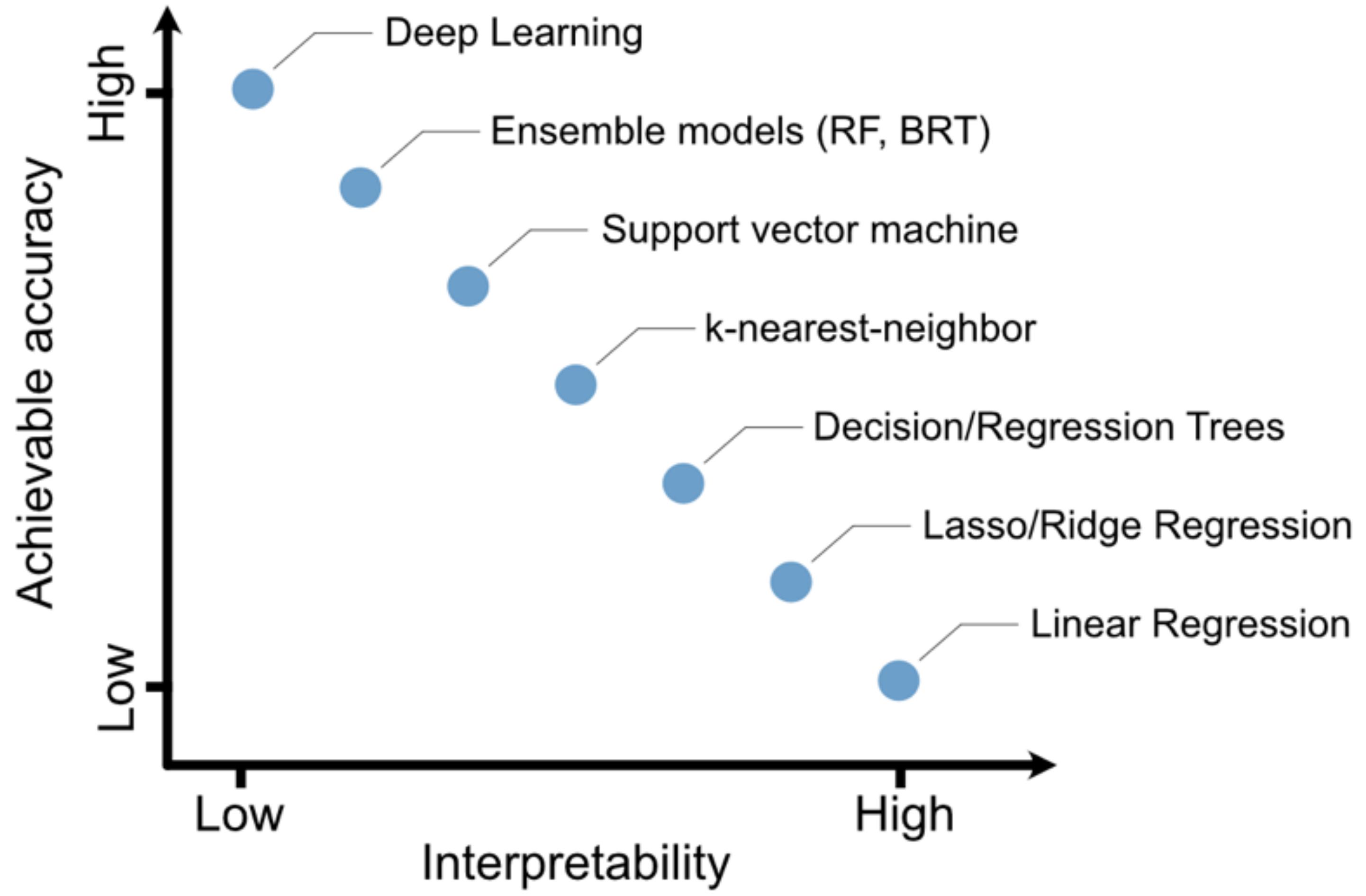
MOO in almost every real-world application



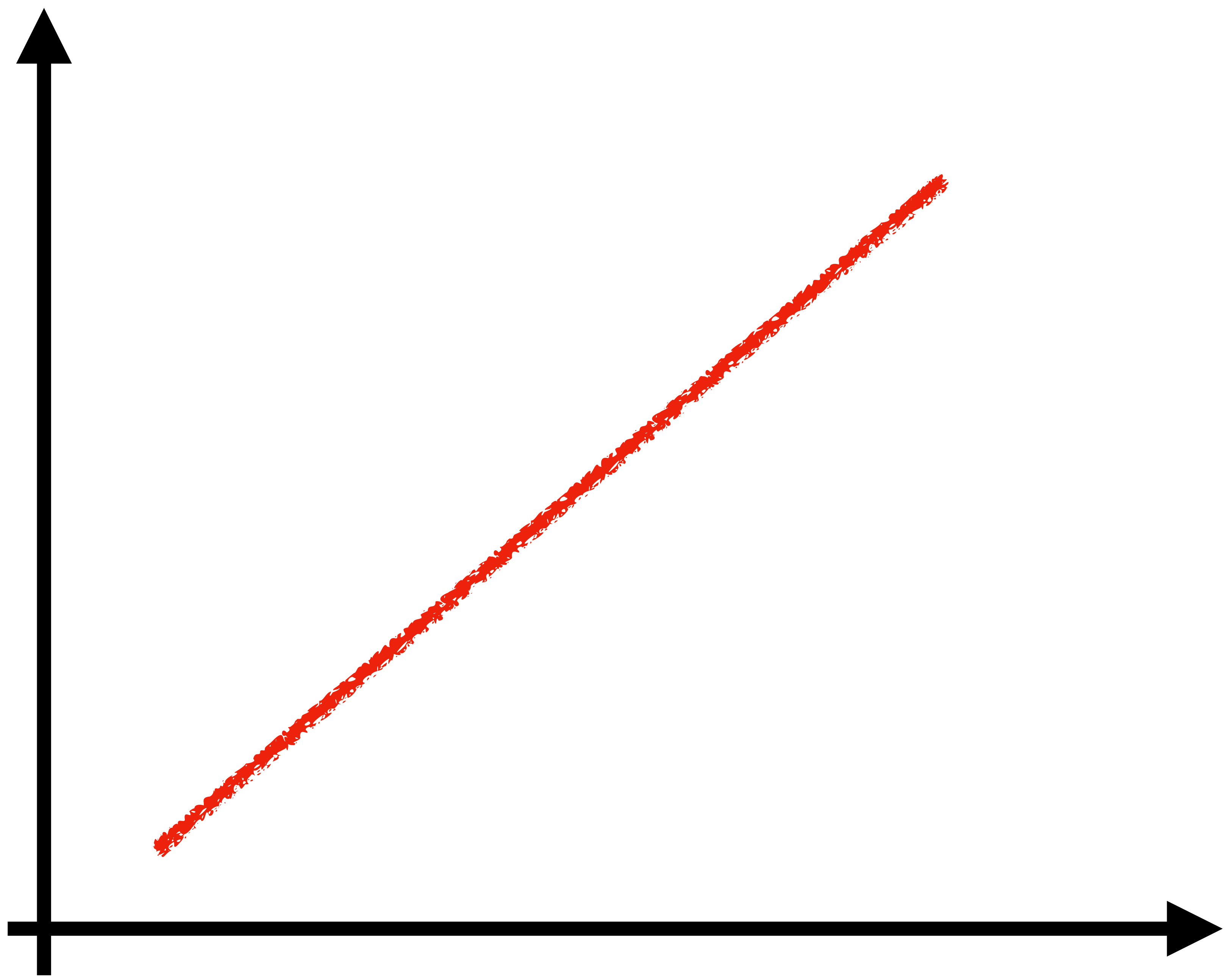




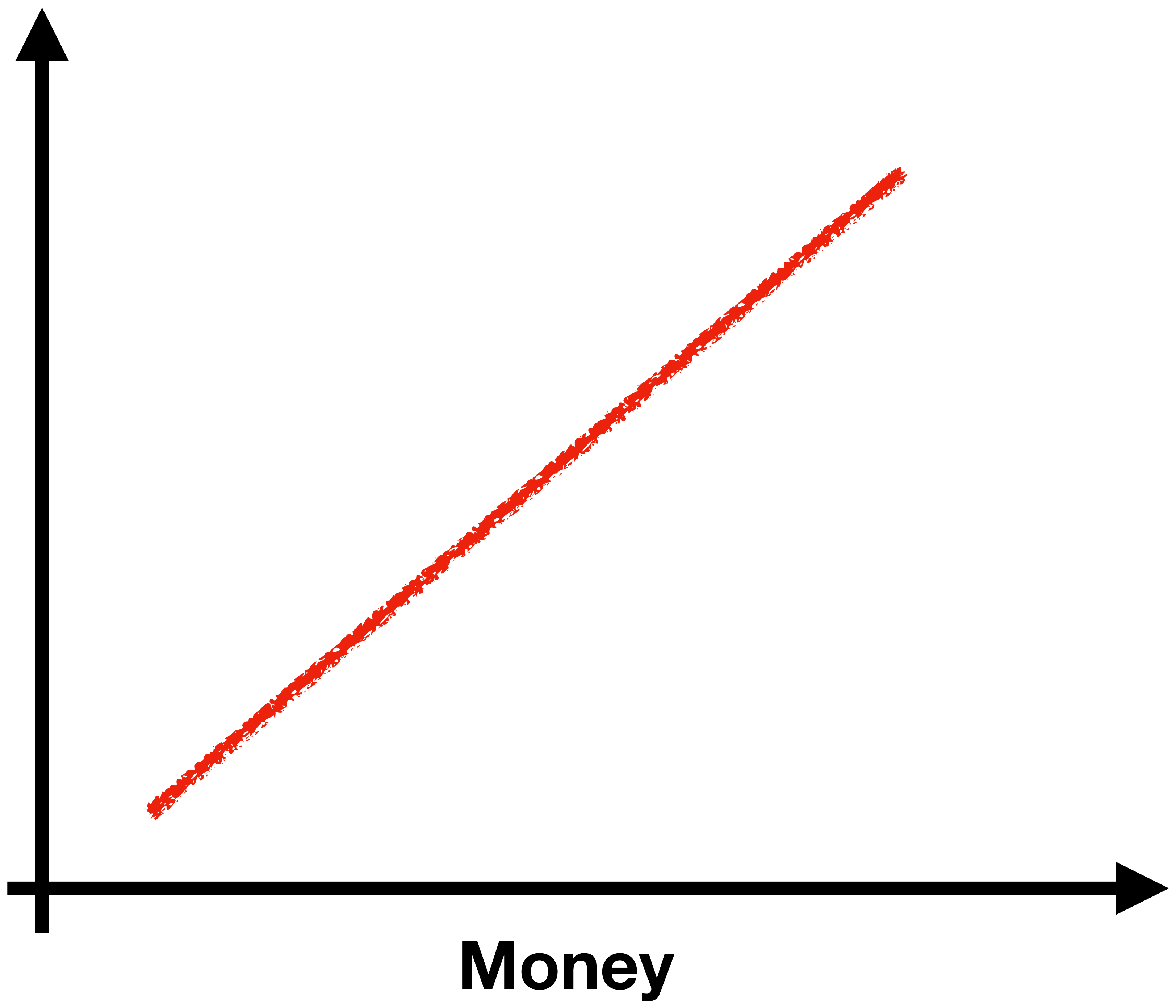




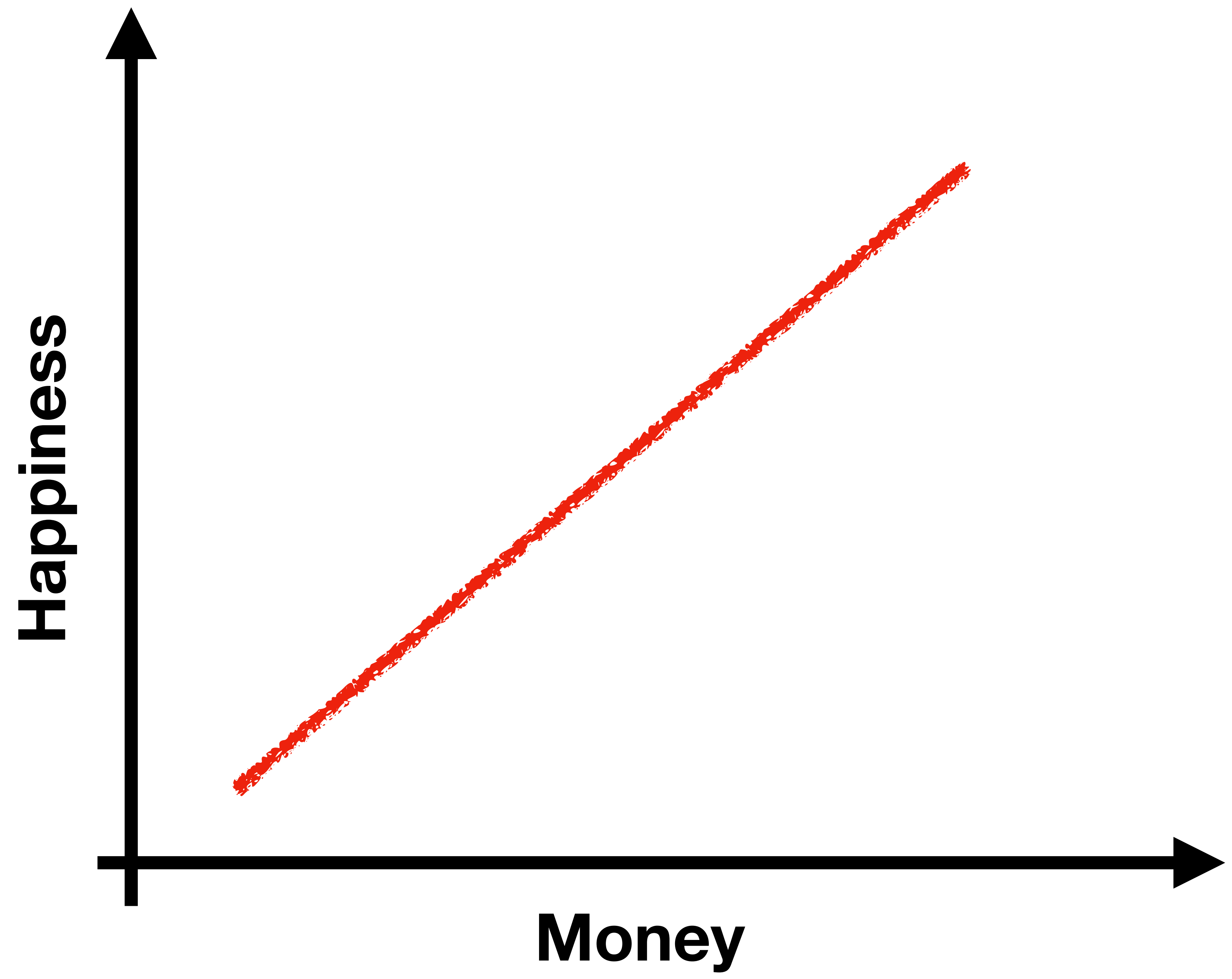
Objectives conflict to some degree



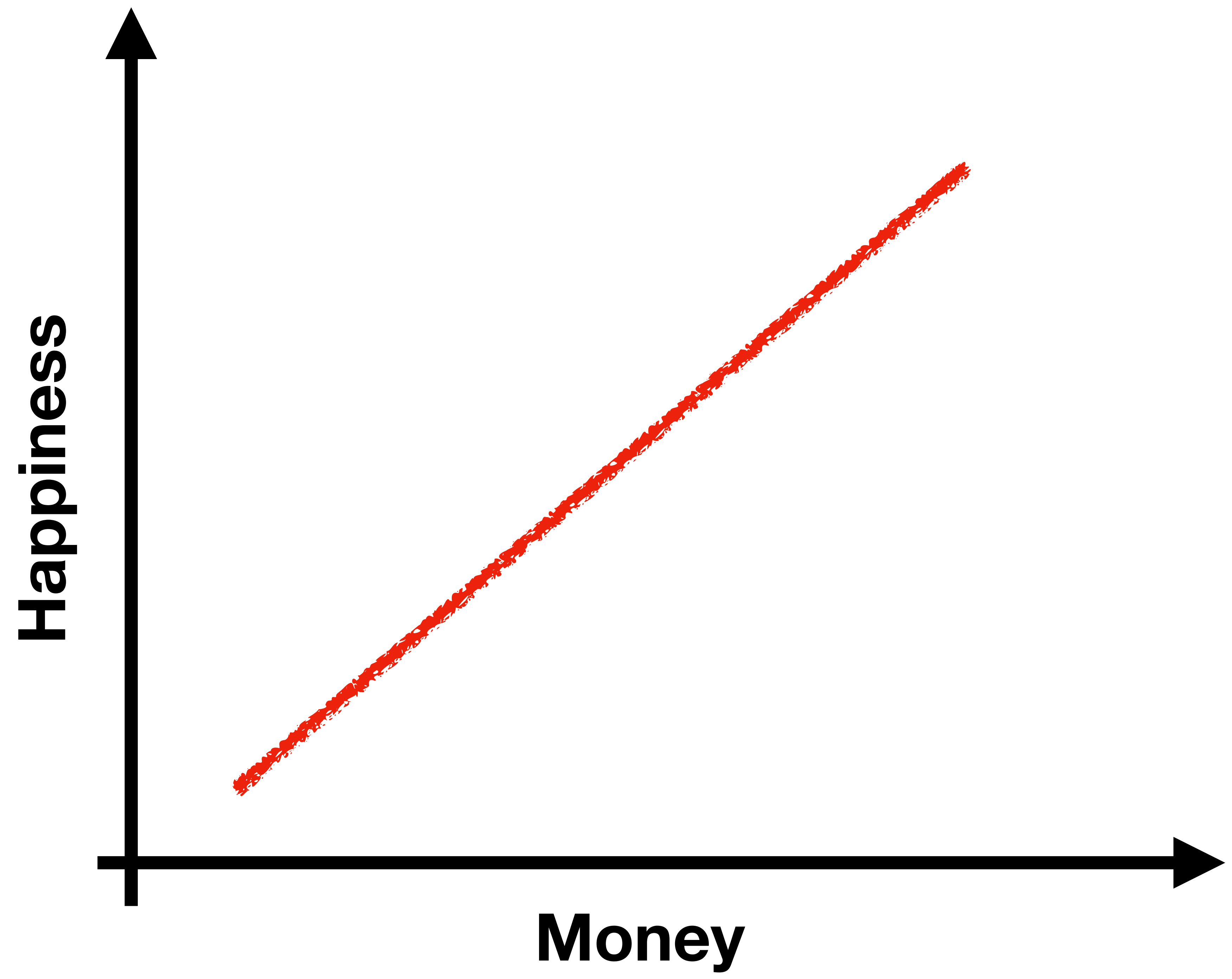
Objectives conflict to some degree

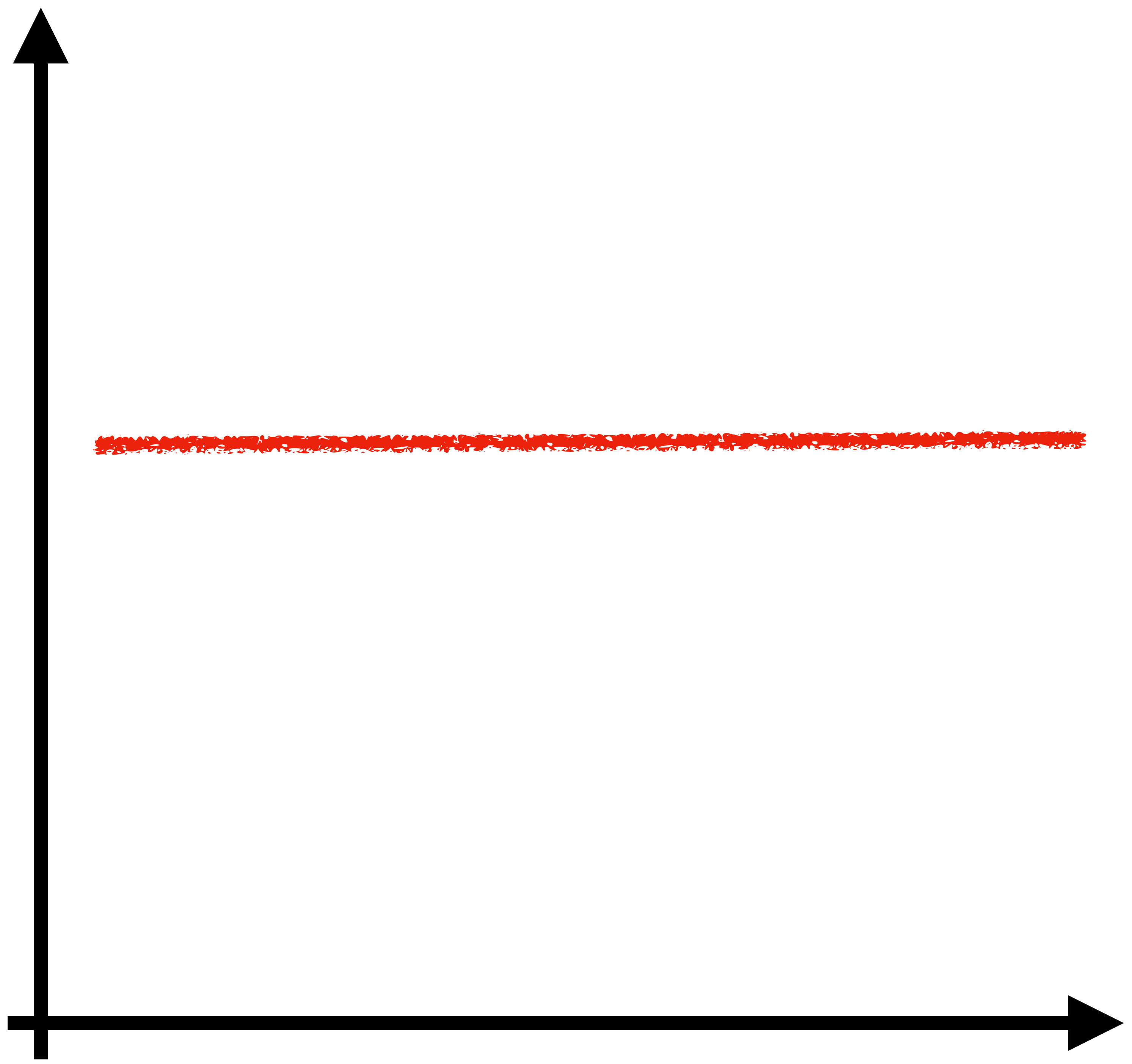


Objectives conflict to some degree

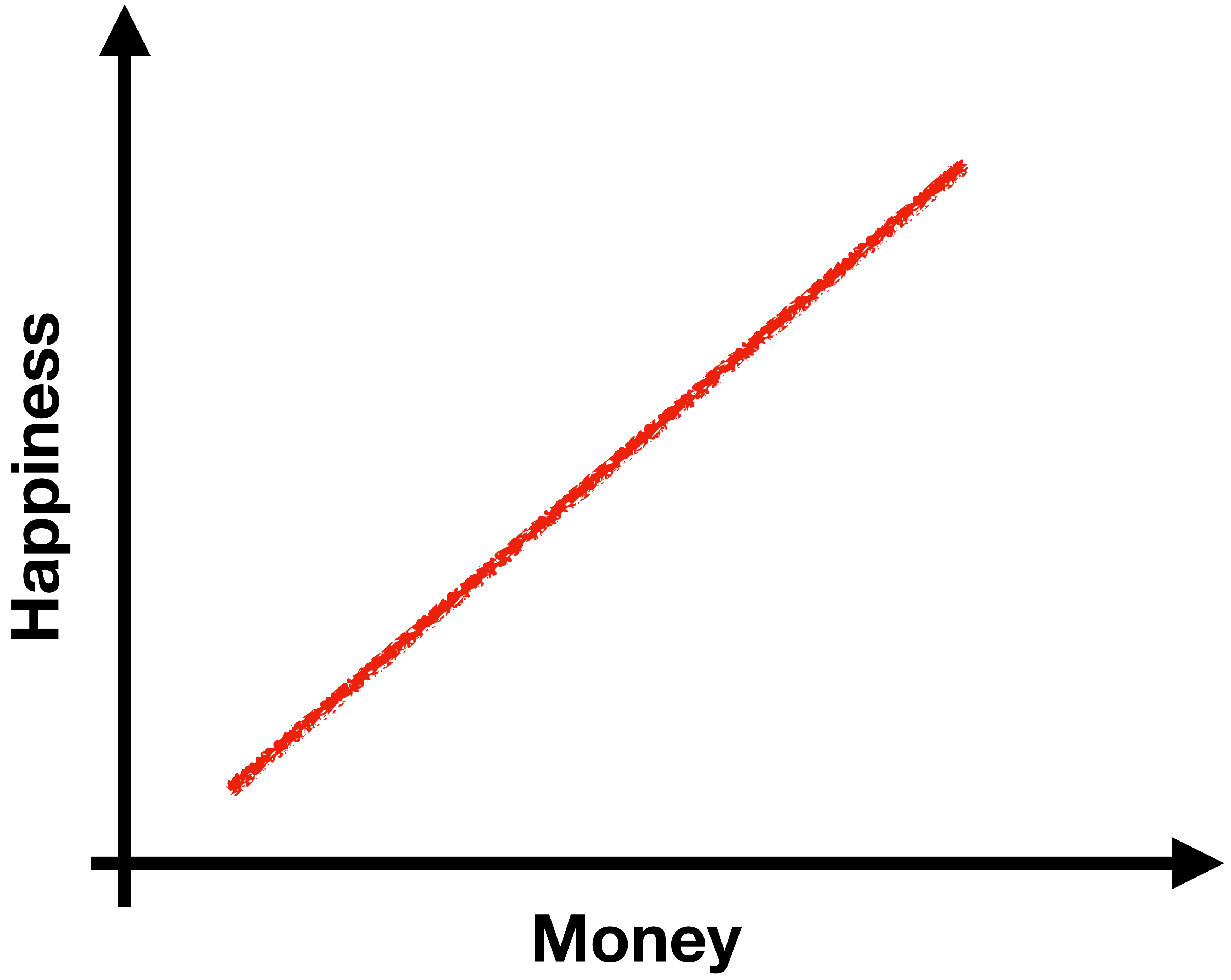


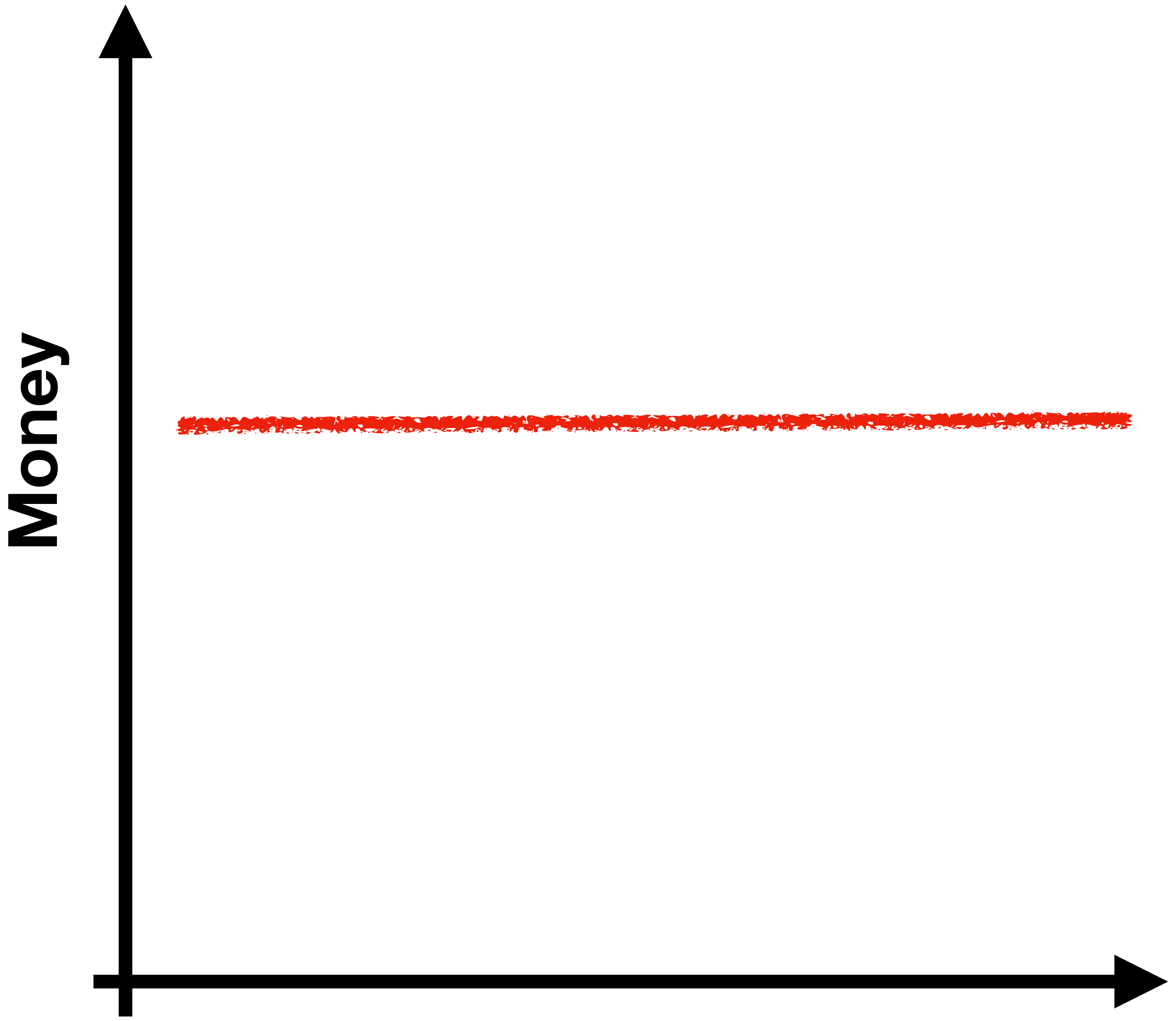
Objectives conflict to some degree



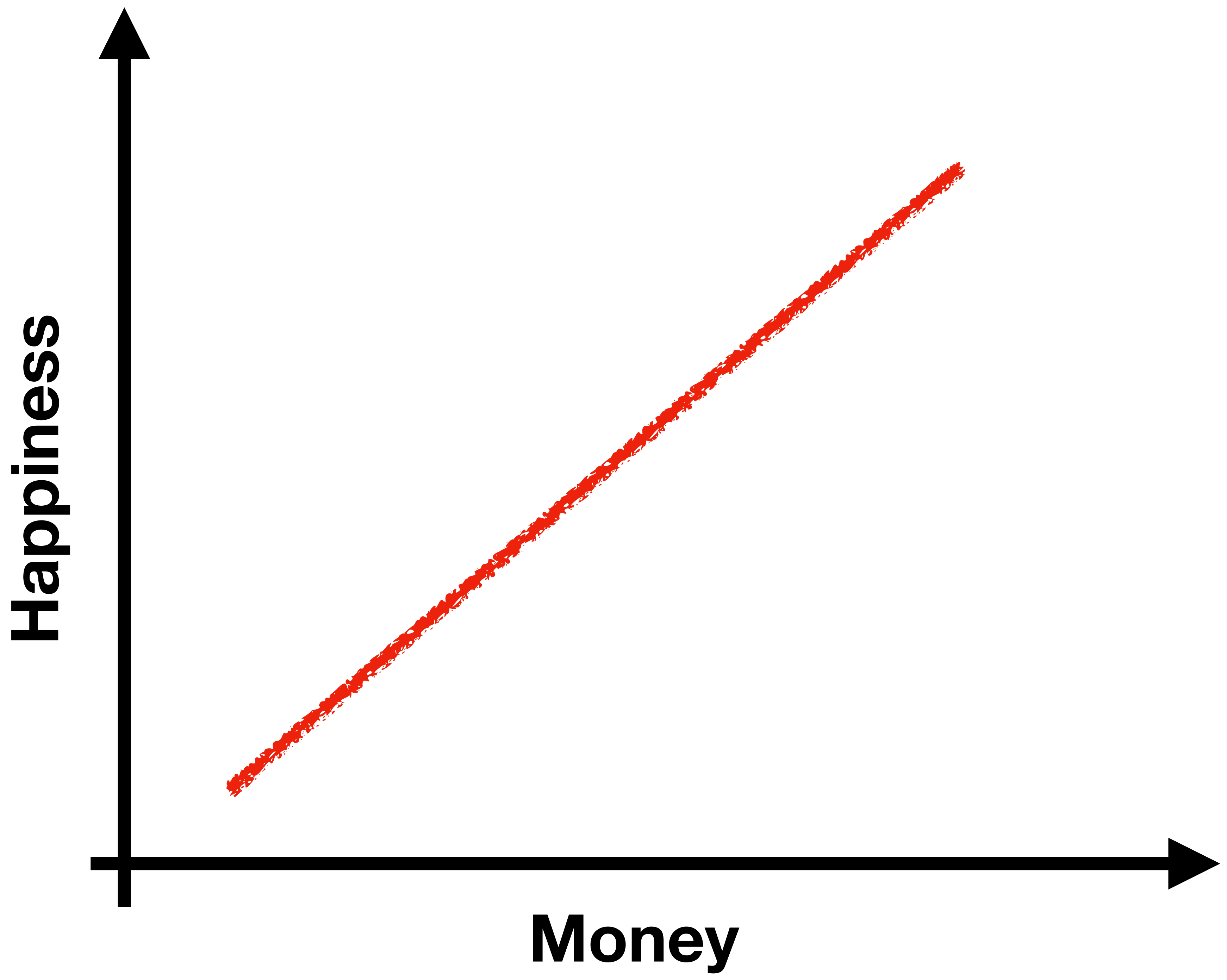


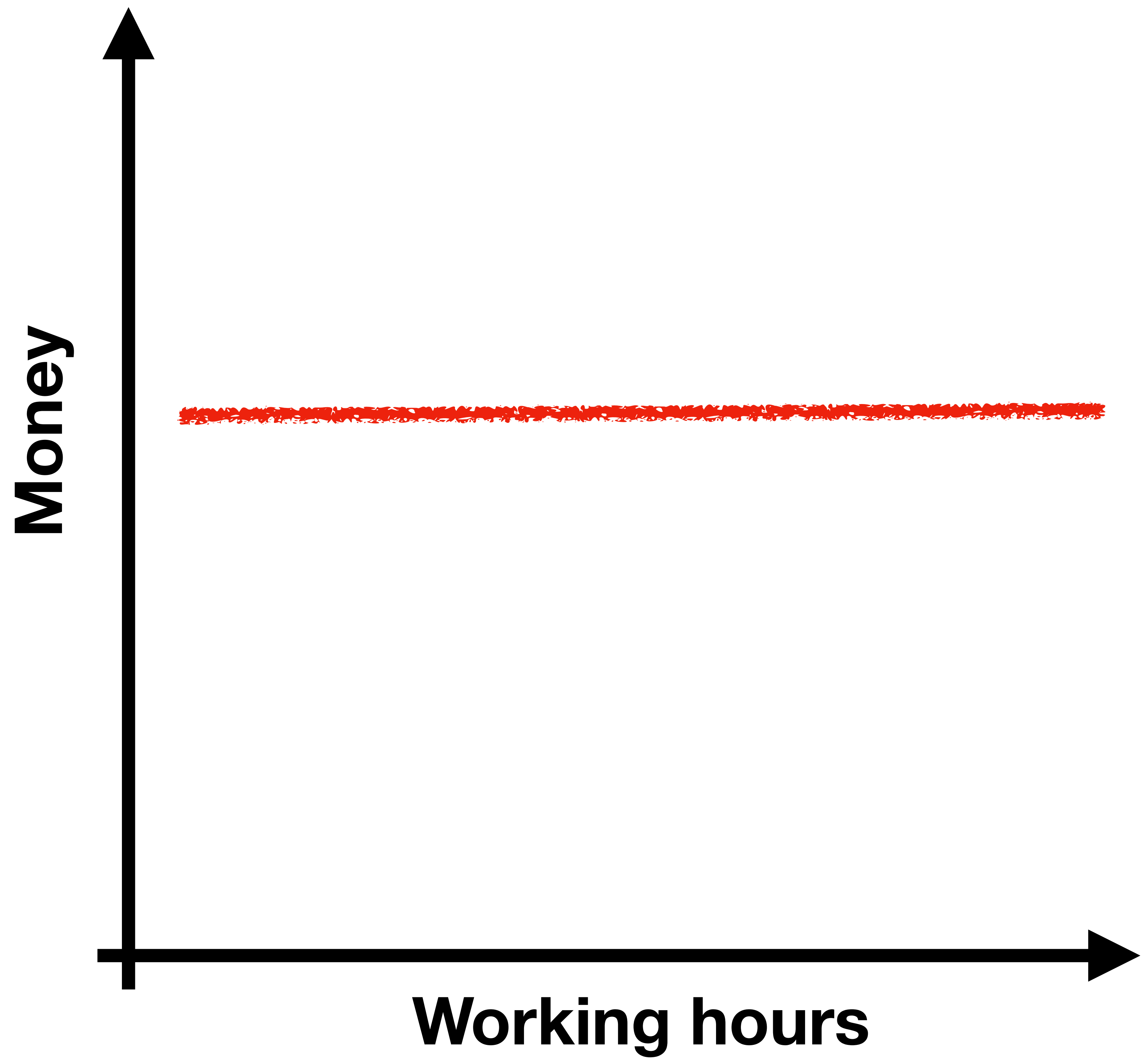
Objectives conflict to some degree





Objectives conflict to some degree





... at least in Academia

MOO getting famous?

Getting much better over the years…

MOO getting famous?

Getting much better over the years…

… clearly not as much as deep learning …

Founders of MOO

Mathematicians and Economists

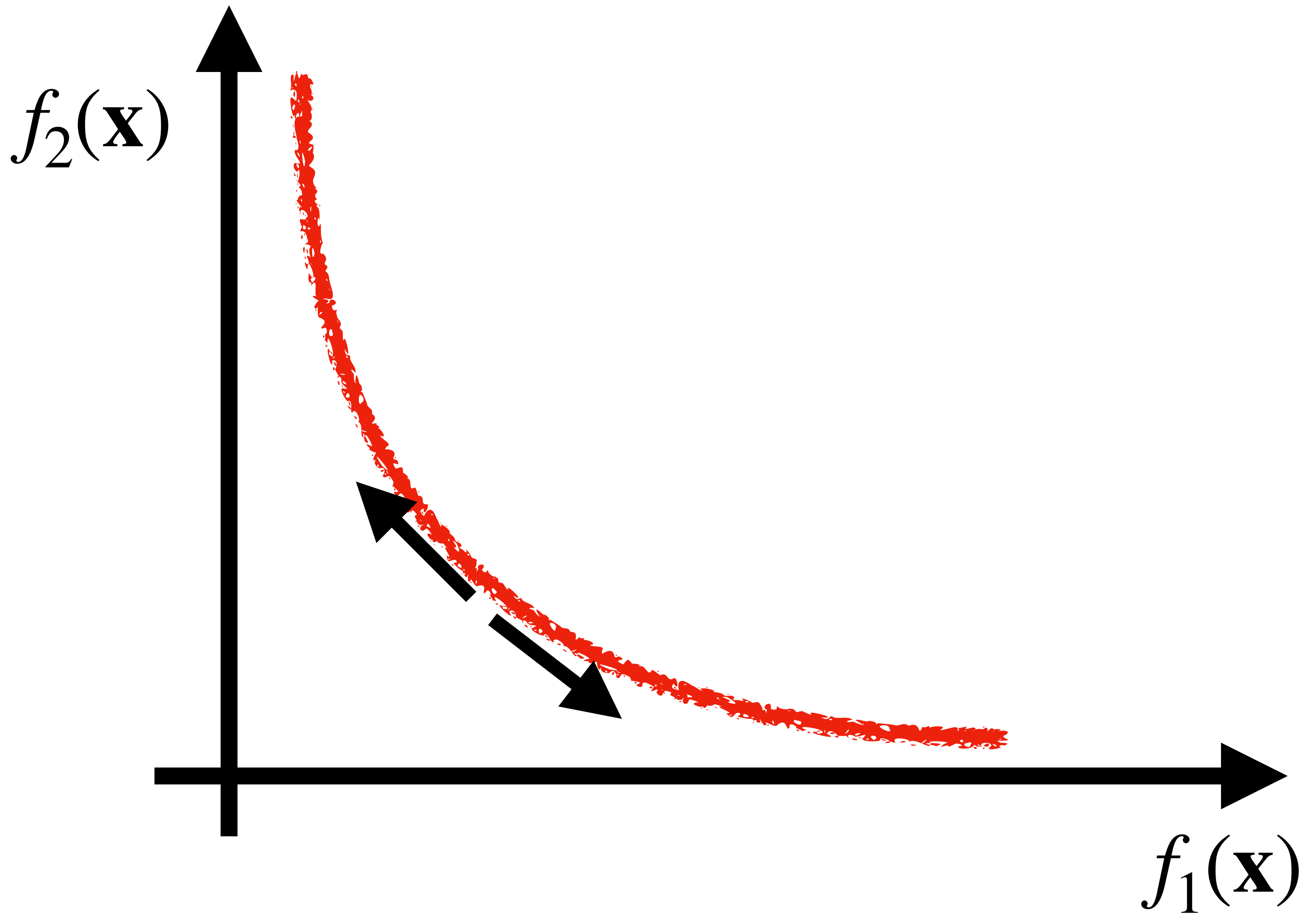
- **Francis Ysidro Edgeworth (1845-1926)**

  Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, published in 1881

  "It is required to find a point (x, y) such that, in whatever direction we take an infinitely small step, $f1$ and $f2$ do not decrease together, but that, while one increases, the other decreases"

Founders of MOO

Mathematicians and Economists



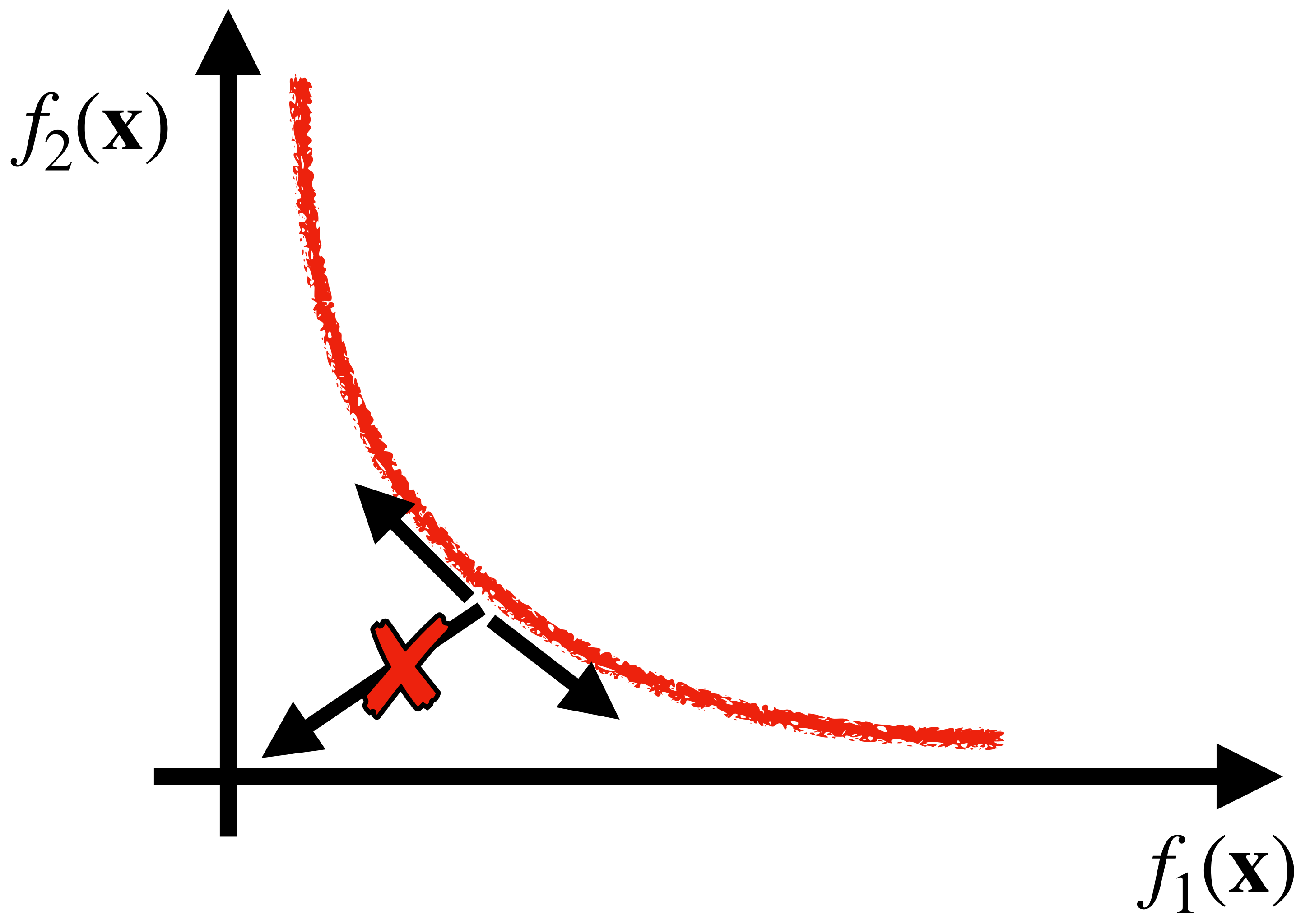
- **Francis Ysidro Edgeworth (1845-1926)**

Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, published in 1881

"It is required to find a point (x, y) such that, in whatever direction we take
an infinitely small step, $f1$ and $f2$ do not decrease together, but that, while one increases, the other decreases"

Founders of MOO

Mathematicians and Economists



- **Francis Ysidro Edgeworth (1845-1926)**

  Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, published in 1881

  “It is required to find a point (x, y) such that, in whatever direction we take an infinitely small step, $f1$ and $f2$ do not decrease together, but that, while one increases, the other decreases”

Founders of MOO

Mathematicians and Economists

- **Francis Ysidro Edgeworth (1845-1926)**

  Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, published in 1881

  "It is required to find a point (x, y) such that, in whatever direction we take an infinitely small step, $f1$ and $f2$ do not decrease together, but that, while one increases, the other decreases"

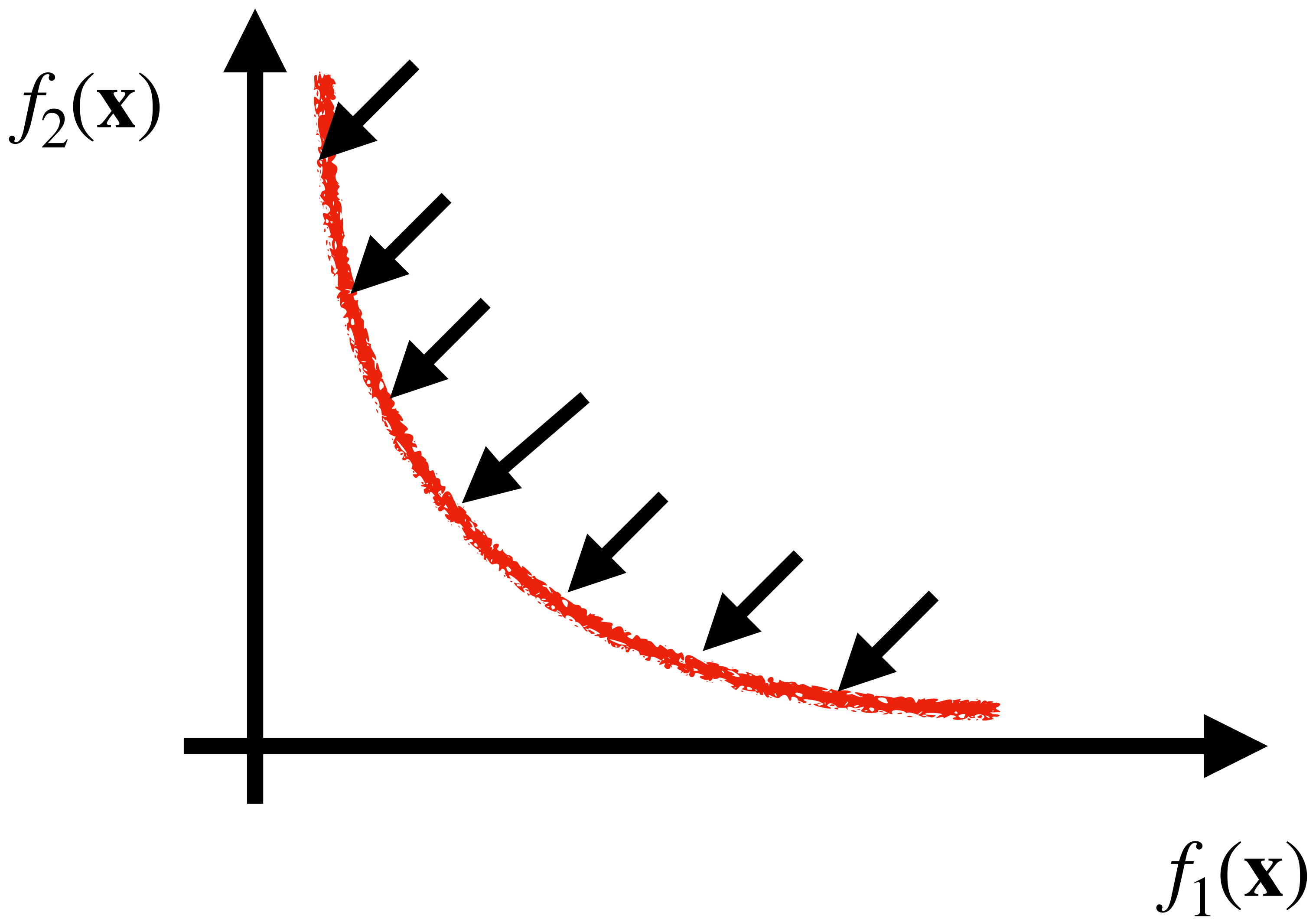
- **Vilfredo Pareto (1848-1923)**

  Manual of Political Economy, published in 1906

  "The optimum allocation of the resources of a society is not attained so long as it is possible to make at least one individual better off in his own estimation while keeping others as well off as before in their own estimation."

Founders of MOO

Mathematicians and Economists



- **Francis Ysidro Edgeworth (1845-1926)**

  Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences, published in 1881

  “It is required to find a point (x, y) such that, in whatever direction we take an infinitely small step, $f1$ and $f2$ do not decrease together, but that, while one increases, the other decreases”

- **Vilfredo Pareto (1848-1923)**

  Manual of Political Economy, published in 1906

  “The optimum allocation of the resources of a society is not attained so long as it is possible to make at least one individual better off in his own estimation while keeping others as well off as before in their own estimation.”

MOO definition

- Decision variables and domain $\mathbf{x} \in \mathbf{R}^n$ $\mathscr{D} = \left\{ \mathbf{x} \in \mathbf{R}^n : x_i^L \le x_i \le x_i^U \; ; i = 1,\ldots,n \right\}$
- Objective functions $(f_1, f_2, \ldots, f_M) \in \mathbf{R}^M$ $f_i(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$
- Constraints $g_j(\mathbf{x}),\, h_k(\mathbf{x}) : \mathbf{R}^n \to \mathbf{R}$
- Problem

$$\min_{\mathbf{x}} \left(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x}) \right)$$

$$\text{s.t.} \quad g_j(\mathbf{x}) \ge 0 \; ; \quad h_k(\mathbf{x}) = 0 \; ; \quad \mathbf{x} \in \mathscr{D}$$

- Feasible solutions

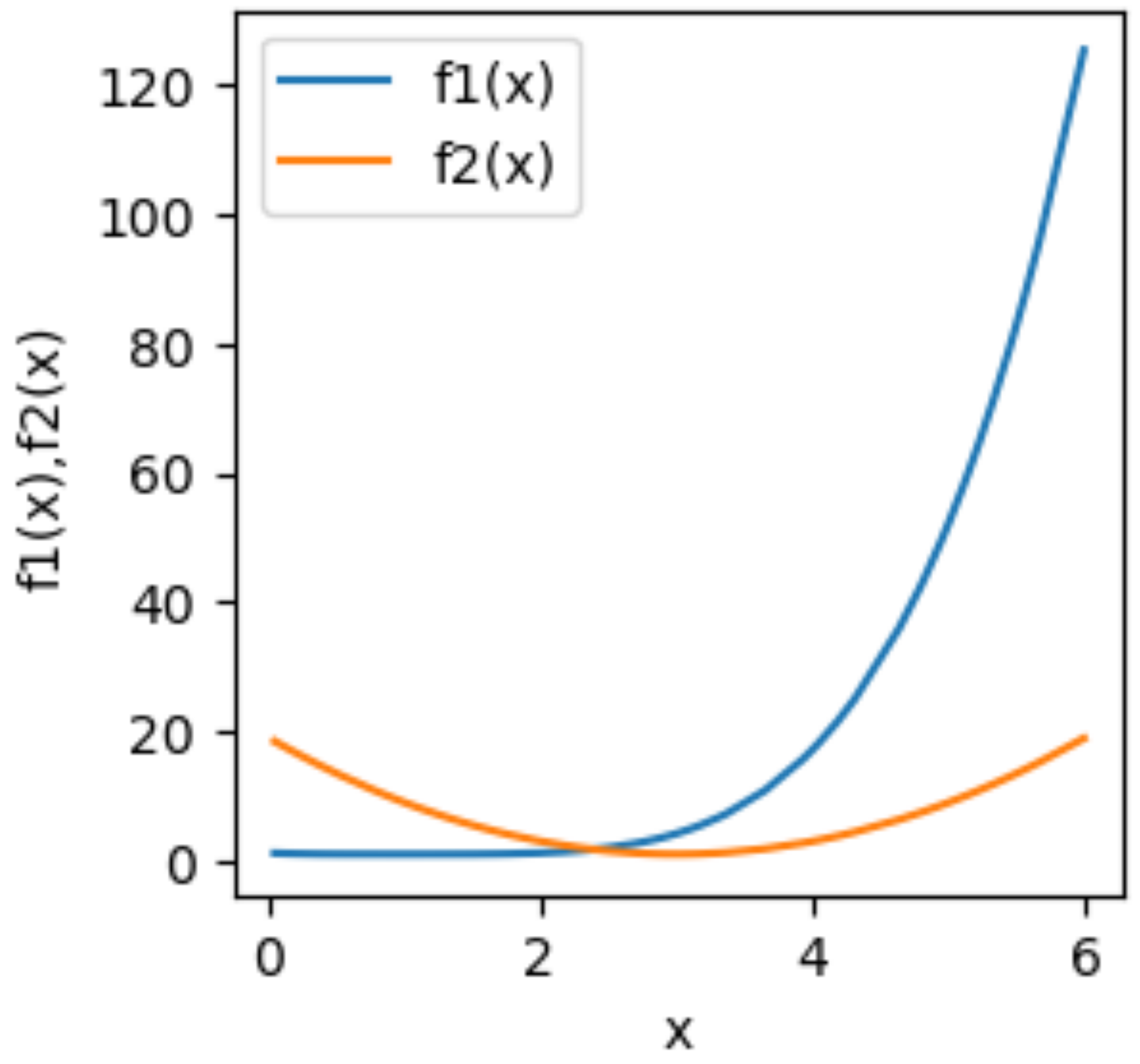
$$S = \left\{ \mathbf{x} \in \mathbf{R}^n : x_i^L \le x_i \le x_i^U \cap g_j(\mathbf{x}) \ge 0 \; \cap h_k(\mathbf{x}) = 0 \right\}$$

Example

$$f_1(x) = \frac{1}{5}(x-1)^4 + 1$$

$$f_2(x) = 2(x-3)^2 + 1$$

$$x \in [0,6] \subset \mathbf{R}$$

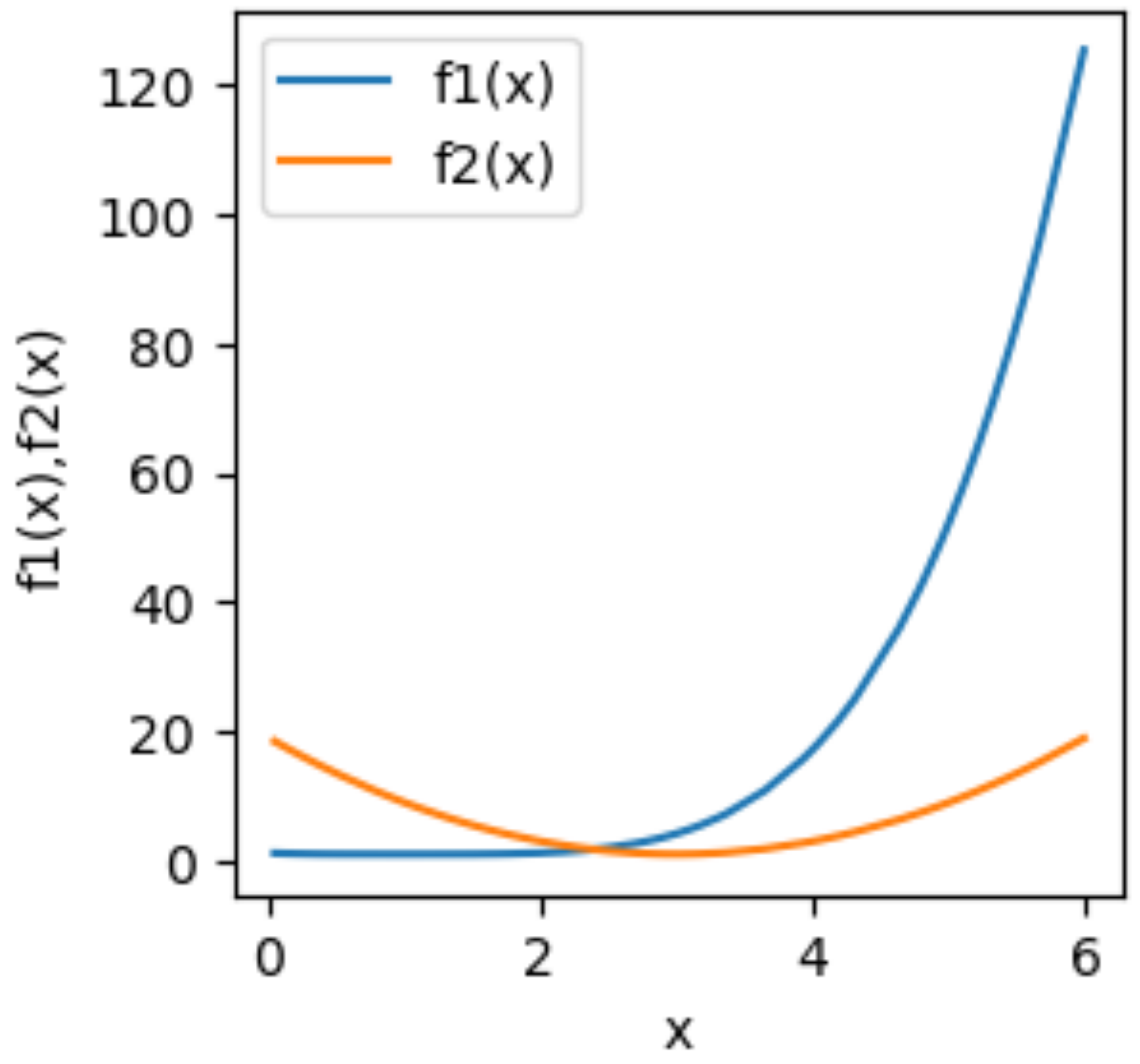


Example

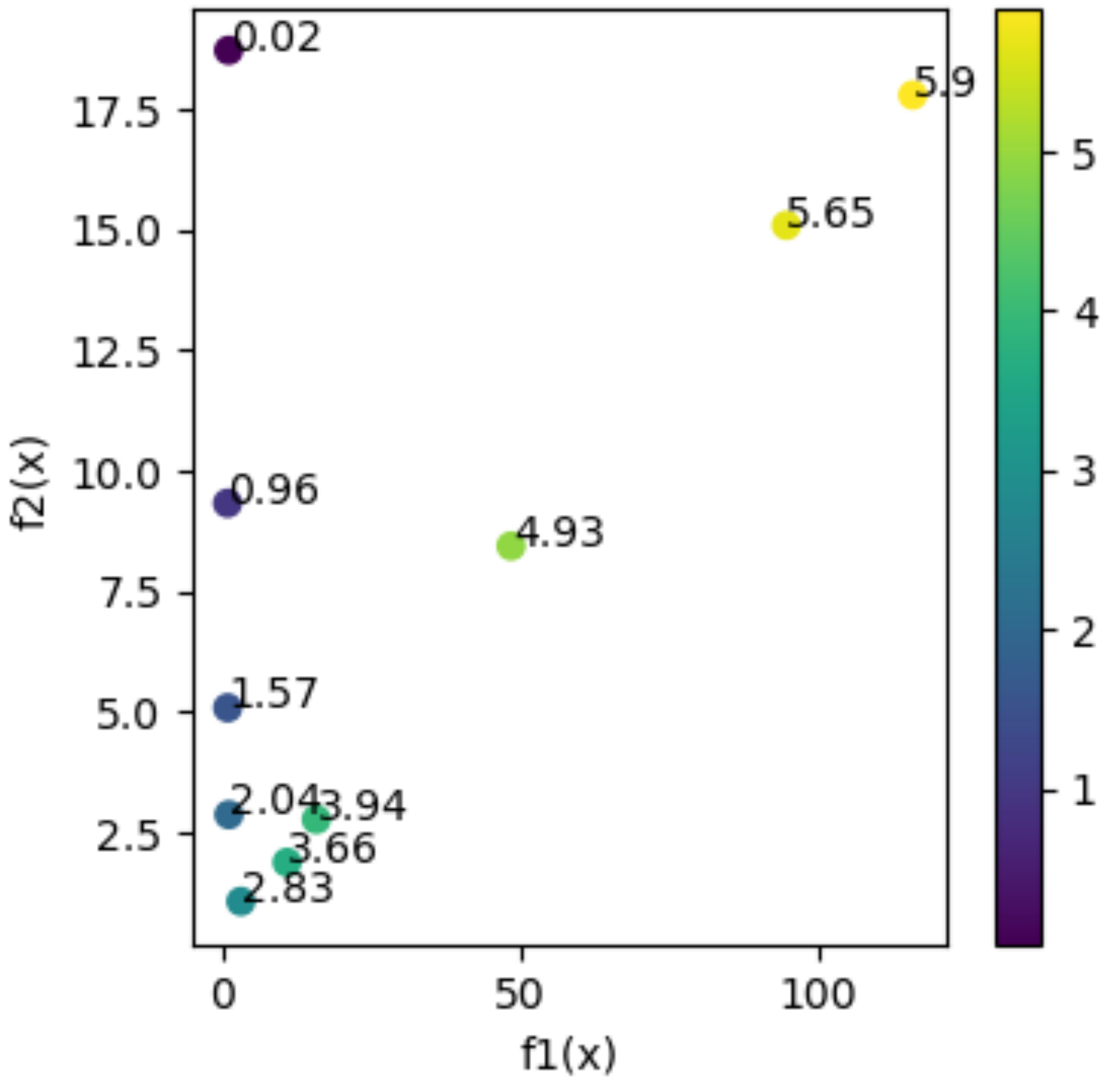
$$f_1(x) = \frac{1}{5}(x-1)^4 + 1$$

$$f_2(x) = 2(x-3)^2 + 1$$

$$x \in [0,6] \subset \mathbf{R}$$



Visualization in low dimensions: from solution to objective space

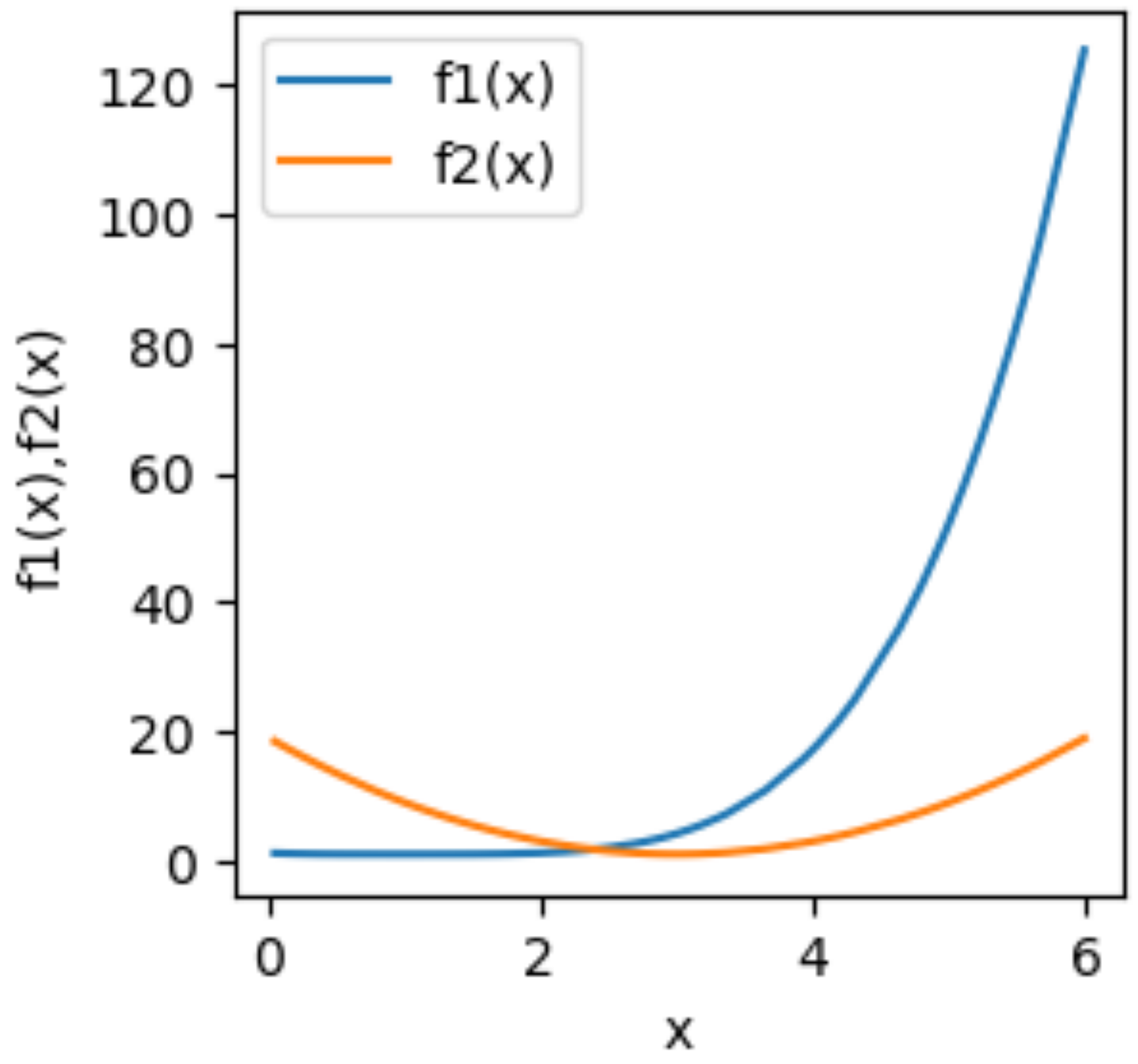


Example

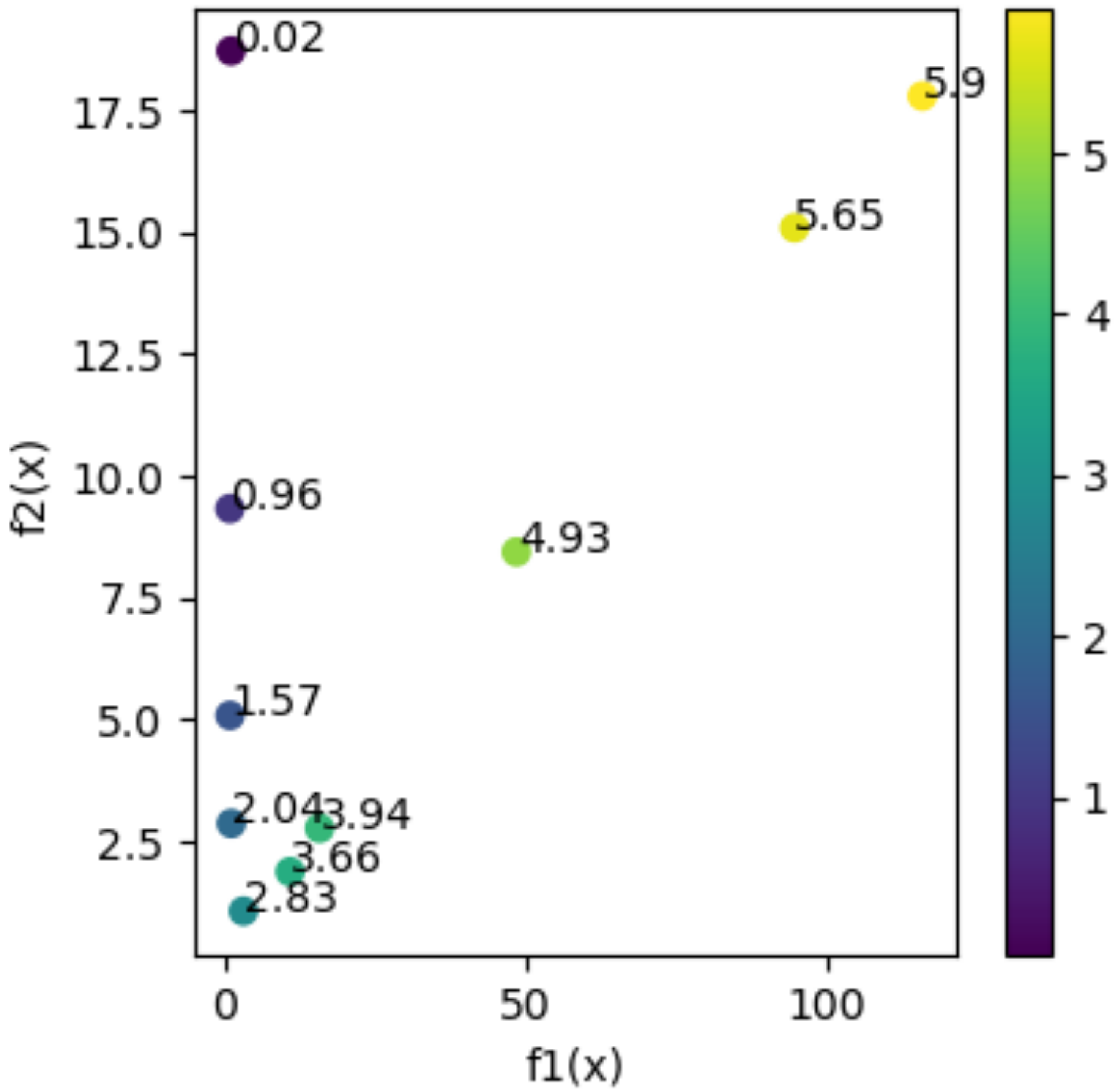
$$f_1(x) = \frac{1}{5}(x-1)^4 + 1$$

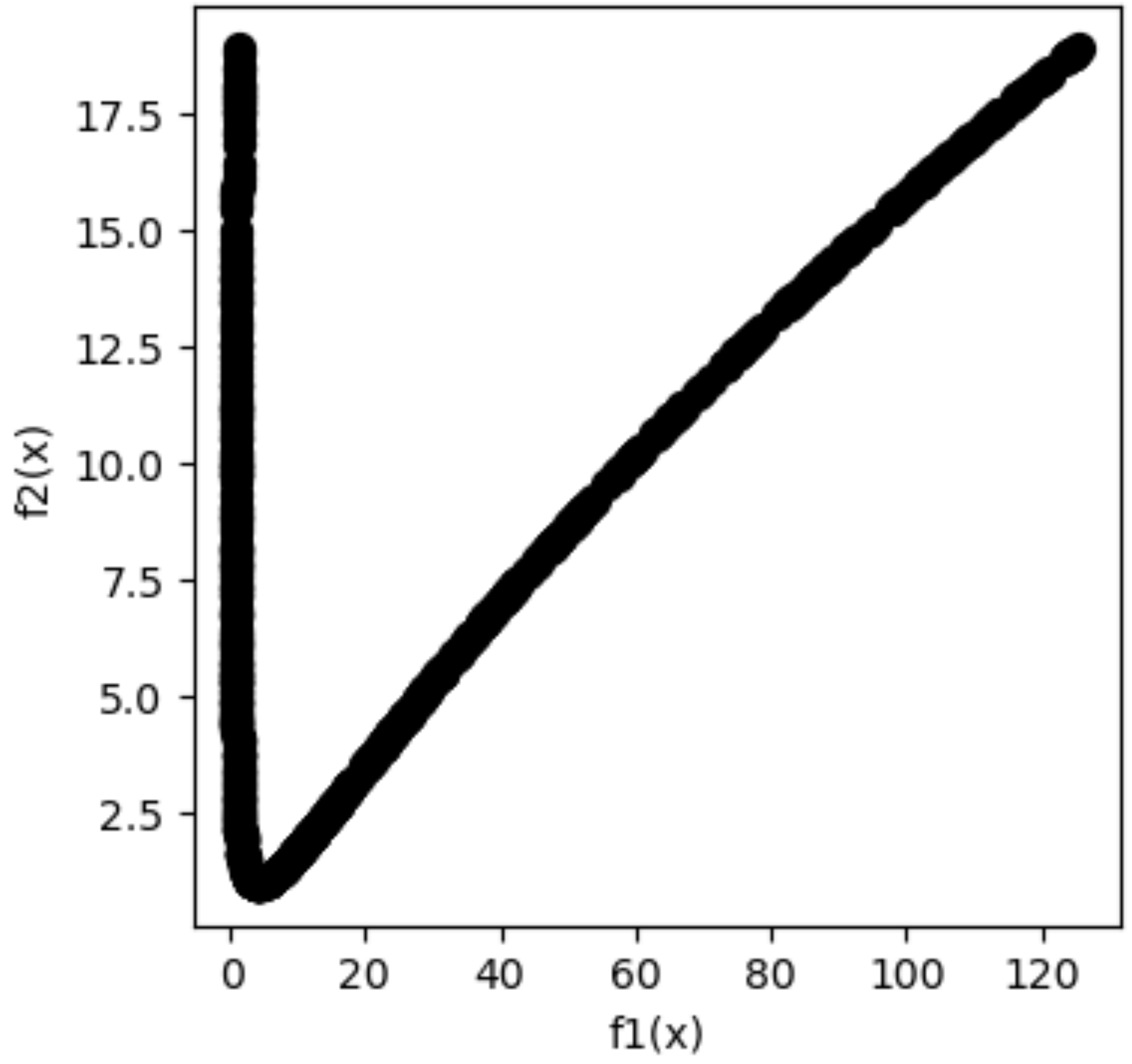
$$f_2(x) = 2(x-3)^2 + 1$$

$$x \in [0,6] \subset \mathbf{R}$$



Visualization in low dimensions: from solution to objective space





Pareto Dominance

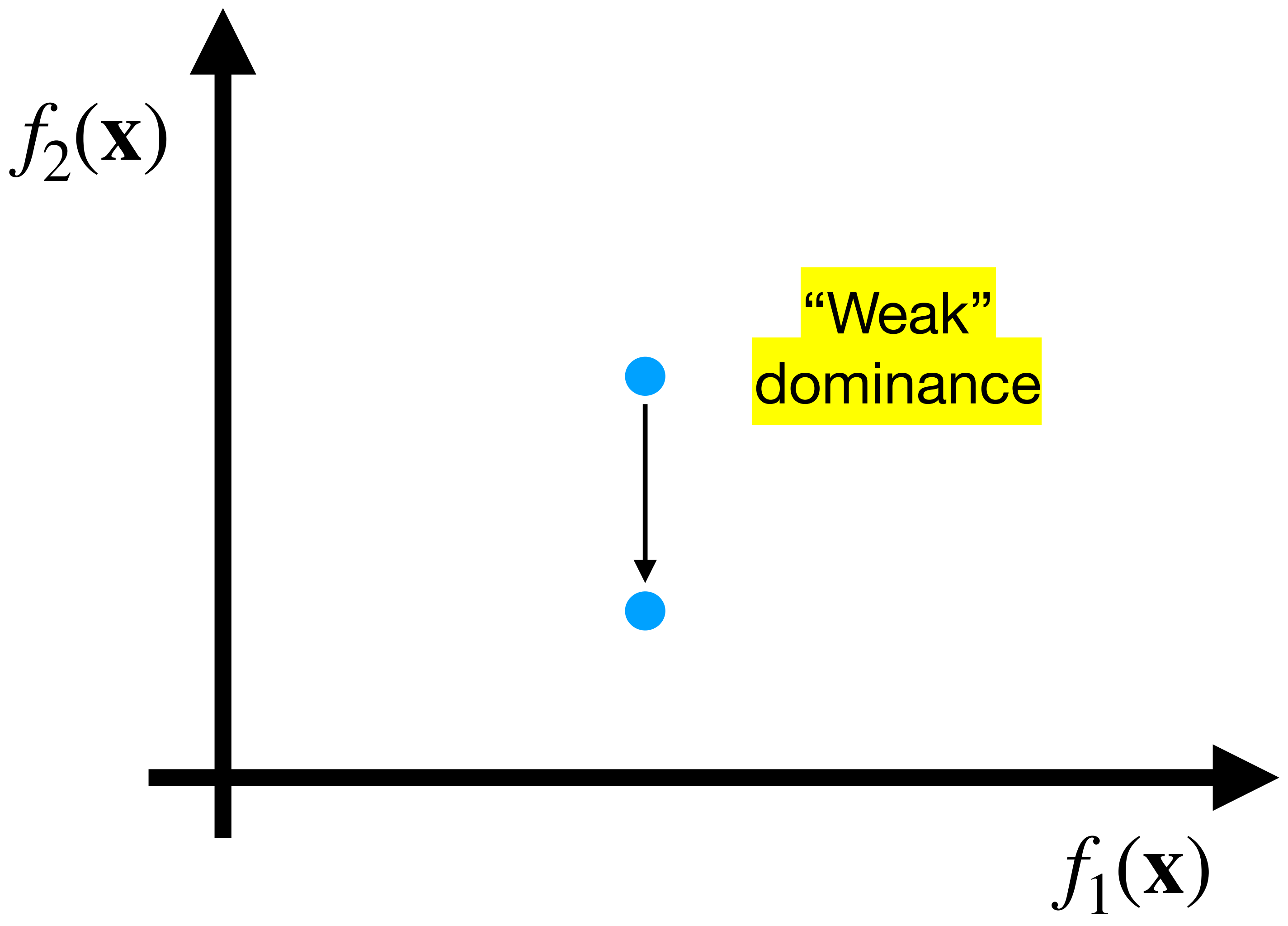
Definition

Pareto dominance: a solution $\mathbf{x}$ is said to dominate a solution $\mathbf{x}^*$ $(\mathbf{x} \preceq \mathbf{x}^*)$ if and only if

$$f_k(\mathbf{x}) \leq f_k(\mathbf{x}^*) \quad \forall k = 1, \ldots, M$$

and

$$\exists\, j \in 1, \ldots, M \quad s.t. \quad f_j(\mathbf{x}) < f_j(\mathbf{x}^*)$$



Pareto Dominance

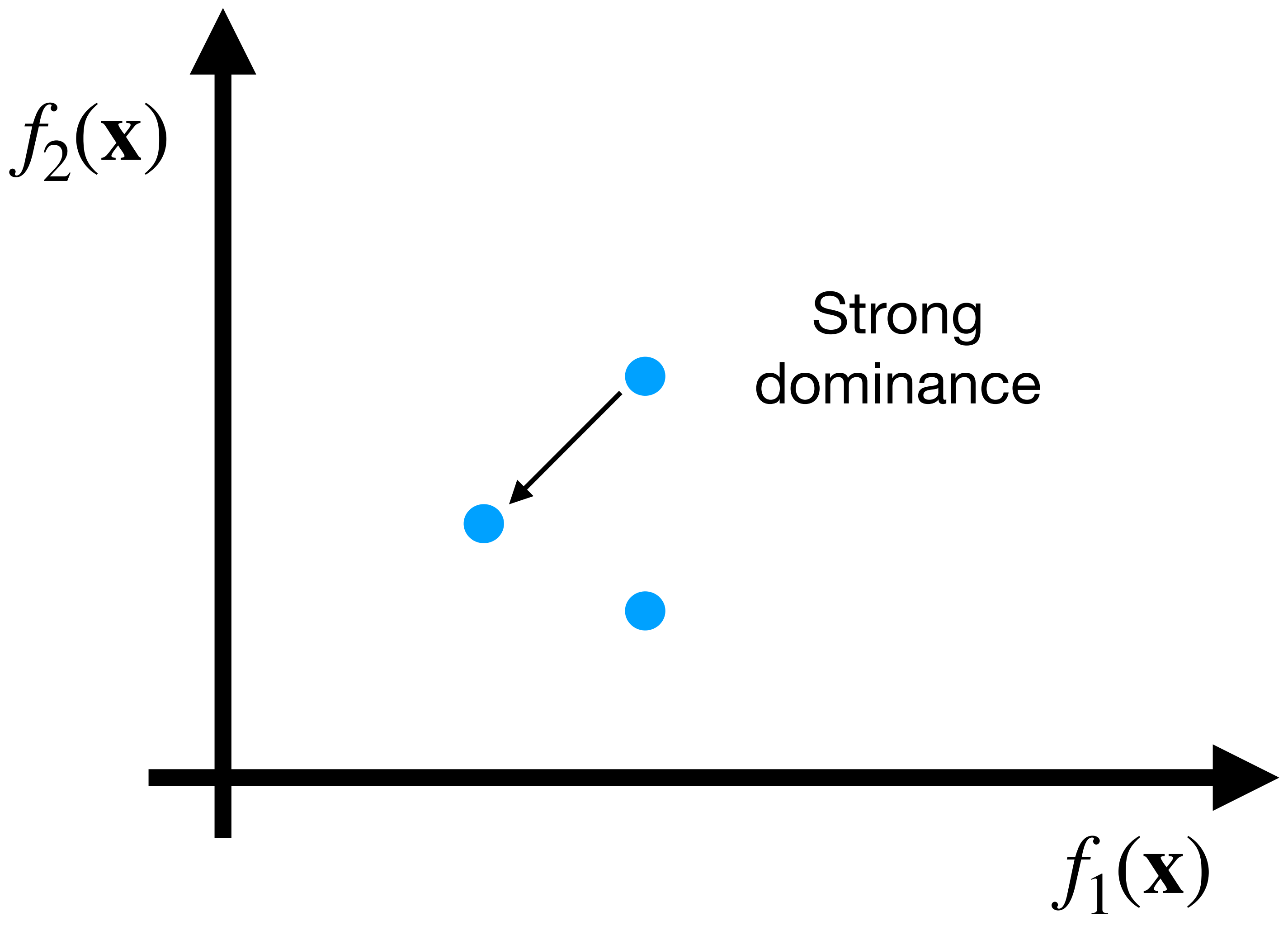
Definition

Pareto dominance: a solution $\mathbf{x}$ is said to dominate a solution $\mathbf{x}^*$ $(\mathbf{x} \preceq \mathbf{x}^*)$ if and only if

$$f_k(\mathbf{x}) \preceq f_k(\mathbf{x}^*) \quad \forall k = 1,\ldots,M$$

and

$$\exists\, j \in 1,\ldots,M \quad s.t. \quad f_j(\mathbf{x}) \prec f_j(\mathbf{x}^*)$$



Strong Pareto dominance: a solution $\mathbf{x}$ is said to strongly dominate a solution $\mathbf{x}^*$ $(\mathbf{x} \prec \mathbf{x}^*)$ if it is strictly better than $\mathbf{x}^*$ in all the objectives.

$$f_k(\mathbf{x}) \prec f_k(\mathbf{x}^*) \quad \forall k = 1,\ldots,M$$

Pareto Dominance

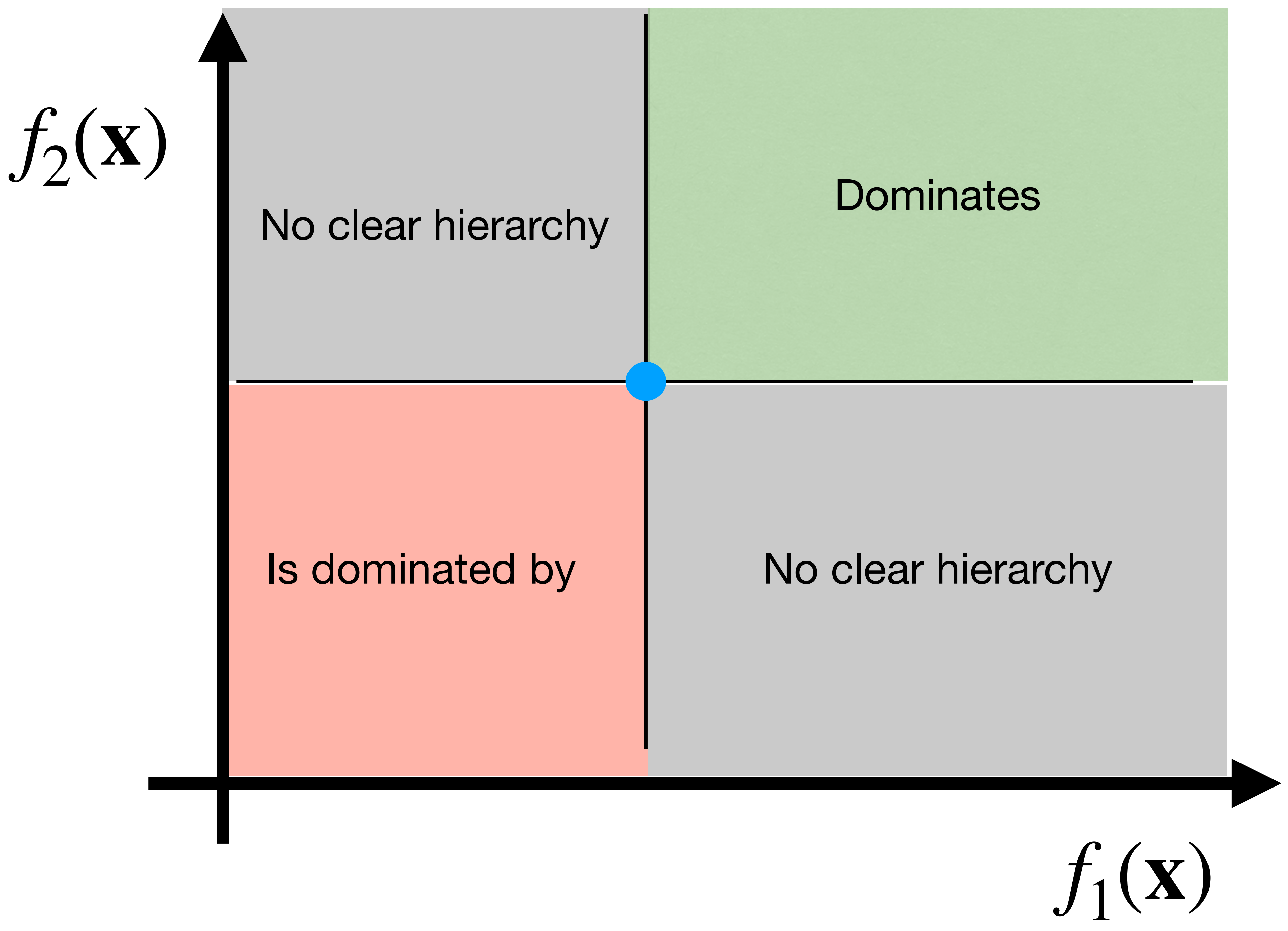
Definition

Pareto dominance: a solution $\mathbf{x}$ is said to dominate a solution $\mathbf{x}^*$ $(\mathbf{x} \preceq \mathbf{x}^*)$ if and only if

$$f_k(\mathbf{x}) \leq f_k(\mathbf{x}^*) \quad \forall k = 1,\ldots,M$$

and

$$\exists\, j \in 1,\ldots,M \quad s.t. \quad f_j(\mathbf{x}) \prec f_j(\mathbf{x}^*)$$



Strong Pareto dominance: a solution $\mathbf{x}$ is said to strongly dominate a solution $\mathbf{x}^*$ $(\mathbf{x} \prec \mathbf{x}^*)$ if it is strictly better than $\mathbf{x}^*$ in all the objectives.

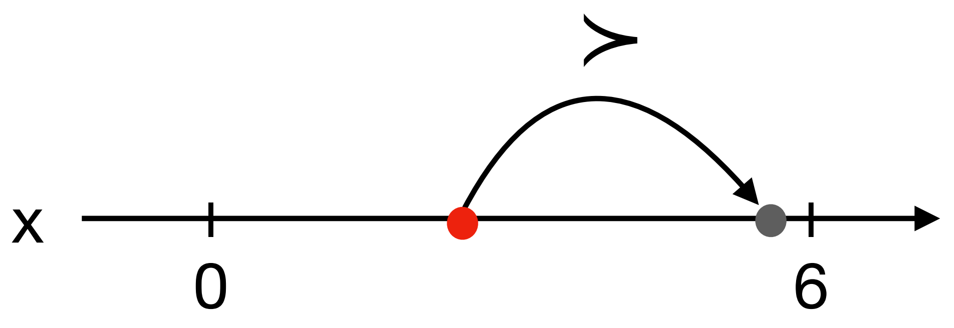
$$f_k(\mathbf{x}) \prec f_k(\mathbf{x}^*) \quad \forall k = 1,\ldots,M$$

Back to the example

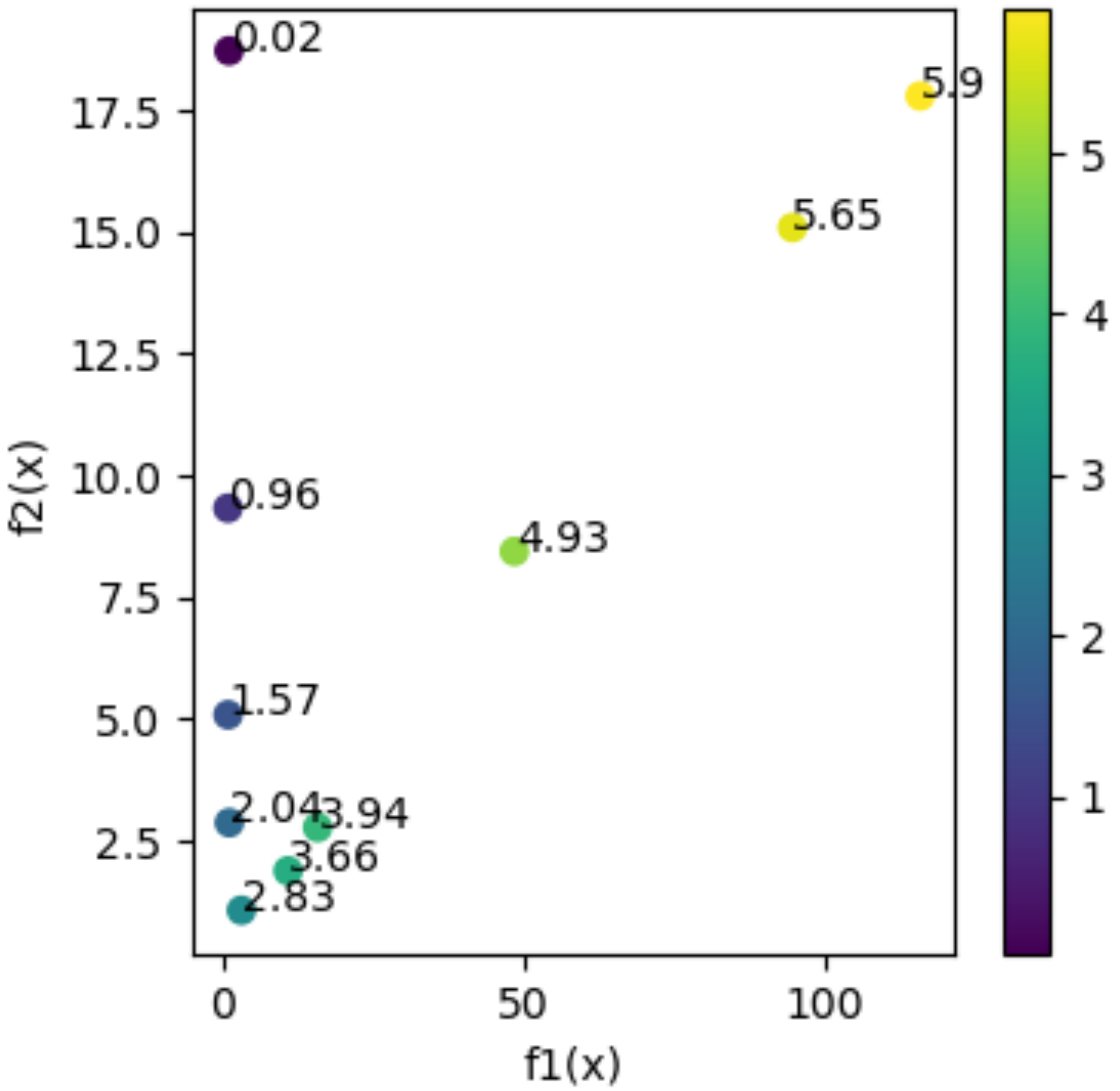
$$f_1(x) = \frac{1}{5}(x-1)^4 + 1$$

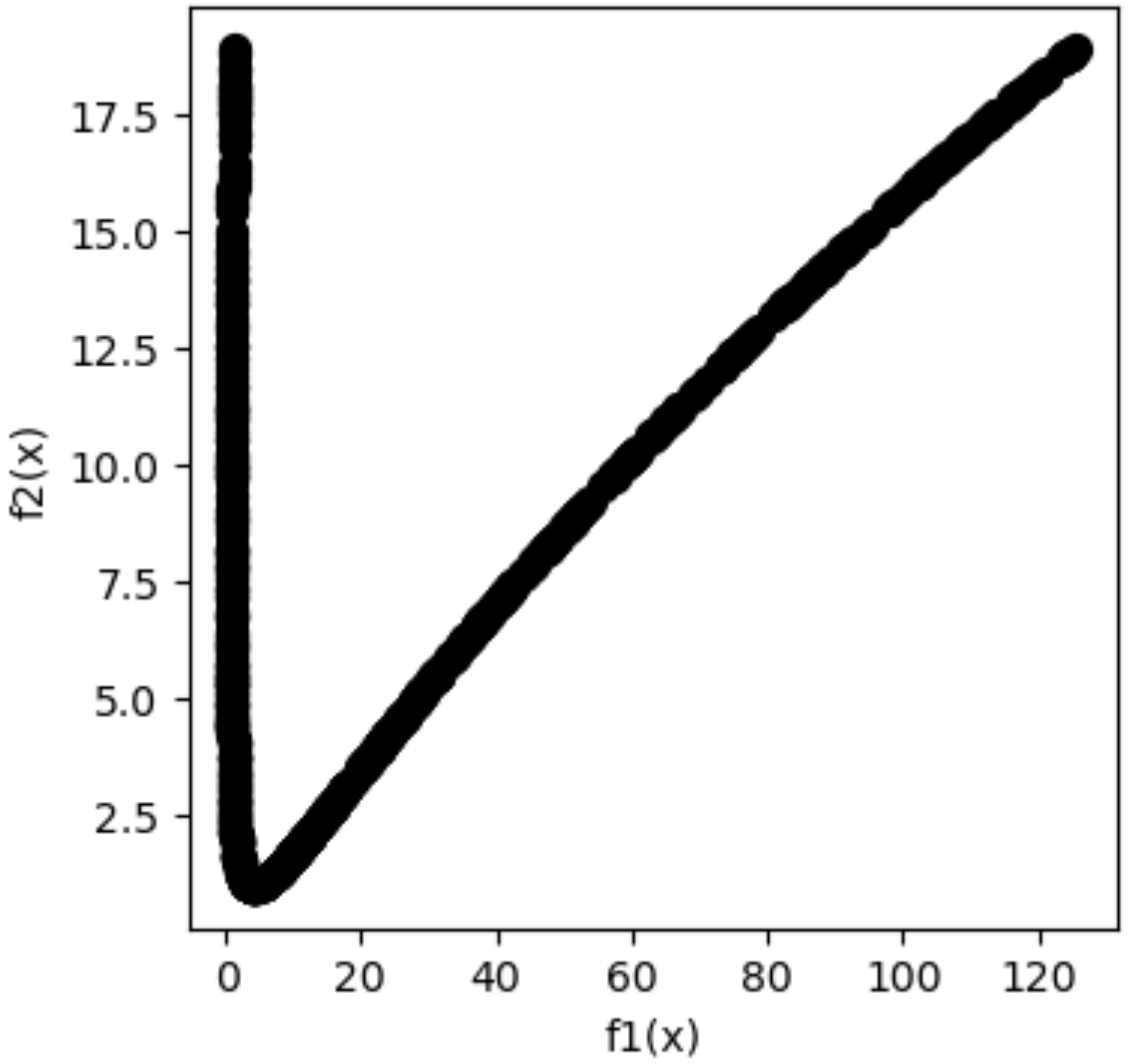
$$f_2(x) = 2(x-3)^2 + 1$$

$$x \in [0,6] \subset \mathbf{R}$$




```
x1 = 2.8 dominates x2 = 5.7:
[f1(x1),f2(x1)]=[(3.1, 1.08)]
[f1(x2),f2(x2)]=[(98.59, 15.58)]
```



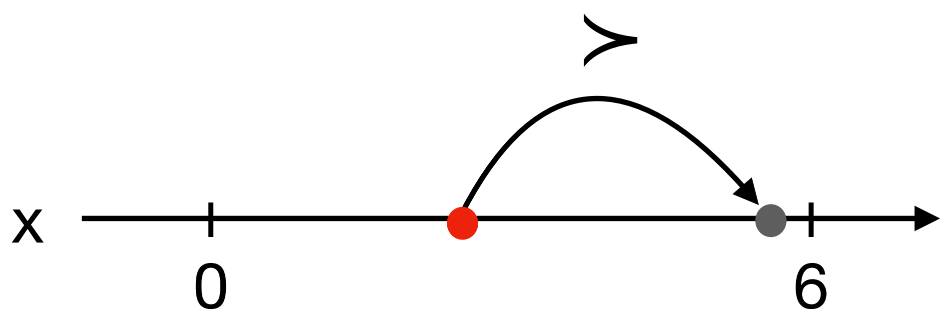


Back to the example

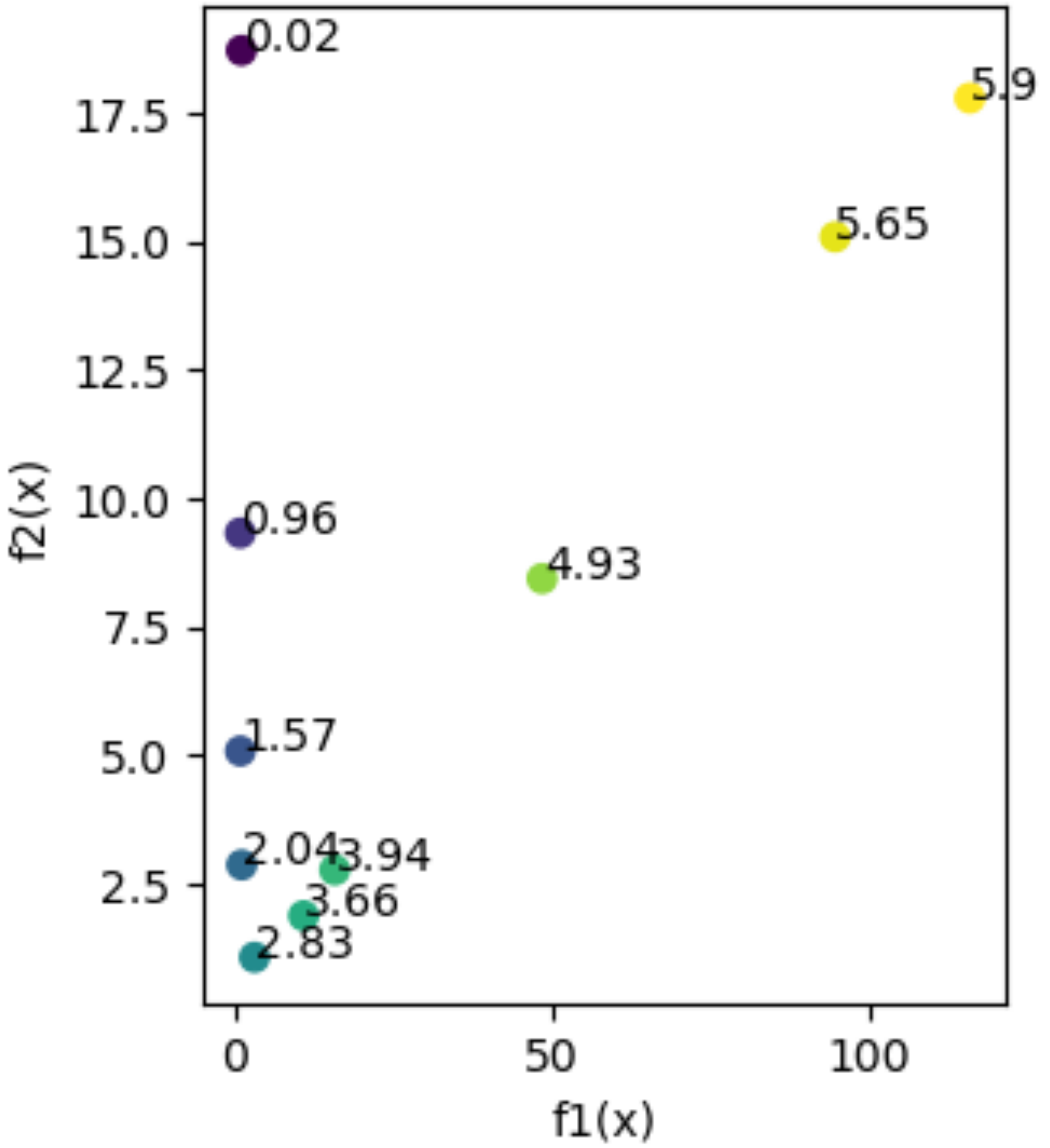
$$f_1(x) = \frac{1}{5}(x-1)^4 + 1$$

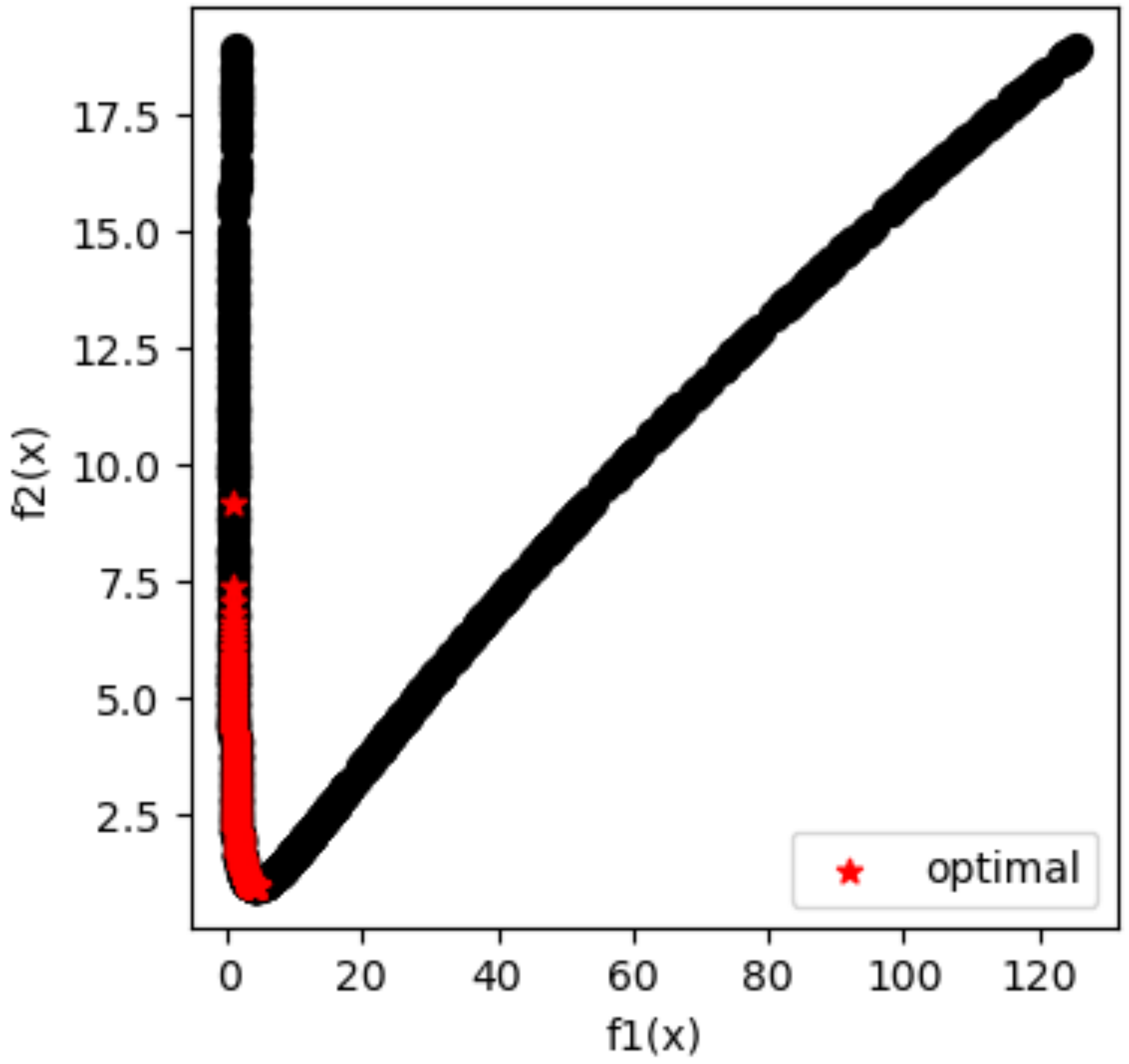
$$f_2(x) = 2(x-3)^2 + 1$$

$$x \in [0,6] \subset \mathbf{R}$$




```
x1 = 2.8 dominates x2 = 5.7:
[f1(x1),f2(x1)]=[(3.1, 1.08)]
[f1(x2),f2(x2)]=[(98.59, 15.58)]
```





Pareto Dominance

Properties

PD induces a **strong partial ordering**:

Partial ordering

- Reflexivity

  $$\mathbf{x} \preceq \mathbf{x} \quad \forall \mathbf{x} \in \mathscr{D}$$

- Antisymmetry

  If $\mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{x}$

  $$\implies \mathbf{x} = \mathbf{y} \qquad \forall \mathbf{x}, \mathbf{y} \in \mathscr{D}$$

- Transitivity

  if $\mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{z}$

  $$\implies \mathbf{x} \preceq \mathbf{z} \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathscr{D}$$

Pareto Dominance

Properties

PD induces a **strong partial ordering**:

- Not reflexive

  $$\mathbf{x} \not\preceq \mathbf{x} \quad \forall \mathbf{x} \in \mathscr{D}$$

- Asymmetry

  $$\nexists \mathbf{x}, \mathbf{y} \quad s.t. \quad \mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{x}$$

- Transitivity

  if $\mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{z}$

  $$\implies \mathbf{x} \preceq \mathbf{z} \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathscr{D}$$

Partial ordering

- Reflexivity

  $$\mathbf{x} \preceq \mathbf{x} \quad \forall \mathbf{x} \in \mathscr{D}$$

- Antisymmetry

  If $\mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{x}$

  $$\implies \mathbf{x} = \mathbf{y} \qquad \forall \mathbf{x}, \mathbf{y} \in \mathscr{D}$$

- Transitivity

  if $\mathbf{x} \preceq \mathbf{y} \cap \mathbf{y} \preceq \mathbf{z}$

  $$\implies \mathbf{x} \preceq \mathbf{z} \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathscr{D}$$

Pareto Optimality

Nomenclature

- **Non-dominated (Pareto optimal) solutions**

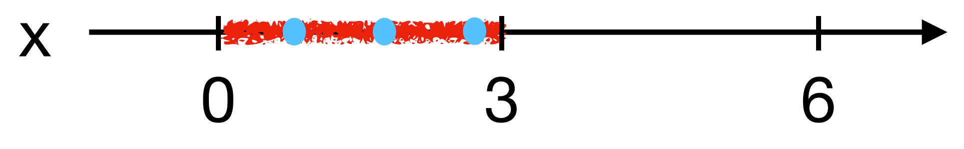
  $\mathbf{x} \in \mathcal{D} \quad s.t. \quad \vec{f}(\mathbf{x})$ is not dominated by any other solution. Blue dots.

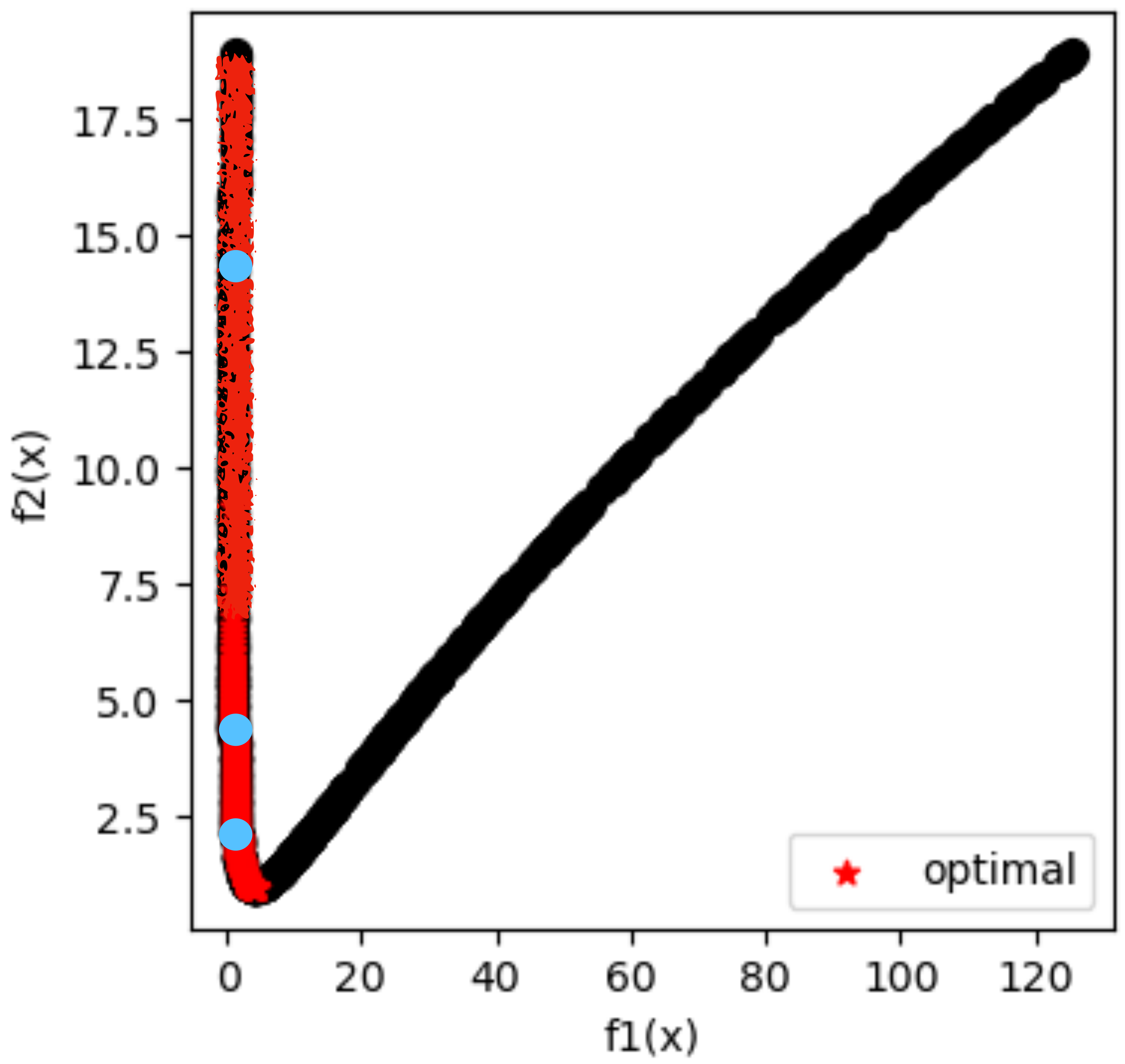
- **Pareto front** (in objective space)

  All $\mathbf{x} \in \mathcal{D} \quad s.t. \quad \vec{f}(\mathbf{x})$ belongs to the red curve in the objective space

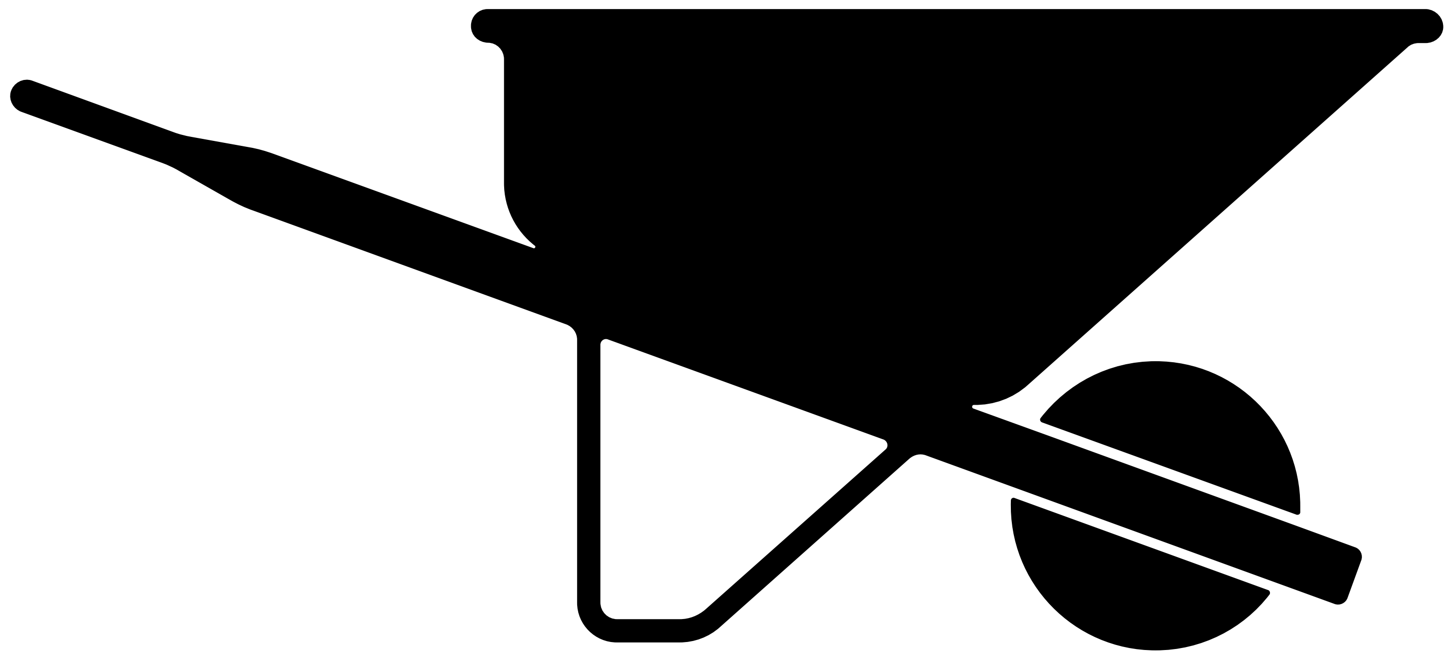
- **Pareto optimal set** (in variables space)

  All $\mathbf{x} \in \mathcal{D}$ belonging to the red curve in the decision variable space





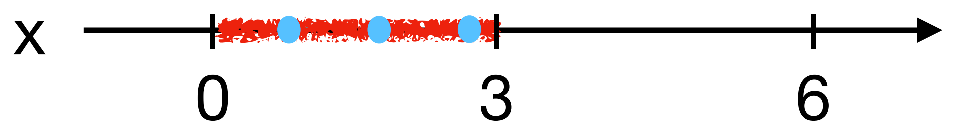
Hands on

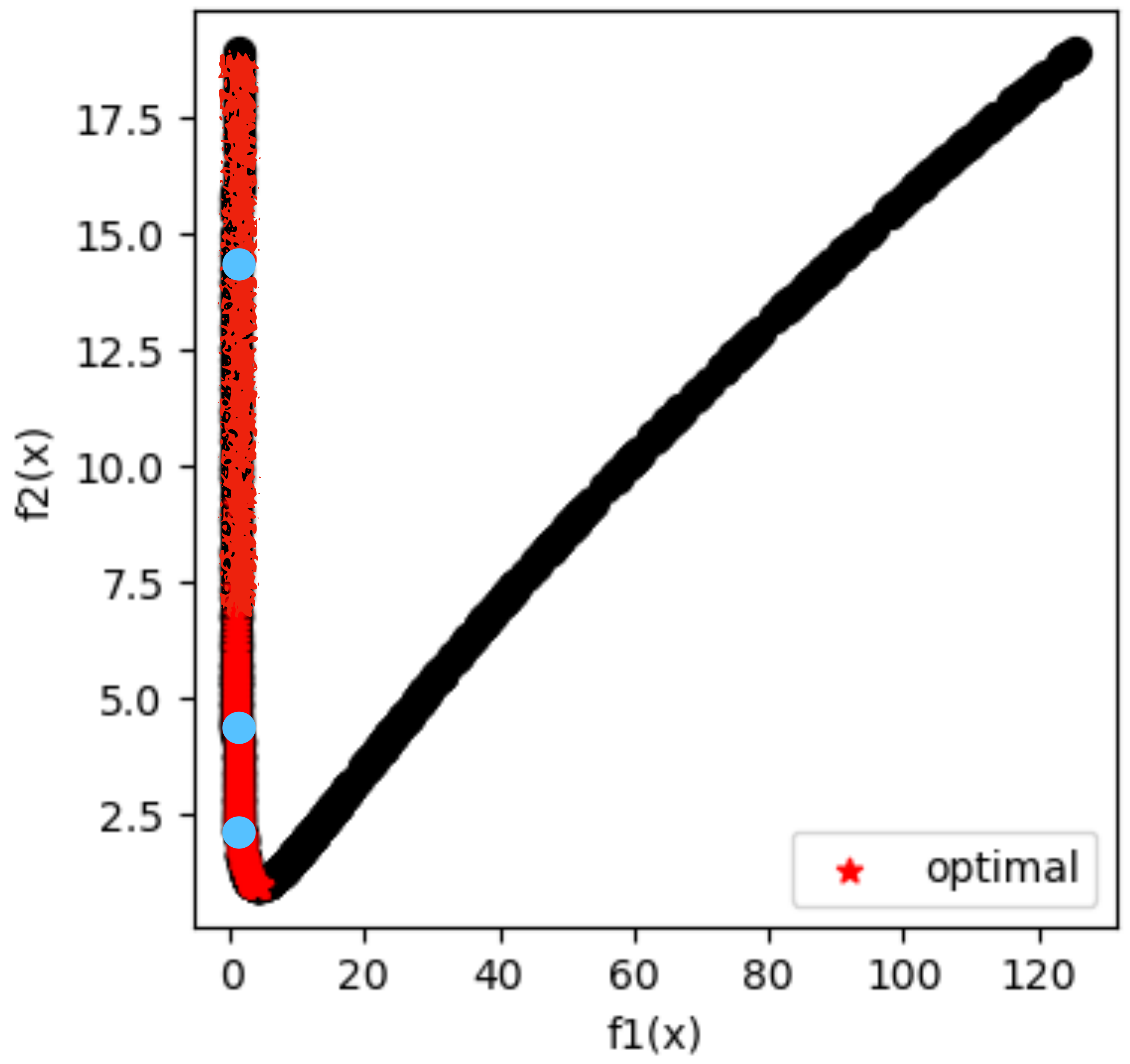


Approaches to MOO

Goals

- Find a non-dominated solution
- Find all non-dominated solution (Pareto set)

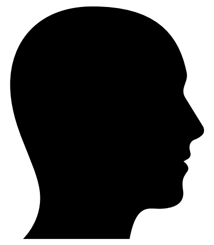




Approaches to MOO

Goals

- Find a non-dominated solution
- Find all non-dominated solution (Pareto set)

Decision making  ***<- decision maker***

- A priori: Define  —> MOO, No Pareto set —> Single solution

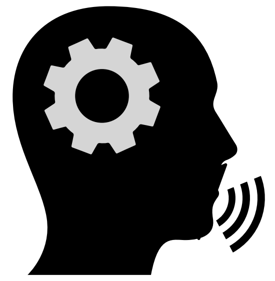
Approaches to MOO

Goals

- Find a non-dominated solution
- Find all non-dominated solution (Pareto set)

Decision making

 ***<- decision maker***

- A priori: Define  —> MOO, No Pareto set —> Single solution
- A posteriori: Define —> MOO, Pareto set —> Single solution 

Approaches to MOO

Goals

- Find a non-dominated solution
- Find all non-dominated solution (Pareto set)

Decision making

<- decision maker

- A priori: Define —> MOO, No Pareto set —> Single solution
- A posteriori: Define —> MOO, Pareto set —> Single solution
- Interactive: Define —> MOO, interactive (HITL) —> Single solution

Approaches to MOO

Goals

- Find a non-dominated solution
- Find all non-dominated solution (Pareto set)

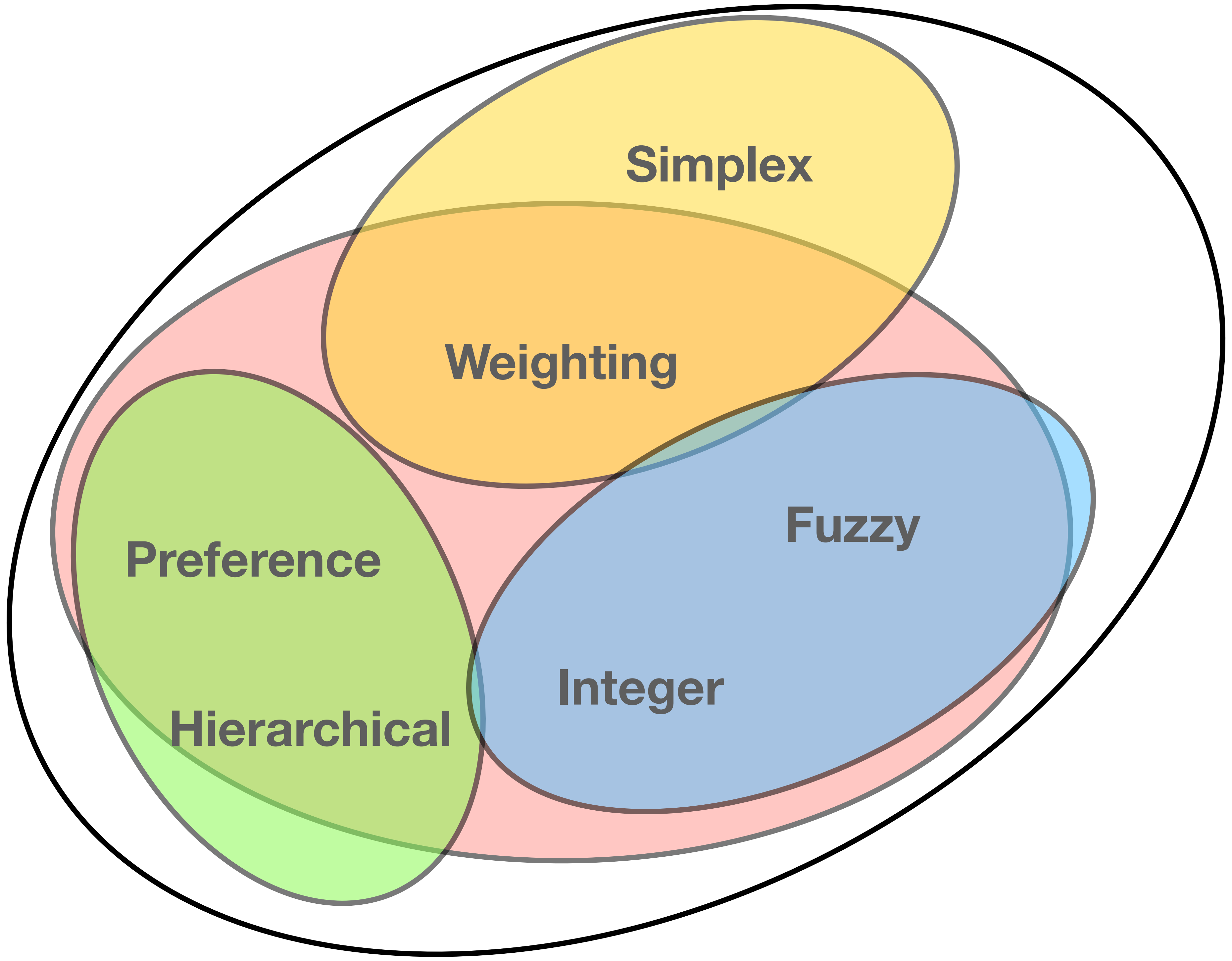
Decision making

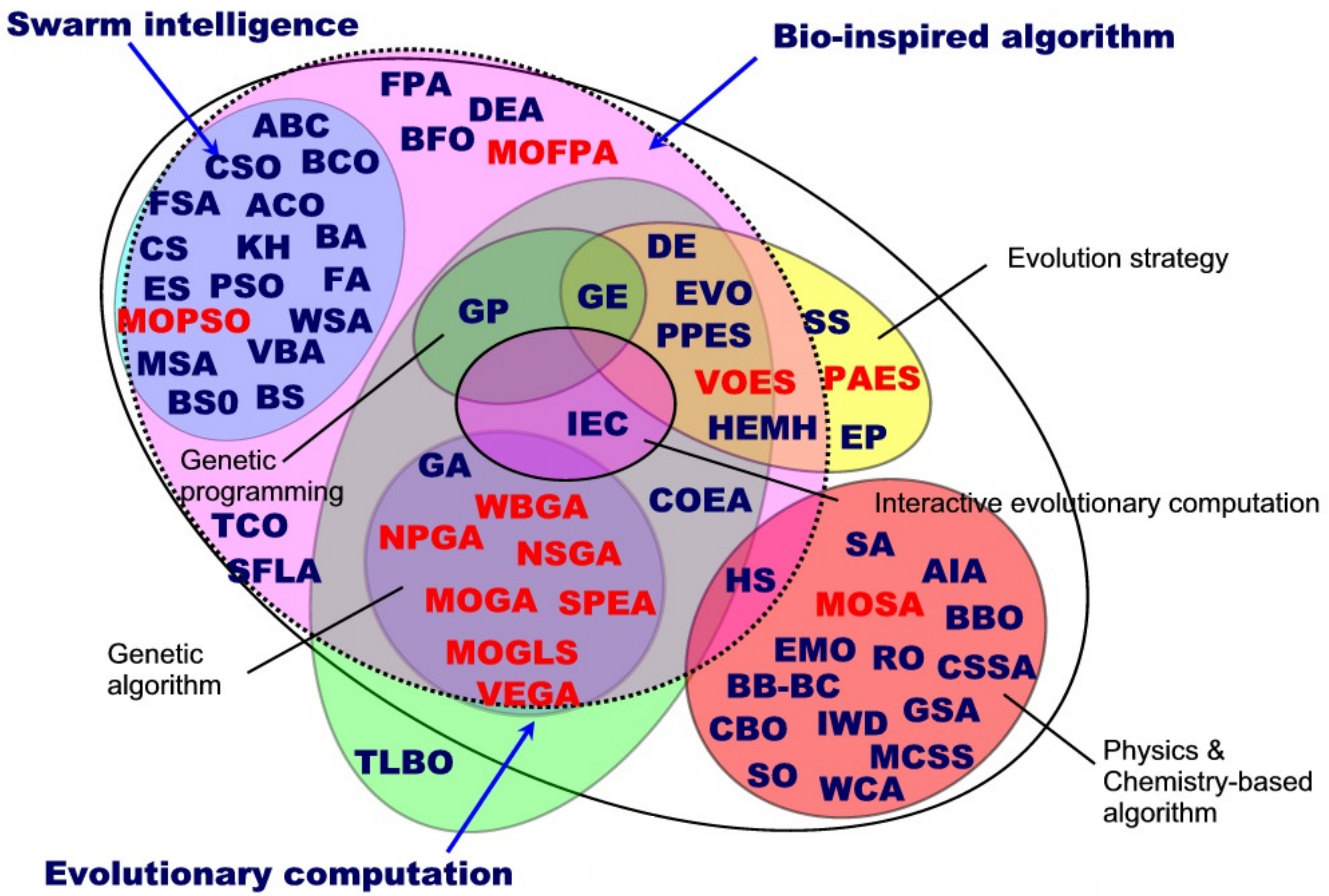
<- decision maker

- A priori: Define —> MOO, No Pareto set —> Single solution
- A posteriori: Define —> MOO, Pareto set —> Single solution
- Interactive: Define —> MOO, interactive (HITL) —> Single solution
- No DM / preference: Define —> Run —> Pareto set

Taxonomy

Classical and meta-heuristic methods





Taken from (Keller, 2017)

Scalarization Algorithms

Basic idea

- Combine multiple objectives into a single utility function
- Solve the single objective problem
- Find one Pareto optimal solution
- Repeat to find the whole Pareto optimal set

Weighted sum methods

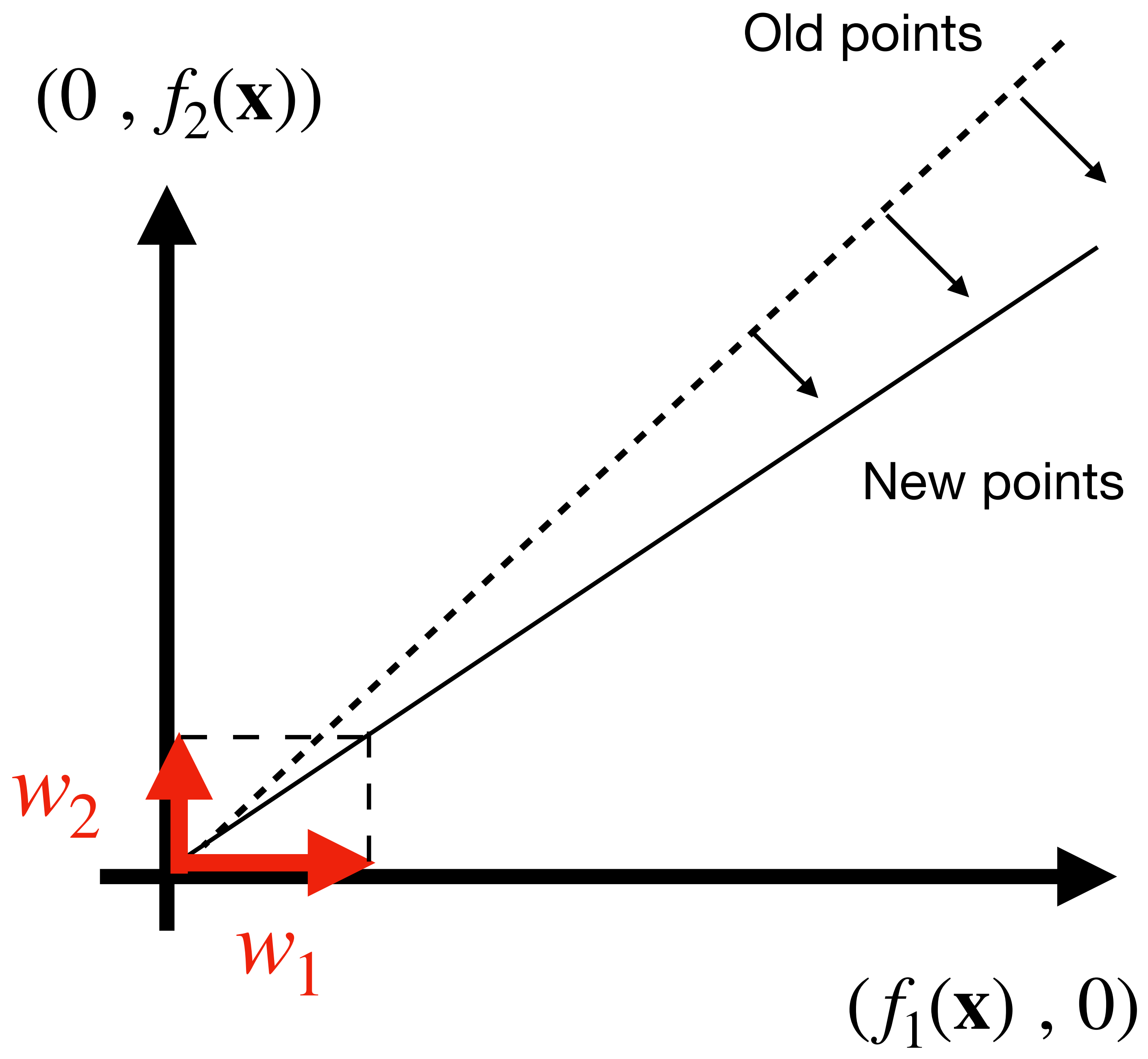
Definition

- Identify a weight vector in the objective space

$$\min_{\mathbf{x}} \sum_{i=1}^{M} w_i f_i(\mathbf{x})$$

$$\mathbf{x} \in \mathscr{D} \qquad w_i > 0 \quad \forall i \in 1,\ldots,M$$

$$\sum_{i=1}^{M} w_i = 1$$



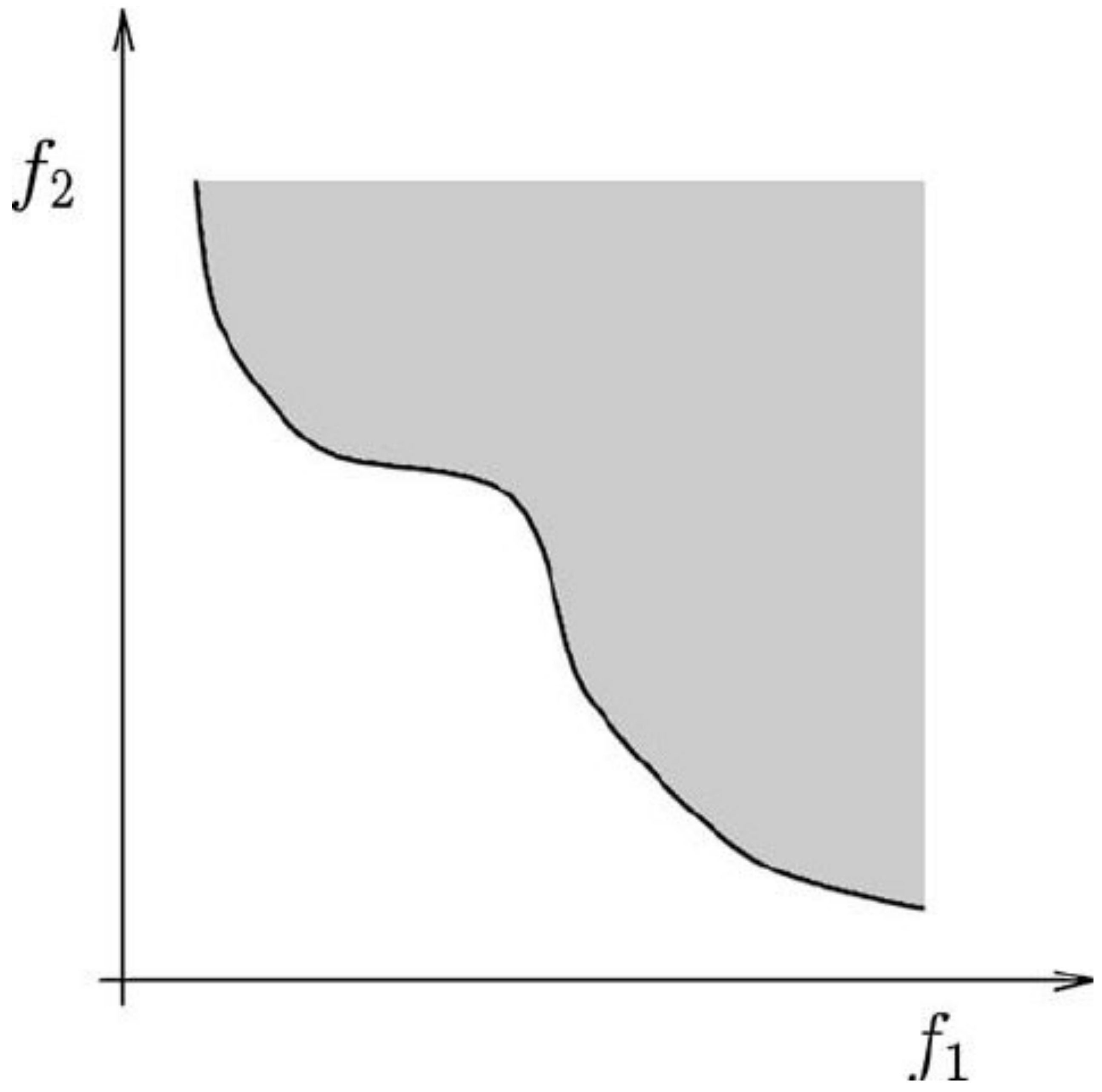
- Normalized weights ensure Pareto optimality

Weighted sum methods
Definition

Can we get all optimal points?

It depends on:

- Convexity of the domain
- Convexity of objectives



Censor, Y. (1977). Pareto optimality in multiobjective problems. Applied Mathematics and Optimization, 4(1), 41-59.

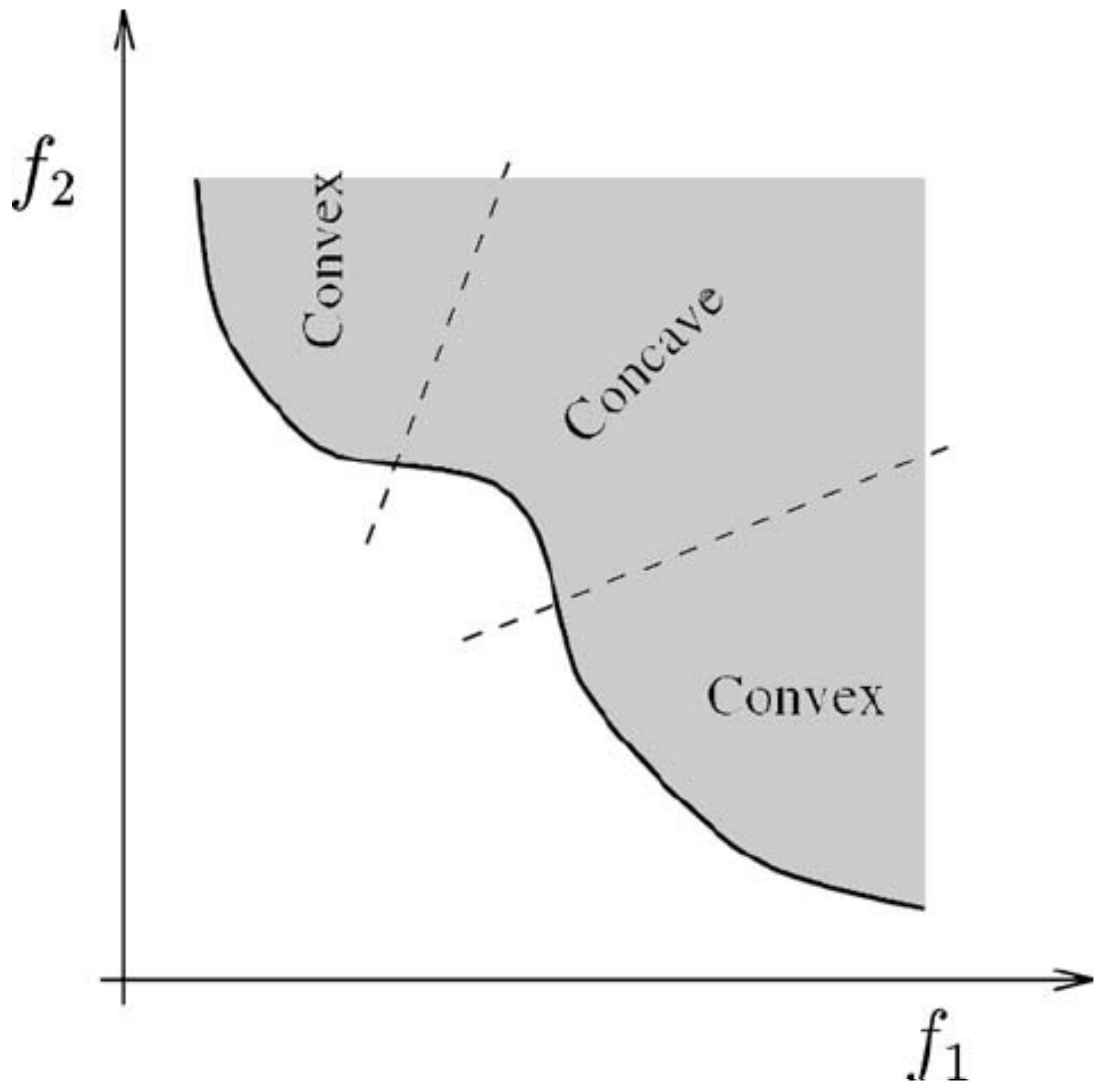
Weighted sum methods

Definition

Can we get all optimal points?

It depends on:

- Convexity of the domain
- Convexity of objectives



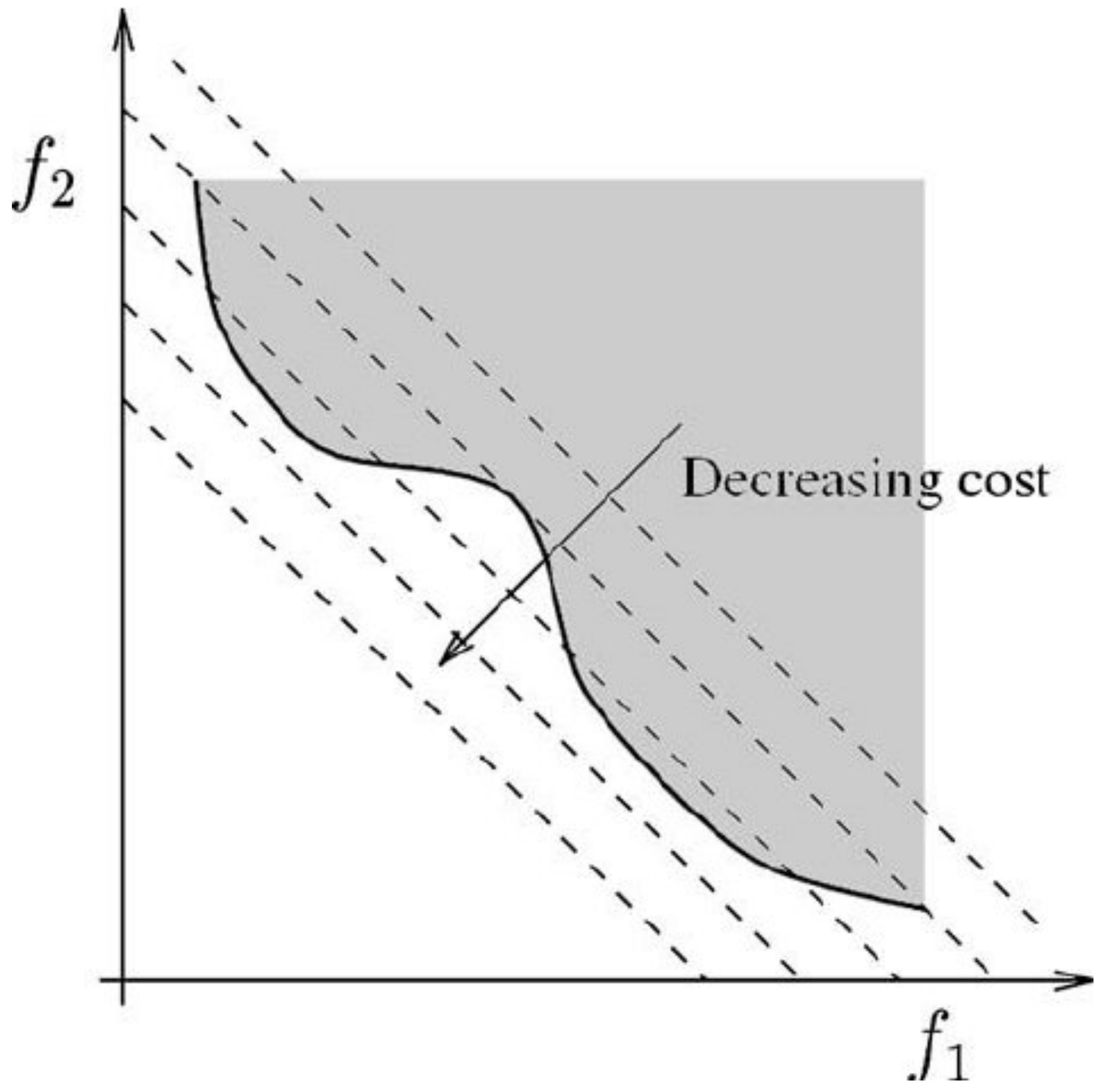
Censor, Y. (1977). Pareto optimality in multiobjective problems. Applied Mathematics and Optimization, 4(1), 41-59.

Weighted sum methods
Definition

Can we get all optimal points?

It depends on:

- Convexity of the domain
- Convexity of objectives



Censor, Y. (1977). Pareto optimality in multiobjective problems. Applied Mathematics and Optimization, 4(1), 41-59.

Example

Convex to non-convex

$$f_1(x_1, x_2) = x_1$$

$$f_2(x_1, x_2) = 1 - x_1 - \alpha \sin(\beta\pi x_1) + x_2^2$$

Solve

$$\min_{\mathbf{x}} \left(w_1 f_1(x_1, x_2) + w_2 f_2(x_1, x_2) \right)$$

$$w_1 + w_2 = 1 \ , \qquad w_1, w_2 > 0$$

Example

Convex to non-convex

$$f_1(x_1, x_2) = x_1$$

$$f_2(x_1, x_2) = 1 - x_1 - \alpha \sin(\beta \pi x_1) + x_2^2$$

Solve

$$\min_{\mathbf{x}} \left(w_1 f_1(x_1, x_2) + w_2 f_2(x_1, x_2) \right)$$

$$w_1 + w_2 = 1 \ , \qquad w_1, w_2 > 0$$

$$\boxed{\begin{aligned} & x_1^{opt} = \frac{1}{\beta\pi} \cos^{-1} \left[\frac{1}{\alpha\beta\pi} \left(\frac{w_1}{w_2} - 1 \right) \right] \\ & x_2^{opt} = 0 \\ & 1 - \alpha\beta\pi \leq \frac{w1}{w_2} \leq 1 + \alpha\beta\pi \end{aligned}}$$

Example

Convex to non-convex

$$f_1(x_1, x_2) = x_1$$

$$f_2(x_1, x_2) = 1 - x_1 - \alpha \sin(\beta \pi x_1) + x_2^2$$

Solve

$$\min_{\mathbf{x}} \left(w_1 f_1(x_1, x_2) + w_2 f_2(x_1, x_2) \right)$$

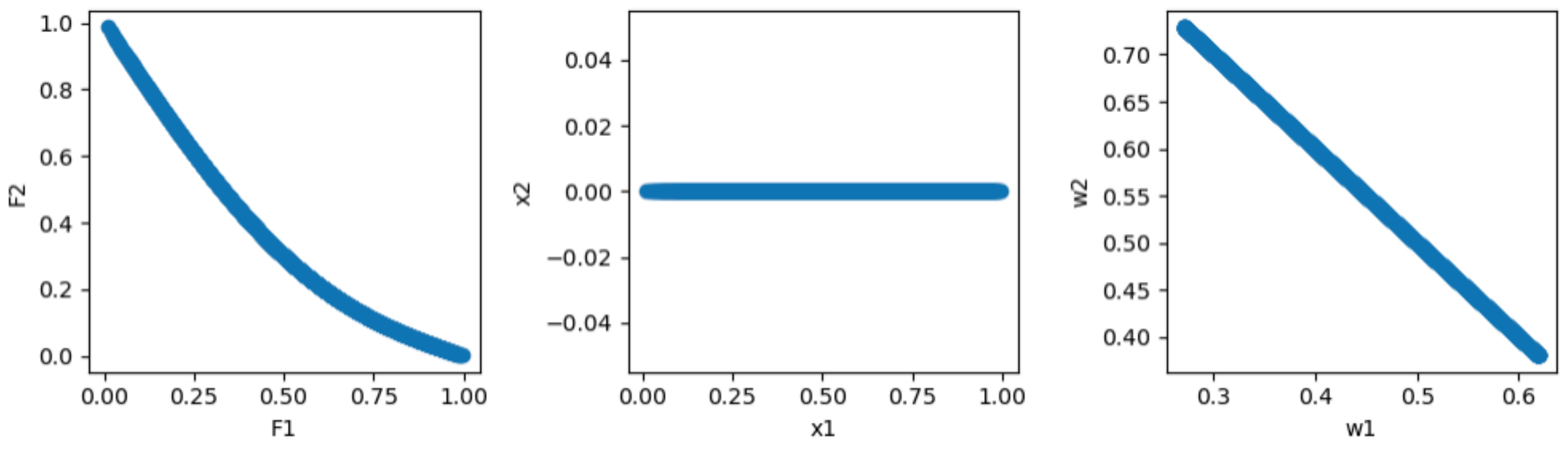
$$w_1 + w_2 = 1 \ , \qquad w_1, w_2 > 0$$

$$x_1^{opt} = \frac{1}{\beta\pi} \cos^{-1} \left[\frac{1}{\alpha\beta\pi} \left(\frac{w_1}{w_2} - 1 \right) \right]$$

$$x_2^{opt} = 0$$

$$1 - \alpha\beta\pi \leq \frac{w1}{w_2} \leq 1 + \alpha\beta\pi$$

$\alpha = 1/5 \qquad \beta = 1$



Example

Convex to non-convex

$$f_1(x_1, x_2) = x_1$$

$$f_2(x_1, x_2) = 1 - x_1 - \alpha \sin(\beta \pi x_1) + x_2^2$$

Solve

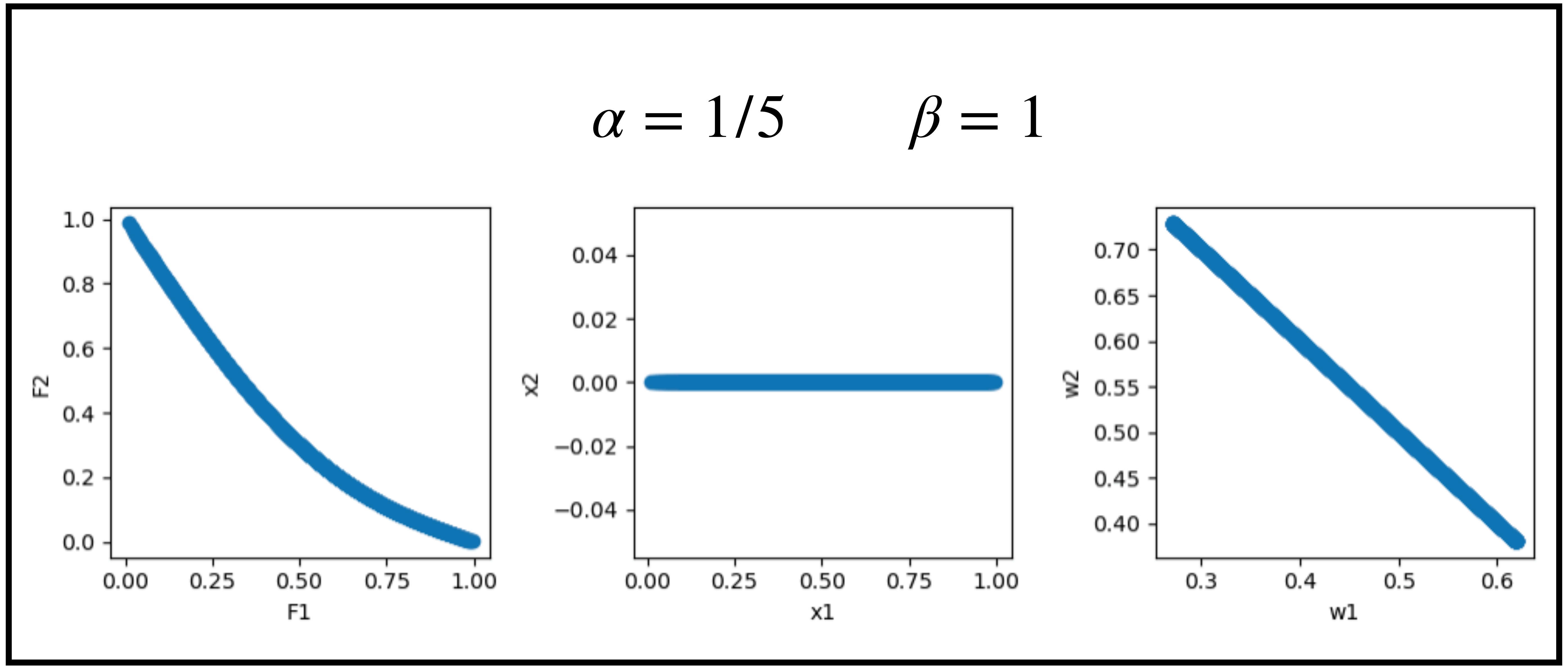
$$\min_{\mathbf{x}} \left(w_1 f_1(x_1, x_2) + w_2 f_2(x_1, x_2) \right)$$

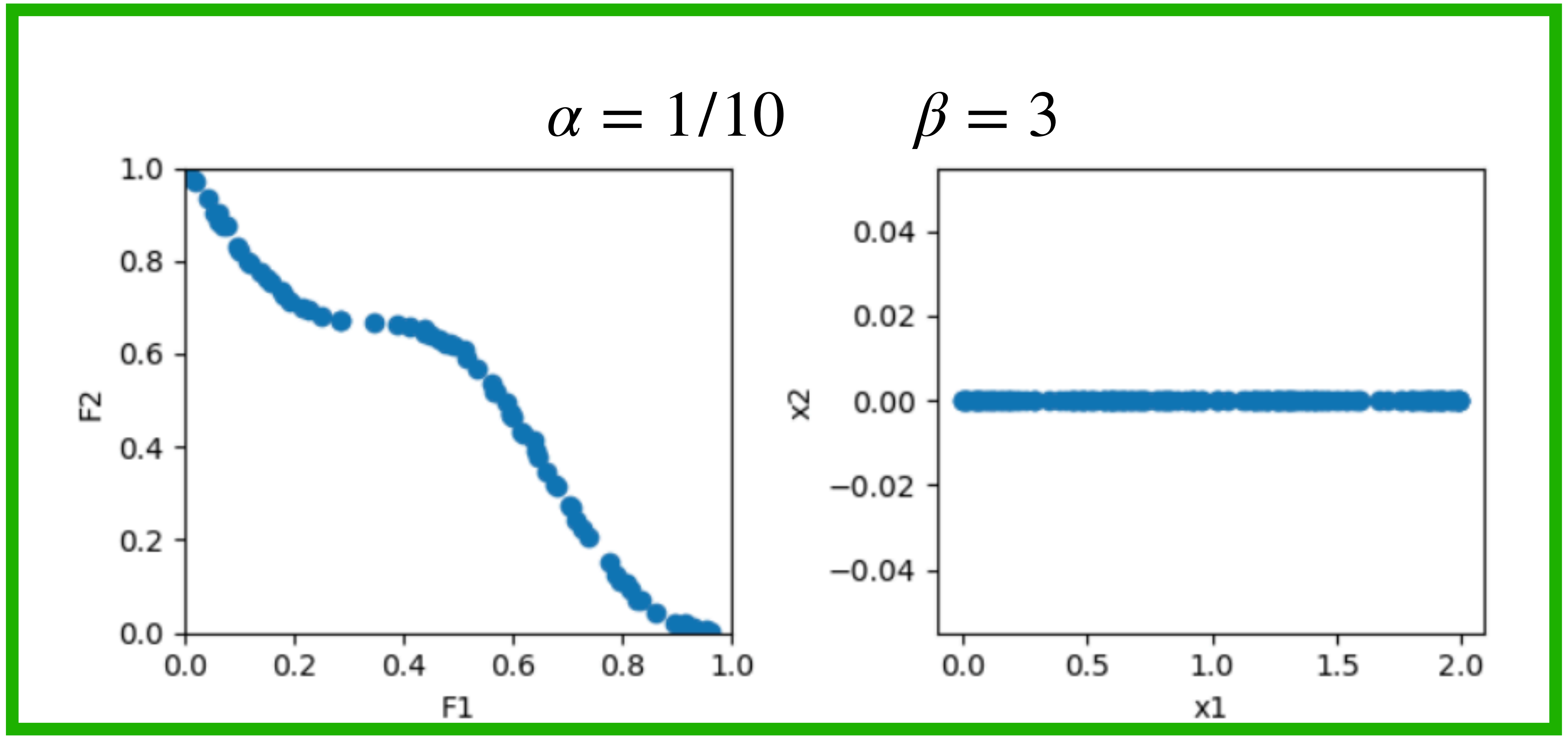
$$w_1 + w_2 = 1 \,, \qquad w_1, w_2 > 0$$

$$x_1^{opt} = \frac{1}{\beta\pi} \cos^{-1}\left[\frac{1}{\alpha\beta\pi} \left(\frac{w_1}{w_2} - 1 \right) \right]$$

$$x_2^{opt} = 0$$

$$1 - \alpha\beta\pi \leq \frac{w1}{w_2} \leq 1 + \alpha\beta\pi$$





Weighted sum methods

We can now answer

Can we get all optimal points?

… not guaranteed …

A sufficient condition is:

- The domain $\mathscr{D}, S$ is convex
- Each objective $f_k(\mathbf{x})$ is convex

Censor, Y. (1977). Pareto optimality in multiobjective problems. Applied Mathematics and Optimization, 4(1), 41-59.

Weighted sum methods

We can now answer

Can we get all optimal points?

… not guaranteed …

Other methods can solve this limitation

A sufficient condition is:

- The domain $\mathscr{D}, S$ is convex
- Each objective $f_k(\mathbf{x})$ is convex

Censor, Y. (1977). Pareto optimality in multiobjective problems. Applied Mathematics and Optimization, 4(1), 41-59.

Weighted exponential *(P. L. Yu, 1973)*

The scalar objective to optimize becomes:

$$f_s = \sum_{i=1}^{M} w_i [f_i(\mathbf{x})]^p$$

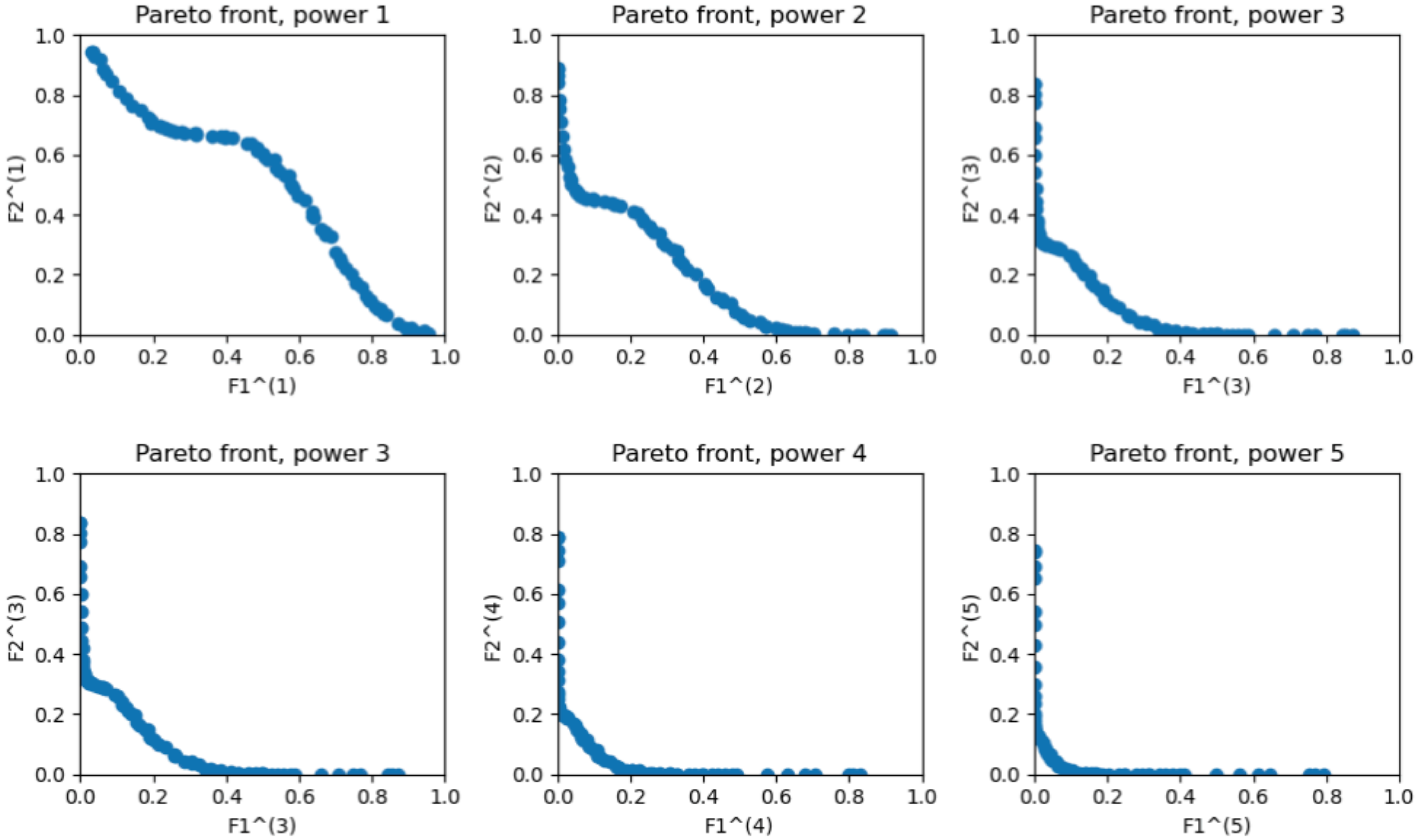
where $f_i(\mathbf{x}) > 0 \quad \forall i = 1,\ldots,M \qquad 1 \le p < \infty$

$$\mathbf{x} \in \mathscr{D} \qquad w_i > 0 \quad \forall i \in 1,\ldots,M \qquad \sum_{i=1}^{M} w_i = 1$$

- The condition on weights ensures optimality
- Bigger p, bigger effectiveness
- Bigger p, may give non-Pareto solutions

Weighted exponential *(P. L. Yu, 1973)*

Pictorial view of why it works:



Weighted metric *(P. L. Yu and G. Leitmann, 1974)*

Define the ideal point for each objective $f^* = (f_1{}^*, \ldots, f_M{}^*)$

1) Utopian point - data-driven - min value of each objective

2) Goal point - decision maker 

$$f_s = \left[\sum_{i=1}^{M} w_i^p \, | f_i(\mathbf{x}) - f_i{}^* |^p \right]^{1/p}$$

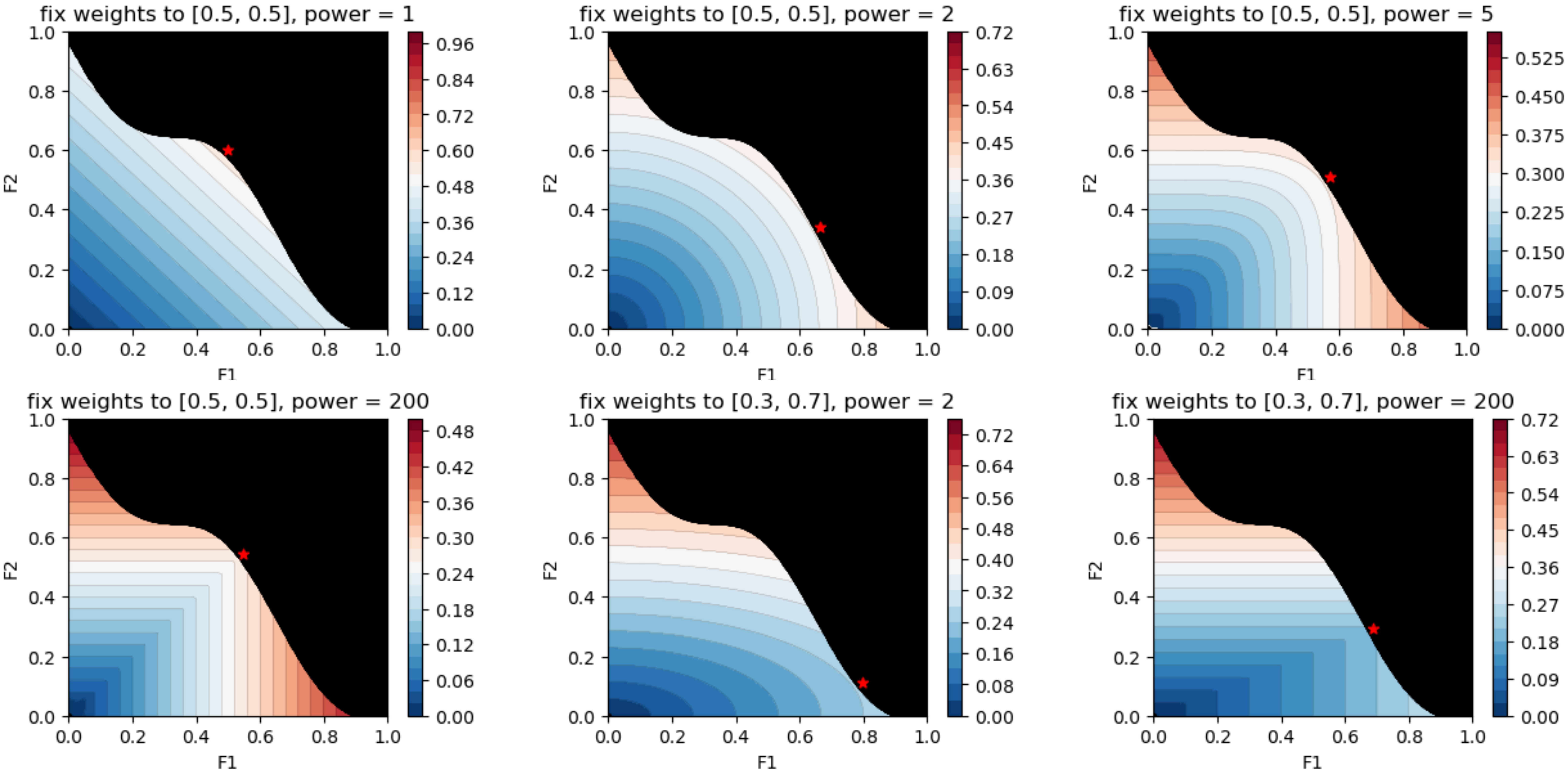
where

$$\mathbf{x} \in \mathscr{D} \qquad w_i > 0 \quad \forall i \in 1,\ldots,M$$

$$\sum_{i=1}^{M} w_i = 1, \quad 1 \leq p < \infty$$

- The condition on weights ensures optimality
- Bigger p, bigger effectiveness but may be non Pareto
- Can play with ideal point

Weighted metric *(P. L. Yu and G. Leitmann, 1974)*



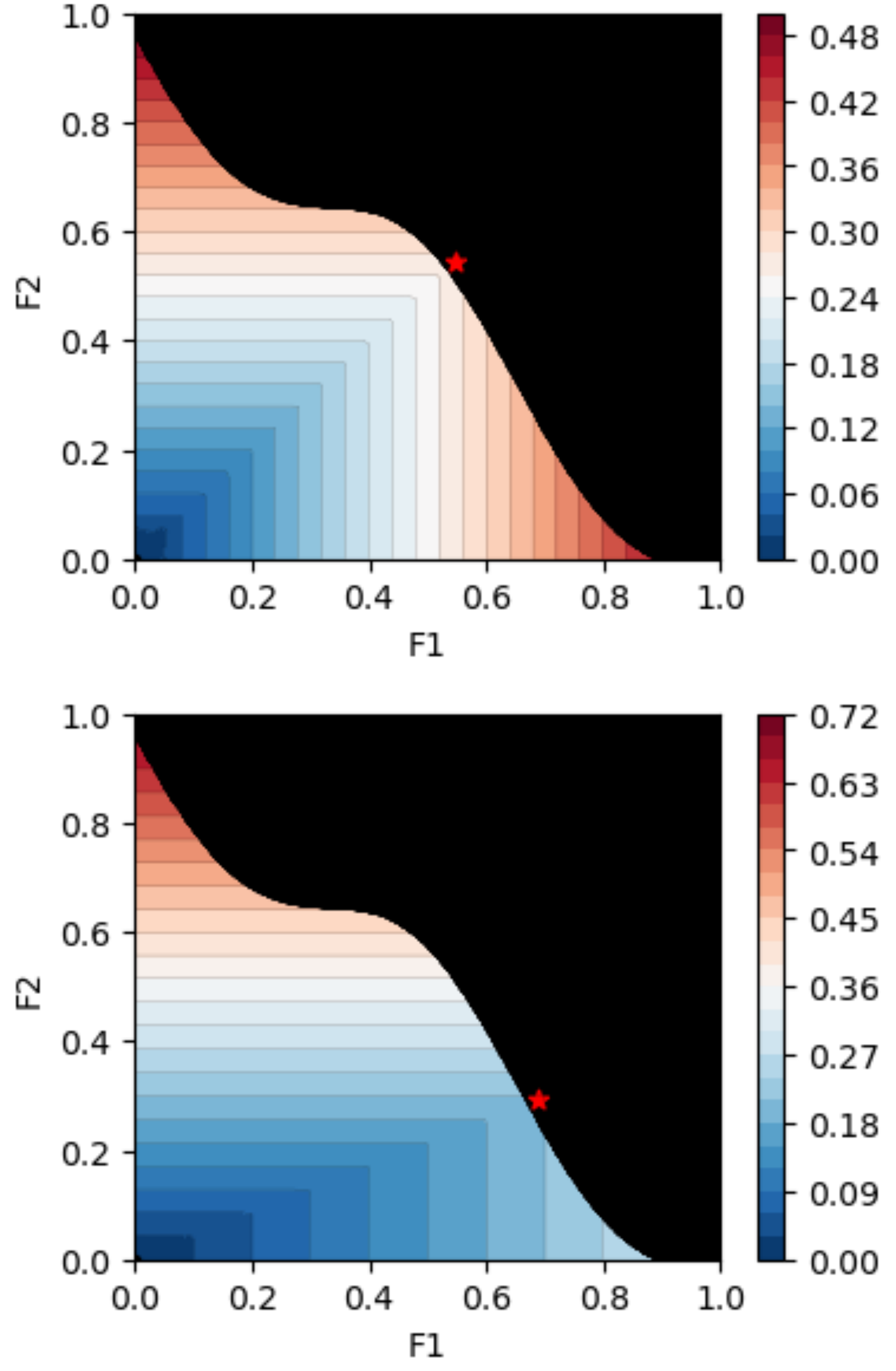
Weighted Chebyshev method *(Lightner, 1981; Messac, 2000)*

$$f_s = \max_{i \in \{1,\ldots,M\}} w_i \,|\, f_i(\mathbf{x}) - f_i^* \,|$$

$$\mathbf{x} \in \mathcal{D} \qquad w_i > 0 \quad \forall i \in 1,\ldots,M$$

$$\sum_{i=1}^{M} w_i = 1, \quad 1 \le p < \infty$$

- Equivalent to taking $p \to \infty$
- Can overcome lack of convexity conditions
- May get non-Pareto solutions



Other weighted methods

Exponential weighted criterion ***(Athan, Papalambros, 1996)***

$$f_s = \sum_{i=1}^{M} \left(e^{pw_i} - 1\right) e^{pf_i(\mathbf{x})} \qquad p > 1$$

- Can overcome lack of convexity conditions
- Problems with numerical stability

Weighted product method ***(Gerasimov, 1979)***

$$f_s = \prod_{i=1}^{M} |f_i(\mathbf{x})|^{w_i}$$

- Deal with objectives with different magnitude

ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

Algorithm:

- Create a grid of possible upper bounds for each objective $g_{f_k} = \{v_0^k, \ldots, v_G^k\}$
- For each objective, consider the (d-1)-dimensional grid related to the other objectives. A single point of the grid is given by $\overrightarrow{\epsilon}_k = \{v^1, \ldots, v^{k-1}, v^{k+1}, \ldots, v^M\}\ ; v^i \in g_{f_i}$.
- For each point of the grid solve:

$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$
$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Pros/Cons:

- No convex requirement
- Solutions may not be Pareto optimal

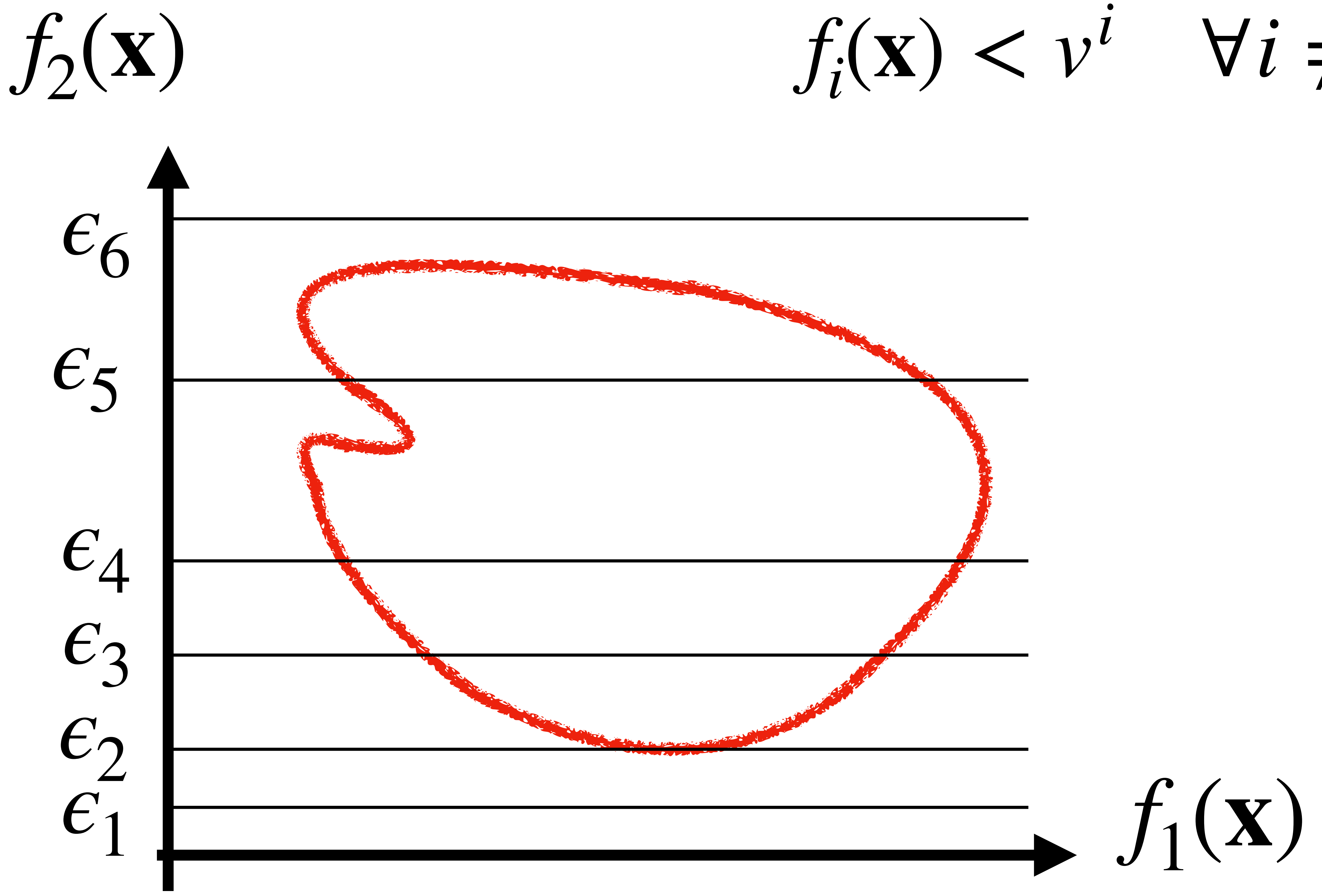
ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

Visual inspection

$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$

$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique



ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

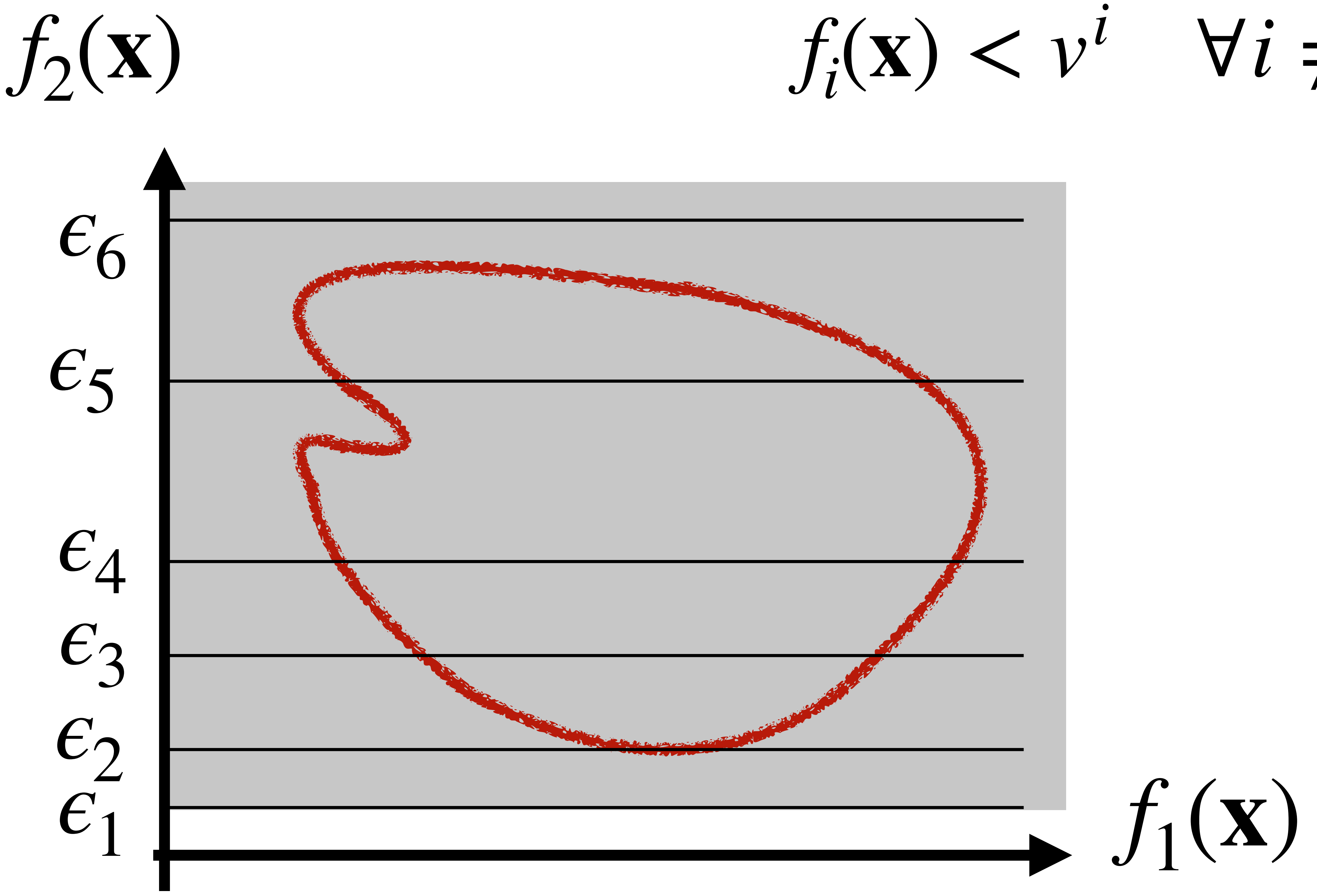
Visual inspection

$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$

$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique

ϵ_1 no solution (PO?)



ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

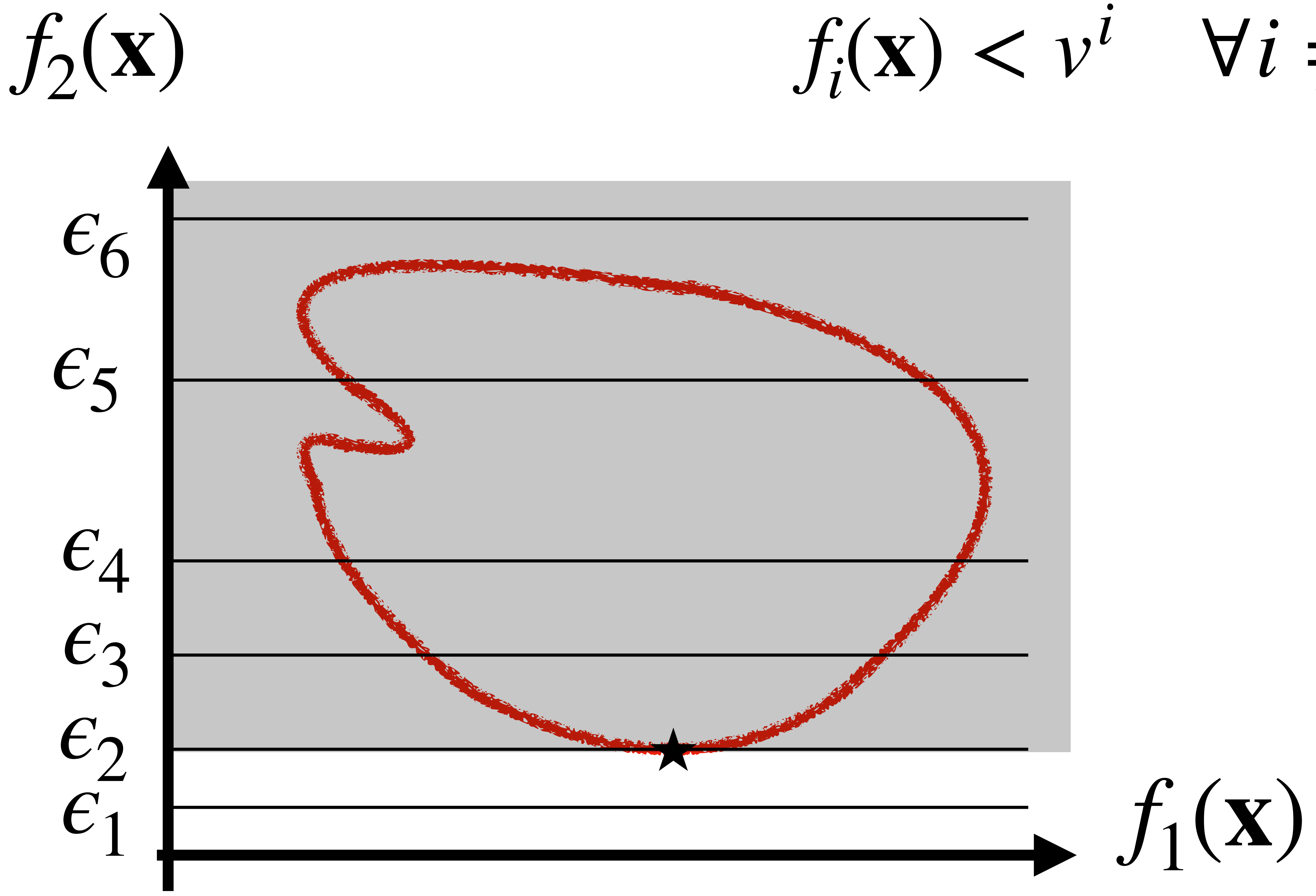
Visual inspection

$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$

$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique

ϵ_1	no solution (PO?)
ϵ_2	1 solution PO



ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

Visual inspection

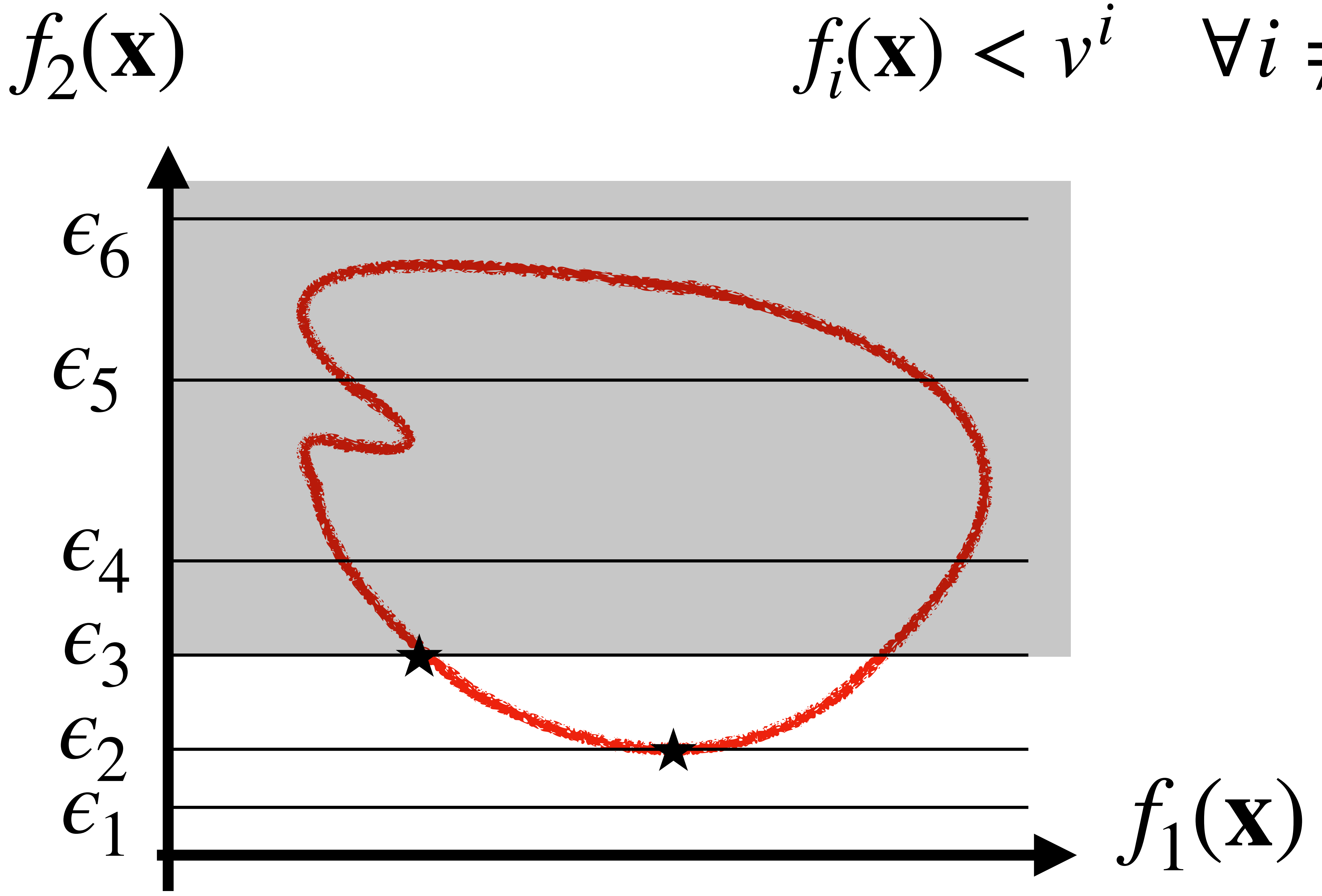
$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$

$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique

ϵ_1 no solution (PO?)

ϵ_2, ϵ_3 1 solution PO



ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

Visual inspection

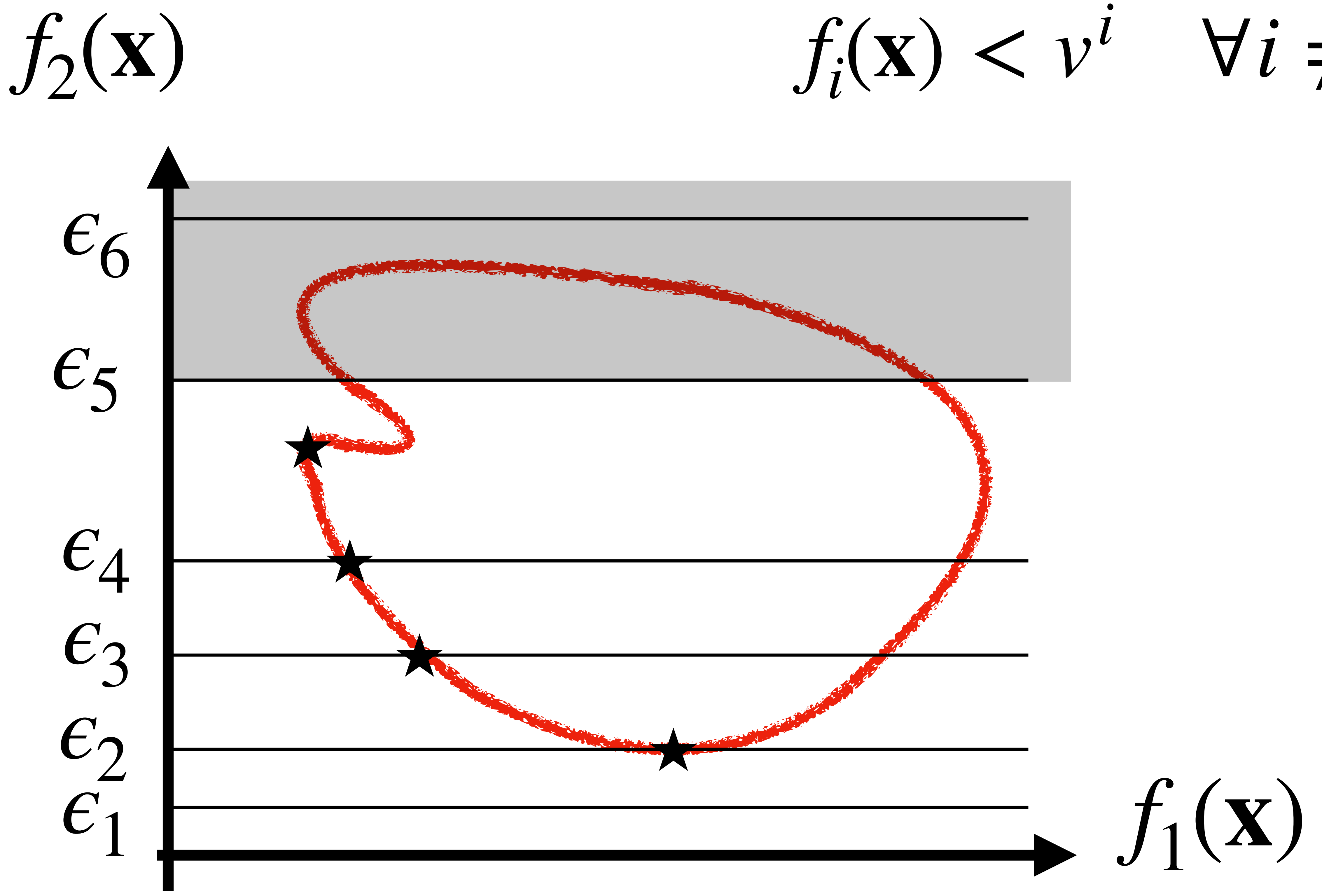
$$f_s(\overrightarrow{\epsilon}_k) = \min f_k(\mathbf{x})$$

$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique

ϵ_1 no solution (PO?)

$\epsilon_2, \ldots, \epsilon_5$ 1 solution PO



ε-constraint method *(Haimes, Lasdon, Wismer, 1971)*

Visual inspection

$$f_s(\vec{\epsilon}_k) = \min f_k(\mathbf{x})$$

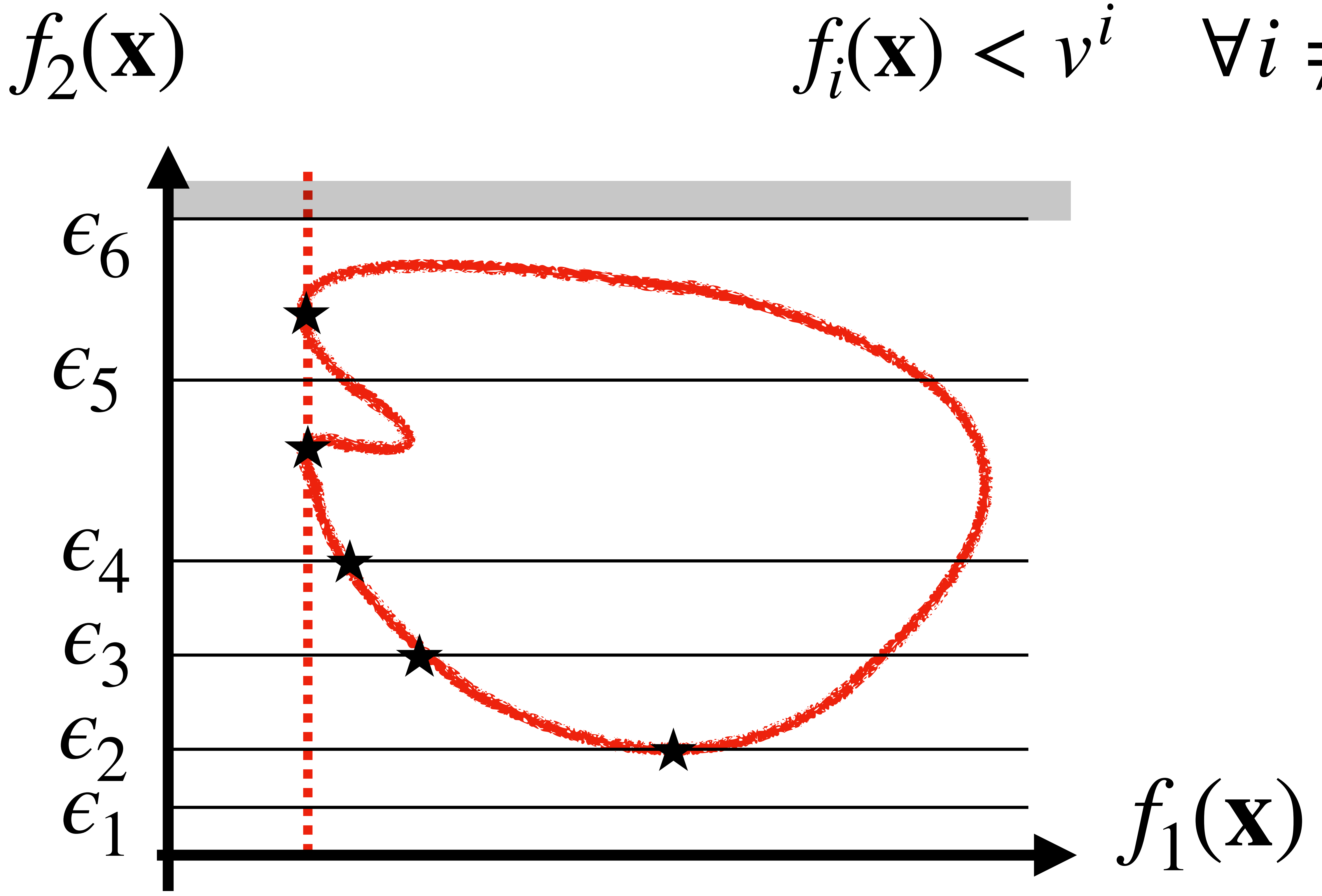
$$f_i(\mathbf{x}) < v^i \quad \forall i \neq k$$

Solutions are PO if unique

ϵ_1 no solution (PO?)

$\epsilon_2, \ldots, \epsilon_5$ 1 solution PO

ϵ_6 2 solutions (PO?)

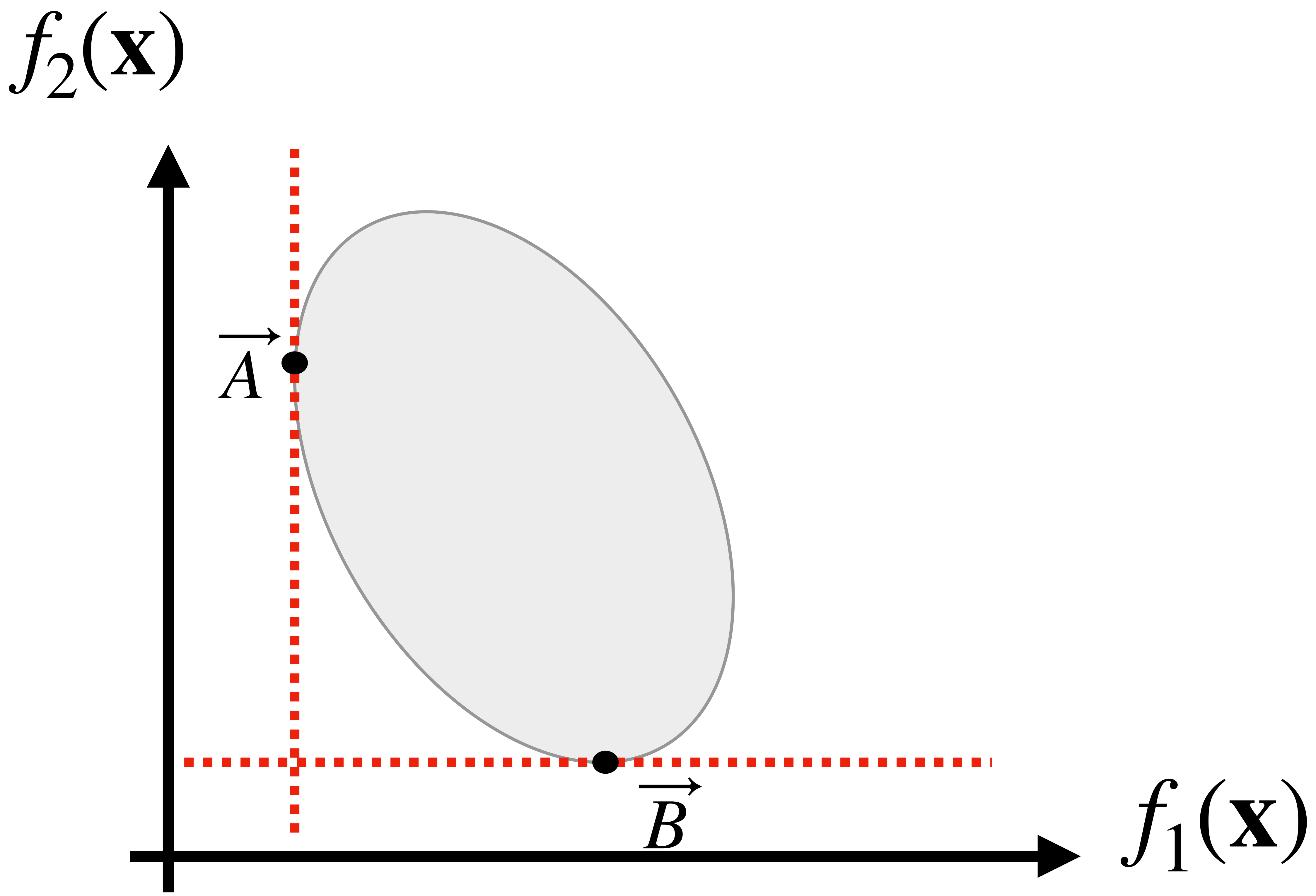


Normal boundary intersection *(Dennis, 1998; Messac, 2003)*

Algorithm:

- Find anchor points

$$\overrightarrow{A} = \{\mathbf{x}_1{}^* = \min_{x \in \mathscr{D}} f_1(\mathbf{x})\}\ ; \qquad \overrightarrow{B} = \{\mathbf{x}_2{}^* = \min_{x \in \mathscr{D}} f_2(\mathbf{x})\}$$



Normal boundary intersection *(Dennis, 1998; Messac, 2003)*

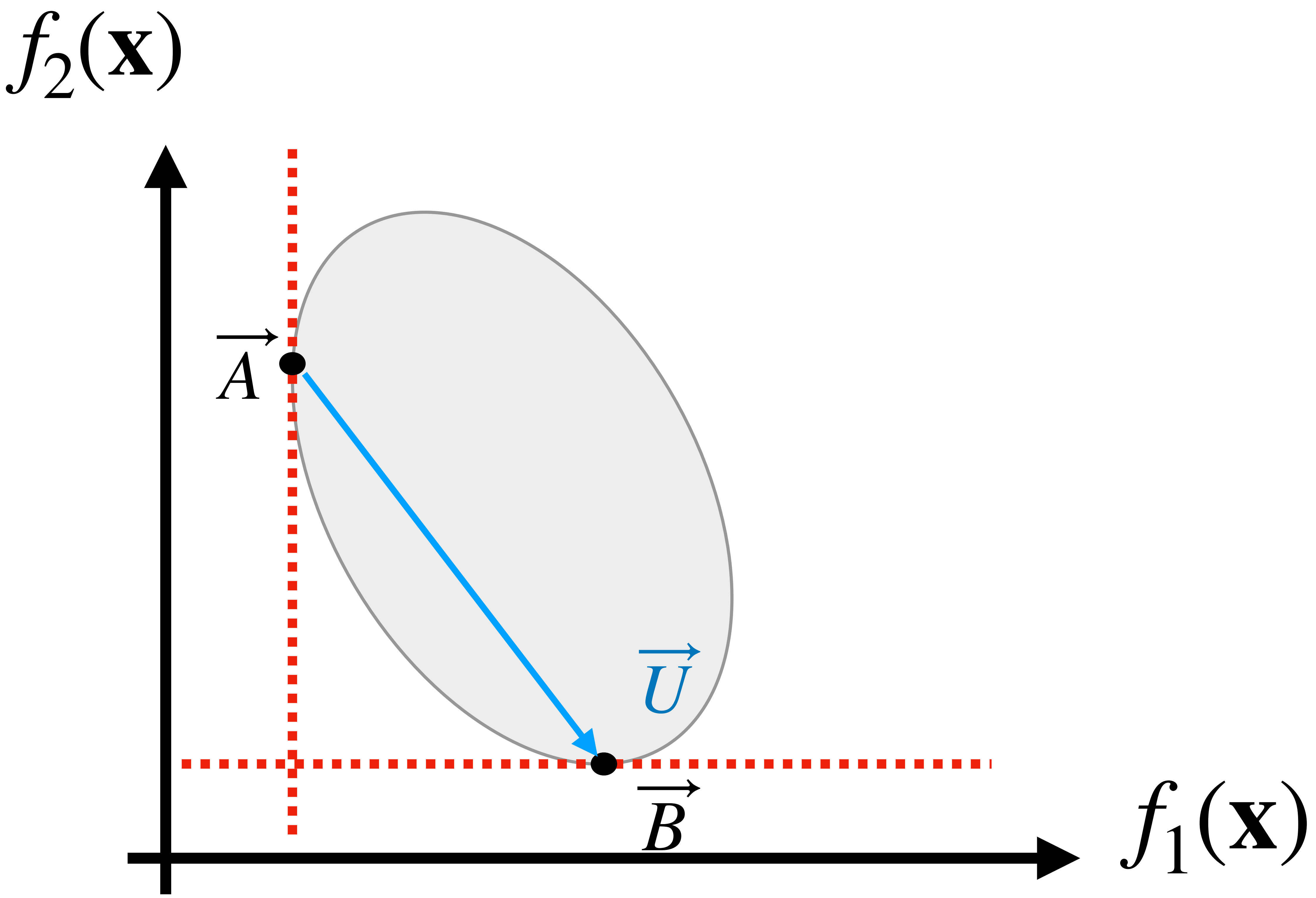
Algorithm:

- Find anchor points

$$\vec{A} = \{\mathbf{x}_1 * = \min_{x \in \mathscr{D}} f_1(\mathbf{x})\} \ ; \quad \vec{B} = \{\mathbf{x}_2 * = \min_{x \in \mathscr{D}} f_2(\mathbf{x})\}$$

- Define utopian hyperplane through anchors

$$\vec{U} = \vec{A} - \vec{B}$$



Normal boundary intersection *(Dennis, 1998; Messac, 2003)*

Algorithm:

- Find anchor points

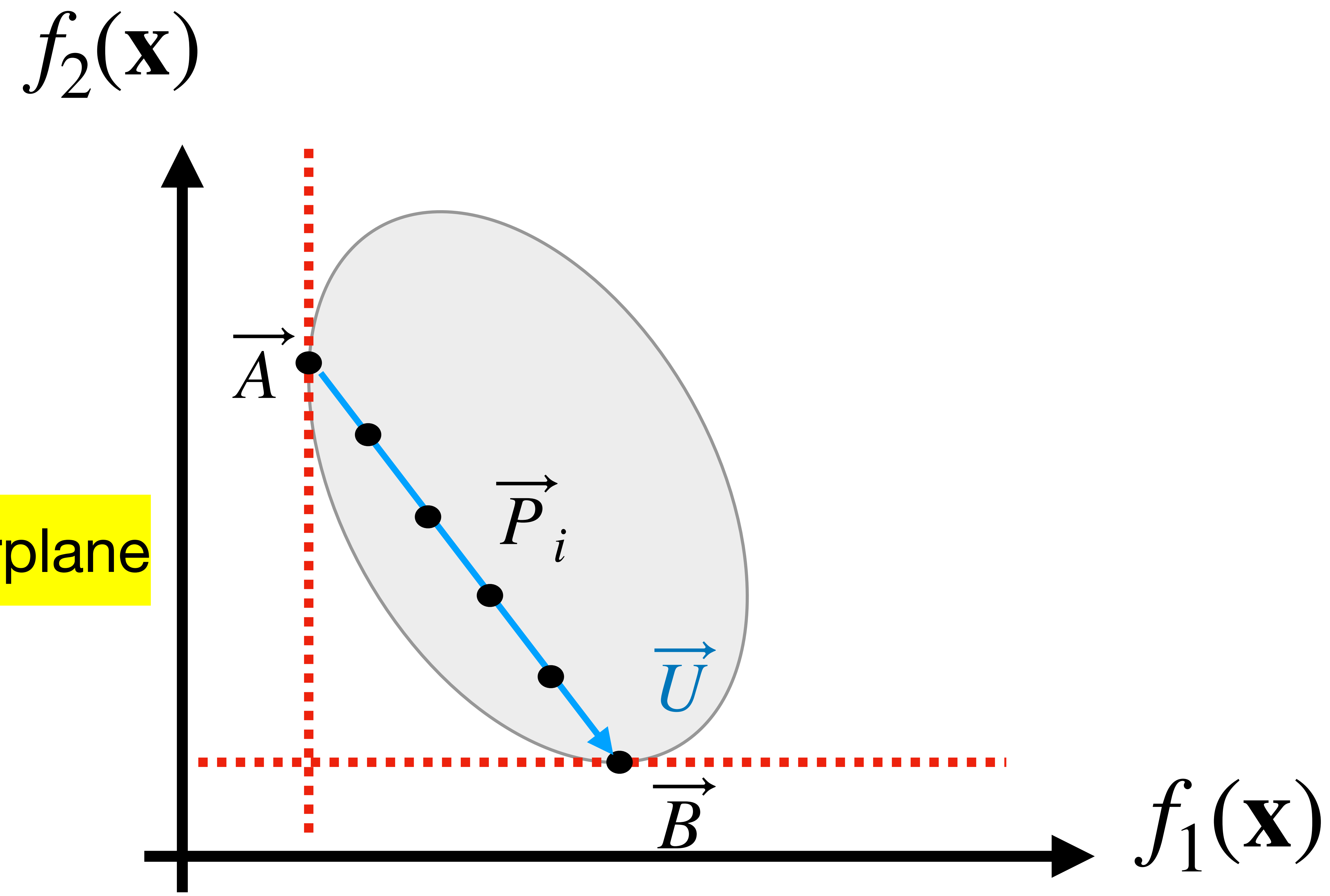
$$\vec{A} = \{\mathbf{x}_1{}^* = \min_{x \in \mathcal{D}} f_1(\mathbf{x})\}\ ; \qquad \vec{B} = \{\mathbf{x}_2{}^* = \min_{x \in \mathcal{D}} f_2(\mathbf{x})\}$$

- Define utopian hyperplane through anchors

$$\vec{U} = \vec{A} - \vec{B}$$

- Define an evenly distributed grid on the utopia hyperplane

$$W = \{\vec{w}_i\} = \{(w_{1i}, w_{2i})\} = \{(\delta i, 1 - \delta i)\}_{i=0,\ldots,1/\delta}$$
$$P = \{\vec{P}_i\} = \{w_{1i}\vec{A} + w_{2i}\vec{B}\}_{i=0,\ldots,1/\delta}$$



Normal boundary intersection *(Dennis, 1998; Messac, 2003)*

Algorithm:

- Find anchor points

$$\vec{A} = \{\mathbf{x}_1{}^* = \min_{x \in \mathscr{D}} f_1(\mathbf{x})\}\ ; \qquad \vec{B} = \{\mathbf{x}_2{}^* = \min_{x \in \mathscr{D}} f_2(\mathbf{x})\}$$

- Define utopian hyperplane through anchors

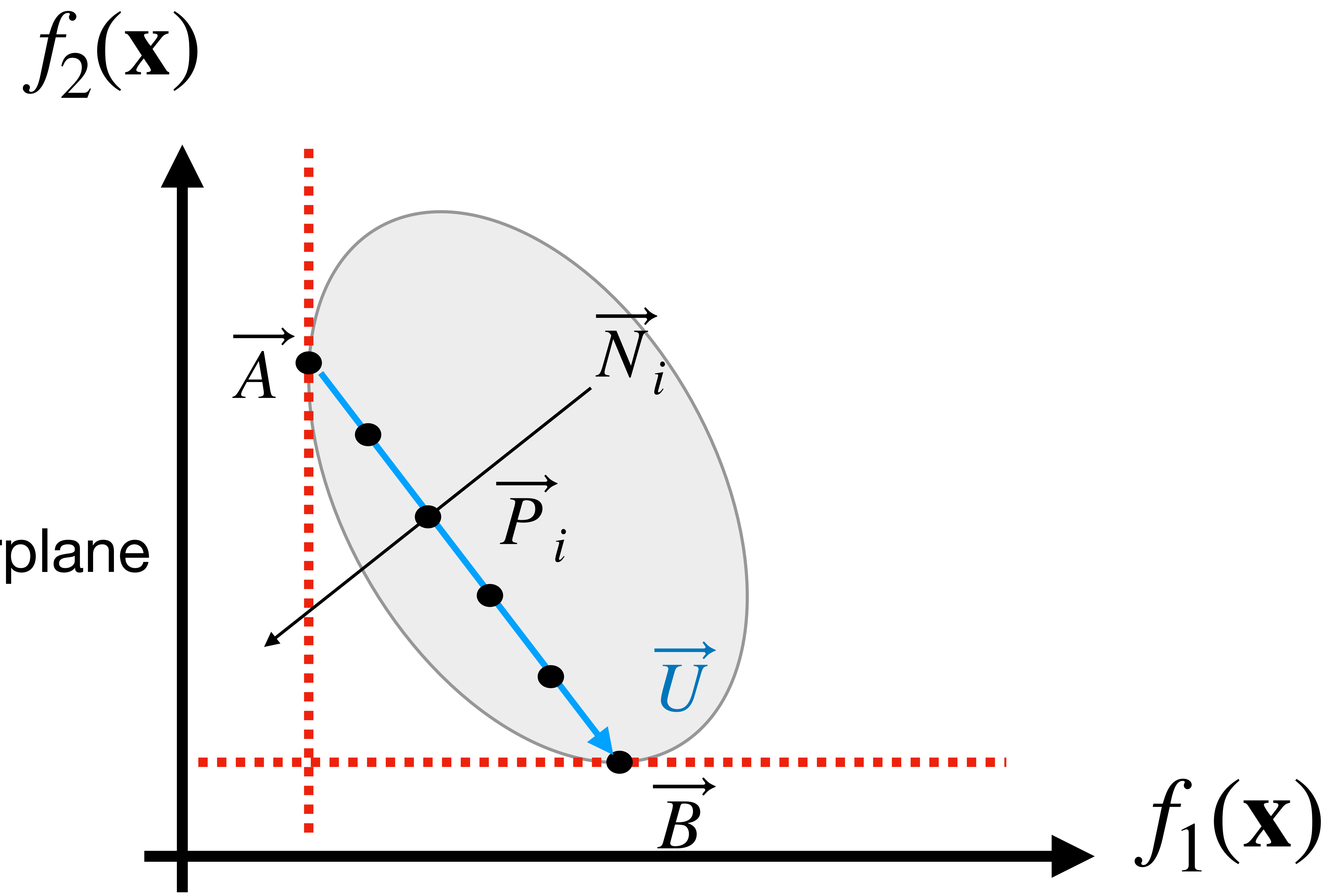
$$\vec{U} = \vec{A} - \vec{B}$$

- Define an evenly distributed grid on the utopia hyperplane

$$W = \{\vec{w}_i\} = \{(w_{1i}, w_{2i})\} = \{(\delta i, 1 - \delta i)\}_{i=0,\ldots,1/\delta}$$
$$P = \{\vec{P}_i\} = \{w_{1i}\vec{A} + w_{2i}\vec{B}\}_{i=0,\ldots,1/\delta}$$

- Define the normal vector at each grid point

$$N = \{\vec{N}_i\} \quad s.t. \quad \vec{P}_i \cdot \vec{N}_i = 0$$



Normal boundary intersection *(Dennis, 1998; Messac, 2003)*

Optimize the objectives along each normal vector

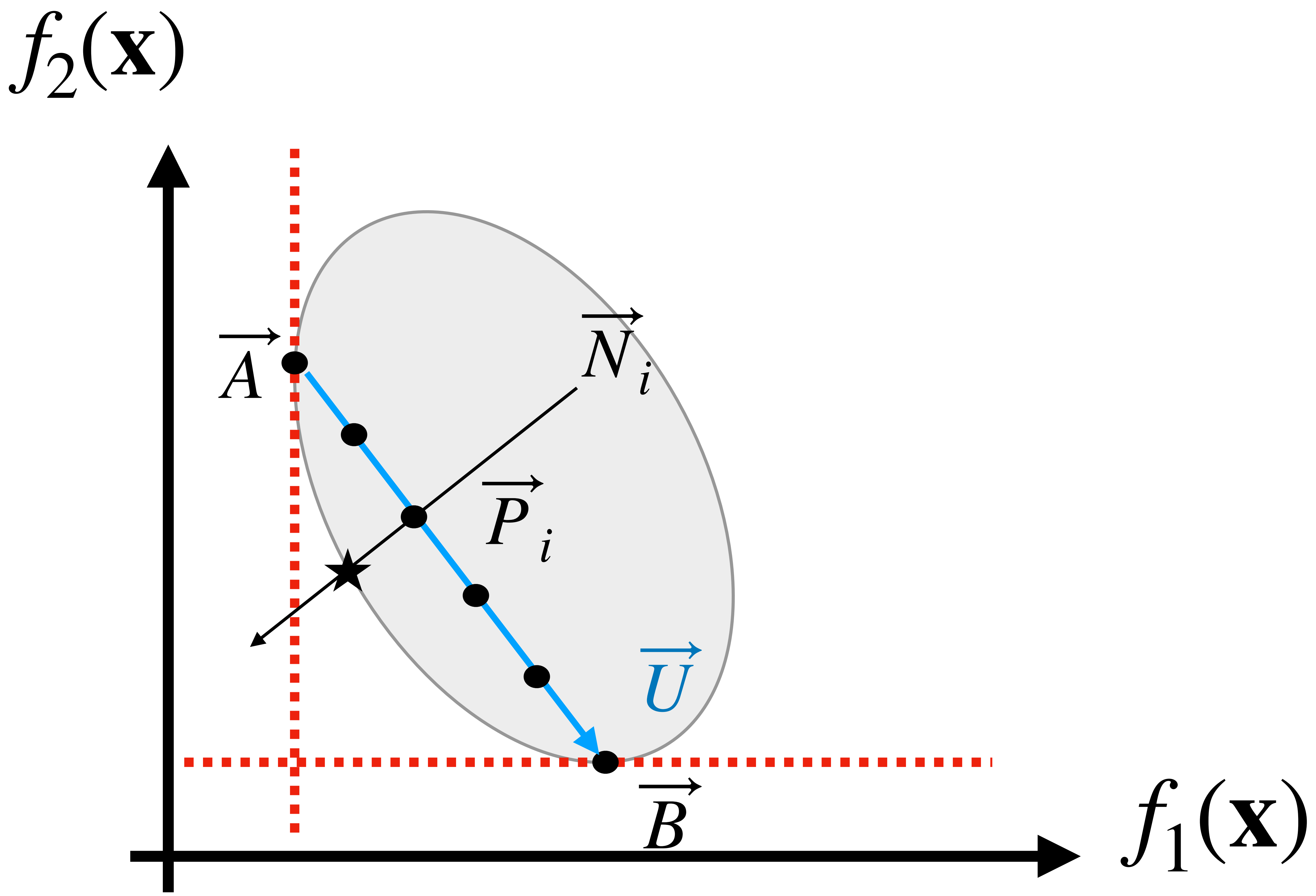
$$f_s = \min_{\mathbf{x} \in \mathscr{D}} f_1(\mathbf{x})$$

$$g_s = \overrightarrow{U} \cdot (\overrightarrow{F}(\mathbf{x}) - P_i) = 0$$

$$\overrightarrow{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))$$

Pros/Cons:

- No convex requirement
- Solutions may not be Pareto optimal



Other classical methods

- Goal programming
- Simplex method
- Physical programming
- Lexicographic methods
- ….

References:

A. *Keller, André A. Multi-objective optimization in theory and practice i: classical methods. Bentham Science Publishers, 2017.*
B. *Miettinen, Kaisa. Nonlinear multiobjective optimization. Vol. 12. Springer Science & Business Media, 1999*
C. *Blank, J., & Deb, K. (2020). Pymoo: Multi-objective optimization in python. Ieee access, 8, 89497-89509.*
D. *Deb, K., Sindhya, K., & Hakanen, J. (2016). Multi-objective optimization. In Decision sciences (pp. 161-200). CRC Press.*

Take home message

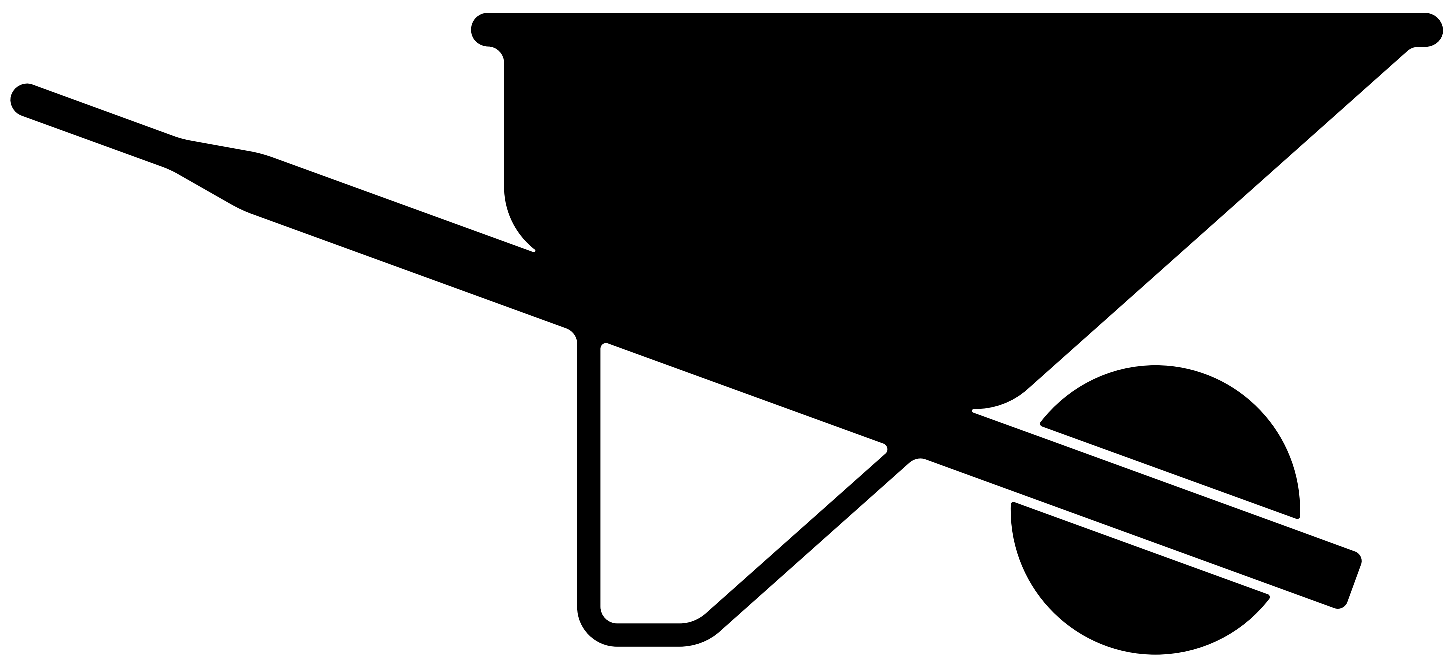
- Classical MOO algorithms usually exploit scalarization methods

- Whenever possible, **rescale objectives**. The trade-off between conflicting objectives reflects the rate of change in one objective resulting from a unit increment in another objective.

- Always check domain and objective convexity

- Scalarization methods robust to non-convexity need Pareto optimality check

Hands on



Case study: Robust Optimization

Robustness: technological insensitivity to variability in industrial and environmental factors.

What about optimization?

What if we have some **noise** in the decision variable space or in the numerical coefficients?

Case study: Robust Optimization

The expected value of the objective becomes

$$F(f, \mathbf{x}, \alpha, \epsilon) = \int \dots \int f(\mathbf{x} + \epsilon_x, \alpha + \epsilon_\alpha) p(\epsilon_x) p(\epsilon_\alpha) \prod_i d\epsilon_x^i \prod_j d\epsilon_\alpha^j$$

In 1D with only noise in x

$$F(f, x, \epsilon) = \int_{-\infty}^{\infty} f(x + \epsilon, \alpha)\, p(\epsilon)\, d\epsilon$$

Case study: Robust Optimization

So, the following constrained minimization problem

$$\min_{\mathbf{x}\in\mathscr{D}} f(\mathbf{x}, \alpha), \qquad \text{satisfying} \qquad g_j(\mathbf{x}, \alpha) \leq 0 \qquad j = 1,\ldots,J$$

Is rephrased as follows in the presence of uncertainties

$$\min_{\mathbf{x},\alpha} F\left[f(\mathbf{x}+\epsilon_x, \alpha+\epsilon_\alpha)\right] \qquad \text{satisfying} \qquad G_j\left[g_j(\mathbf{x}+\epsilon_x, \alpha+\epsilon_\alpha)\right] \leq 0 \qquad j = 1,\ldots,J$$

It incorporates uncertainties!

Case study: Robust Optimization

Let us assume for simplicity that random variables are independent:

$$p(\overrightarrow{\epsilon_x}) = \prod_{i=1}^{n} p_i(\epsilon_{xi}) \qquad p(\overrightarrow{\epsilon_\alpha}) = \prod_{\rho=1}^{P} p_\rho(\epsilon_{\alpha\rho})$$

In general, we assume p to be Gaussian PDFs. In practice, F, G use the mean and variance of f and g_j

$$F = w_1\mu_f + w_2\sigma_f \qquad G_j = \mu_{g_j} + k_j\sigma_{g_j}^2$$

Case study: Robust Optimization

$$F = w_1\mu_f + w_2\sigma_f \qquad G_j = \mu_{g_j} + k_j\sigma_{g_j}^2$$

Where (same for G_j)

$$\mu_f = \mathbf{E}[f(\mathbf{x}, \alpha)] = \int \dots \int f(\mathbf{x} + \epsilon_x, \alpha + \epsilon_\alpha) \prod_i p_i(\epsilon_x^i) d\epsilon_x^i \prod_\rho p_\rho(\epsilon_\alpha^\rho) d\epsilon_\alpha^\rho$$

$$\sigma_f = \mathbf{E}[(f(\mathbf{x}, \alpha) - \mu_f)^2] = \int \dots \int \left(f(\mathbf{x} + \epsilon_x, \alpha + \epsilon_\alpha) - \mu_f\right)^2 \prod_i p_i(\epsilon_x^i) d\epsilon_x^i \prod_\rho p_\rho(\epsilon_\alpha^\rho) d\epsilon_\alpha^\rho$$

We can simplify these formulas by expanding them around they mean value using Taylor series.

Case study: Robust Optimization

$$F = w_1\mu_f + w_2\sigma_f \qquad G_j = \mu_{g_j} + k_j\sigma_{g_j}^2$$

In practice:

$$\mu_f \simeq f(\mu_x, \mu_\alpha) \qquad \mu_{g_j} \simeq g_j(\mu_x, \mu_\alpha)$$

$$\sigma_f^2 \simeq \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \sum_{\rho=1}^{K}\left(\frac{\partial f}{\partial \alpha_\rho}\right)^2 \sigma_{\alpha_\rho}^2 \qquad \sigma_{g_j} \simeq \sum_{i=1}^{n}\left(\frac{\partial g_j}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \sum_{\rho=1}^{K}\left(\frac{\partial g_j}{\partial \alpha_\rho}\right)^2 \sigma_{\alpha_\rho}^2$$

If **tolerance/bounded uncertainties** are specified $|\epsilon_{x_i}|$, $|\epsilon_{\alpha_\rho}|$, the constraints are usually reformulated as

$$G_j = \mu_{g_j} + k_{x_j}\sum_{i=1}^{n}\left|\frac{\partial g_j}{\partial x_i}\right| |\epsilon_{x_i}| + k_{\alpha_j}\sum_{\rho=1}^{K}\left|\frac{\partial g_j}{\partial \alpha_\rho}\right| |\epsilon_{\alpha_\rho}|$$

Case study: Robust Optimization

Summarising, the minimization problem is reformulated as:

$$F = \min_{\mathbf{x}\in\mathscr{D}} \left[w_1\, f(\mu_x, \mu_\alpha) + w_2 \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \sum_{\rho=1}^{K} \left(\frac{\partial f}{\partial \alpha_\rho}\right)^2 \sigma_{\alpha_\rho}^2} \right]$$

subject to

$$G_j = \mu_{g_j} + k_{x_j} \sum_{i=1}^{n} \left| \frac{\partial g_j}{\partial x_i} \right| |\epsilon_{x_i}| + k_{\alpha_j} \sum_{\rho=1}^{K} \left| \frac{\partial g_j}{\partial \alpha_\rho} \right| |\epsilon_{\alpha_\rho}| \leq 0 \qquad j = 1, \ldots K$$

or

$$G_j = \min_{\mathbf{x}\in\mathscr{D}} \left[w_1'\, g_j(\mu_x, \mu_\alpha) + w_2' \sqrt{\sum_{i=1}^{n} \left(\frac{\partial g_j}{\partial x_i}\right)^2 \sigma_{x_i}^2 + \sum_{\rho=1}^{K} \left(\frac{\partial g_j}{\partial \alpha_\rho}\right)^2 \sigma_{\alpha_\rho}^2} \right] \leq 0 \qquad j = 1 \ldots, K$$

Hands on