Multiobjective Optimization

Optimization of Complex Systems – March 9th 2026

Andrea Brilli

Sapienza University of Rome

Introduction

$$\min\ f_1(x), \dots, f_p(x)$$
$$\text{s.t. } x \in \mathcal{F}.$$

$x \in \mathbb{R}^n$, $p > 1$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{F} \subseteq \mathbb{R}^n$.

- $\mathbb{R}^n$ is the **decision space**
- $\mathbb{R}^p$ is the **objective space**

- $x \in \mathbb{R}^n$ is a **decision vector**
- $z \in \mathbb{R}^p$ is an **objective vector**

Introduction

$$\min \; f_1(x), \ldots, f_p(x)$$
$$\text{s.t. } x \in \mathcal{F}.$$

$x \in \mathbb{R}^n$, $p > 1$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{F} \subseteq \mathbb{R}^n$.

- $\mathbb{R}^n$ is the **decision space**
- $\mathbb{R}^p$ is the **objective space**

- $x \in \mathbb{R}^n$ is a **decision vector**
- $z \in \mathbb{R}^p$ is an **objective vector**

Introduction

$$\min\ f_1(x), \ldots, f_p(x)$$
$$\text{s.t. } x \in \mathcal{F}.$$

$x \in \mathbb{R}^n$, $p > 1$, $f_i : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{F} \subseteq \mathbb{R}^n$.

- $\mathbb{R}^n$ is the **decision space**
- $\mathbb{R}^p$ is the **objective space**

- $x \in \mathbb{R}^n$ is a **decision vector**
- $z \in \mathbb{R}^p$ is an **objective vector**

Introduction

We define:

- $f(x) = (f_1(x), \dots, f_p(x)) : \mathbb{R}^n \to \mathbb{R}^p$ **objective vector function**
- $\mathcal{Z} = f(\mathcal{F})$ **feasible objective region**:

$$\mathcal{Z} = \{z \in \mathbb{R}^p : z = f(x),\ x \in \mathcal{F}\}.$$

- z^{id} **ideal objective vector**:

$$z_i^{id} = \min_{x \in \mathcal{F}} f_i(x), \qquad i = 1, \dots, p.$$

Note: we assume $z^{id} \notin \mathcal{Z}$, i.e. the objective functions are *in conflict* with each other.

Ordering in p-dimensional space

We can define on $\mathbb{R}^p$ a **partial strict ordering**, known as **Pareto dominance**.

Named after the Italian economist and sociologist **Vilfredo Pareto** (1848–1923), who first introduced the concept of efficient allocation in economics.

Ordering in p-dimensional space

Given two vectors z^1 and z^2 in $\mathbb{R}^p$, we say that:

z^1 *dominates* (in the Pareto sense) z^2, written $z^1 \preceq z^2$, if

$z_i^1 \leq z_i^2$ for **every** $i = 1, \ldots, p$, and
$z_j^1 < z_j^2$ for **some** $j \in \{1, \ldots, p\}$.

Ordering in p-dimensional space

Given two vectors z^1 and z^2 in $\mathbb{R}^p$, we say that:

z^1 *dominates* (in the Pareto sense) z^2, written $z^1 \preceq z^2$, if

$z_i^1 \leq z_i^2$ for **every** $i = 1, \ldots, p$, and
$z_j^1 < z_j^2$ for **some** $j \in \{1, \ldots, p\}$.

Note: the ordering is **only partial** — there exist pairs z^1, z^2 such that **neither**

$$z^1 \preceq z^2 \quad \text{nor} \quad z^2 \preceq z^1.$$

In this case, z^1 and z^2 are said to be **mutually non-dominated**.

Weak dominance

Given two vectors z^1 and z^2 in $\mathbb{R}^p$, we say that:

z^1 *weakly dominates* (in the Pareto sense) z^2, written $z^1 \prec z^2$, if

$z_i^1 < z_i^2$ for **every** $i = 1, \ldots, p$.

Note: if $z^1 \prec z^2$ then $z^1 \preceq z^2$, but the converse does **not** hold in general.

Pareto optimality

Given the multiobjective problem

$$\min \; f_1(x), \ldots, f_p(x) \quad \text{s.t. } x \in \mathcal{F},$$

a point $x^* \in \mathcal{F}$ is **Pareto optimal** if there exists no other $x \in \mathcal{F}$ such that

$$f(x) \preceq f(x^*).$$

In the objective space, the set of Pareto optimal points Ω_P is known as the:

- **efficient frontier**, or
- **Pareto front**.

Pareto optimality

Given the multiobjective problem

$$\min \; f_1(x), \ldots, f_p(x) \quad \text{s.t. } x \in \mathcal{F},$$

a point $x^* \in \mathcal{F}$ is **Pareto optimal** if there exists no other $x \in \mathcal{F}$ such that

$$f(x) \preceq f(x^*).$$

In the objective space, the set of Pareto optimal points Ω_P is known as the:

- **efficient frontier**, or
- **Pareto front**.

Weak Pareto optimality

A point $x^* \in \mathcal{F}$ is **weakly Pareto optimal** if there exists no other $x \in \mathcal{F}$ such that

$$f(x) < f(x^*).$$

Note: the set of Pareto optimal points (Ω_P) is contained in the set of weakly Pareto optimal points (Ω_D):

$$\Omega_P \subseteq \Omega_D.$$

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

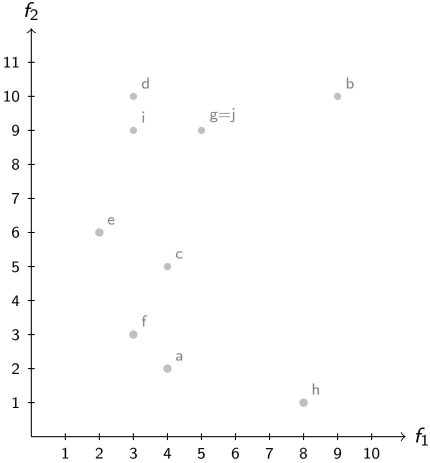
Example (1) — finding non-dominated points

Determine the set of **non-dominated** points among the following:

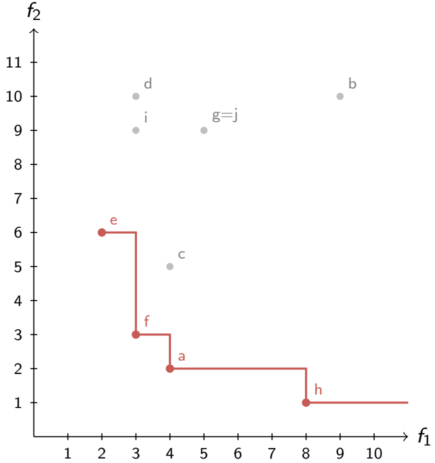
	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

	a	b	c	d	e	f	g	h	i	j
$f_1(x)$	4	9	4	3	2	3	5	8	3	5
$f_2(x)$	2	10	5	10	6	3	9	1	9	9

Example (1) — Pareto front



Example (1) — Pareto front



Non-dominated points:

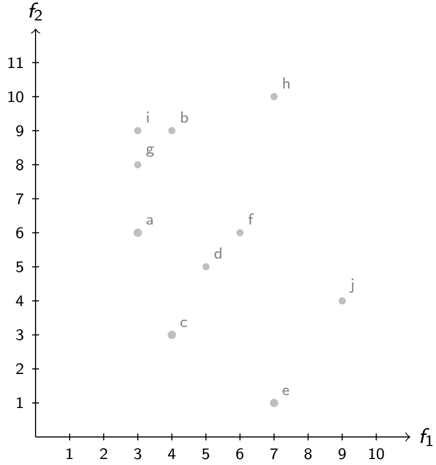
Point	(f_1, f_2)
a	$(4, 2)$
e	$(2, 6)$
f	$(3, 3)$
h	$(8, 1)$

Example (2) — finding non-dominated points

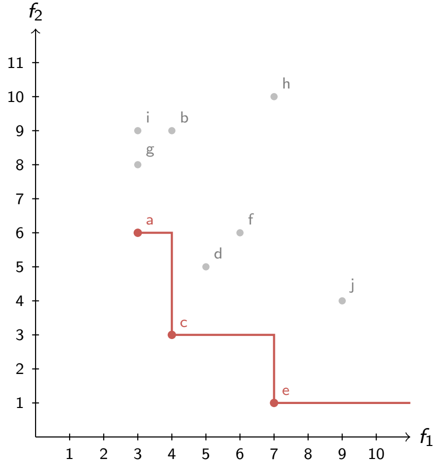
Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
f_1	3	4	4	5	7	6	3	7	3	9
f_2	6	9	3	5	1	6	8	10	9	4

Example (2) — Pareto front



Example (2) — Pareto front



Non-dominated points:

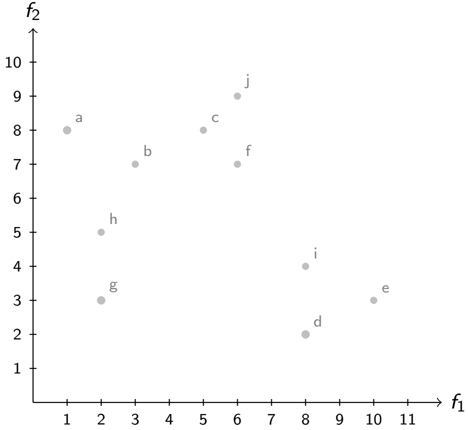
	a	c	e
f_1	3	4	7
f_2	6	3	1

Example (3) — finding non-dominated points

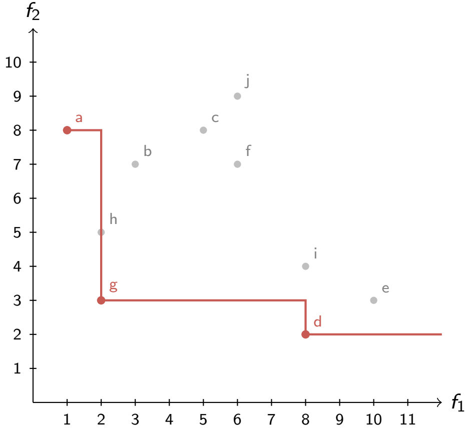
Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
f_1	1	3	5	8	10	6	2	2	8	6
f_2	8	7	8	2	3	7	3	5	4	9

Example (3) — Pareto front



Example (3) — Pareto front



Non-dominated points:

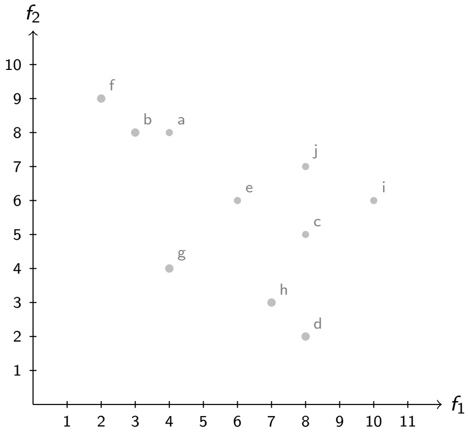
	a	d	g
f_1	1	8	2
f_2	8	2	3

Example (4) — finding non-dominated points

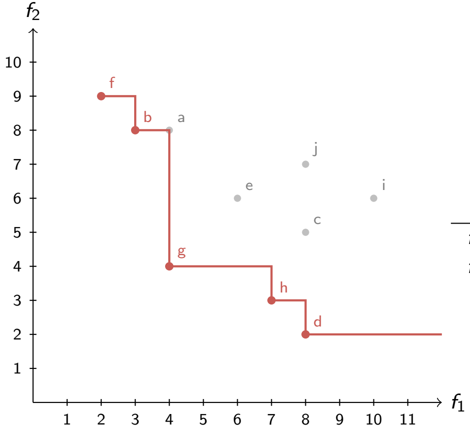
Determine the set of **non-dominated** points among the following:

	a	b	c	d	e	f	g	h	i	j
f_1	4	3	8	8	6	2	4	7	10	8
f_2	8	8	5	2	6	9	4	3	6	7

Example (4) — Pareto front



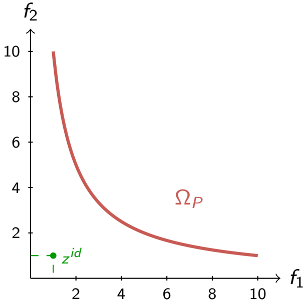
Example (4) — Pareto front



Non-dominated points:

	b	d	f	g	h
f_1	3	8	2	4	7
f_2	8	2	9	4	3

Example (5) — continuous Pareto front

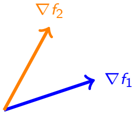


Key observation: in difficult problems the objectives cannot be simultaneously minimized.

Descent directions in the multiobjective case

In single-objective optimization, d is a descent direction at x if $\nabla f(x)^\top d < 0$.

What changes with multiple objectives? We need $\nabla f_i(x)^\top d < 0$ for *all* $i = 1, \ldots, p$ simultaneously. This may or may not be possible depending on the geometry of the gradients.

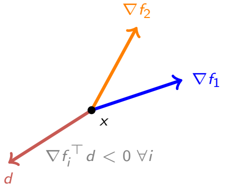


Descent directions in the multiobjective case

In single-objective optimization, d is a descent direction at x if $\nabla f(x)^\top d < 0$.

What changes with multiple objectives? We need $\nabla f_i(x)^\top d < 0$ for *all* $i = 1, \ldots, p$ simultaneously. This may or may not be possible depending on the geometry of the gradients.

Case 1: descent direction exists

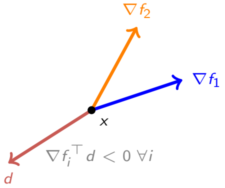


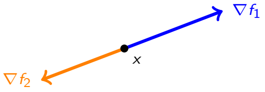
Descent directions in the multiobjective case

In single-objective optimization, d is a descent direction at x if $\nabla f(x)^\top d < 0$.

What changes with multiple objectives? We need $\nabla f_i(x)^\top d < 0$ for *all* $i = 1, \ldots, p$ simultaneously. This may or may not be possible depending on the geometry of the gradients.

Case 1: descent direction exists



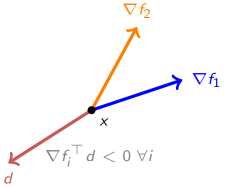


Descent directions in the multiobjective case

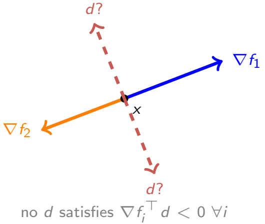
In single-objective optimization, d is a descent direction at x if $\nabla f(x)^\top d < 0$.

What changes with multiple objectives? We need $\nabla f_i(x)^\top d < 0$ for *all* $i = 1, \ldots, p$ simultaneously. This may or may not be possible depending on the geometry of the gradients.

Case 1: descent direction exists



Case 2: no descent direction



Common descent directions

Definition (Common descent direction)

Given $f : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable, a direction $d \in \mathbb{R}^n$, $d \neq 0$, is a **common descent direction** *for f at x if*

$$\nabla f_i(x)^\top d < 0, \qquad i = 1, \ldots, p.$$

Geometric interpretation. d must lie in the open half-space $\{d : \nabla f_i(x)^\top d < 0\}$ for *every i* simultaneously, i.e., in the intersection:

$$\mathcal{D}(x) = \bigcap_{i=1}^{p} \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0\}.$$

Question: When is $\mathcal{D}(x)$ non-empty? What geometric condition on the gradients $\nabla f_1(x), \ldots, \nabla f_p(x)$ guarantees this?

$\mathcal{D}(x) \neq \emptyset$ if and only if 0 **does not belong to the convex hull of the gradients**, i.e.,

$$0 \notin \mathrm{conv}\{\nabla f_1(x), \ldots, \nabla f_p(x)\}.$$

Common descent directions

Definition (Common descent direction)

Given $f : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable, a direction $d \in \mathbb{R}^n$, $d \neq 0$, is a **common descent direction** *for f at x if*

$$\nabla f_i(x)^\top d < 0, \qquad i = 1, \ldots, p.$$

Geometric interpretation. d must lie in the open half-space $\{d : \nabla f_i(x)^\top d < 0\}$ for *every* i simultaneously, i.e., in the intersection:

$$\mathcal{D}(x) = \bigcap_{i=1}^{p} \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0\}.$$

Question: When is $\mathcal{D}(x)$ non-empty? What geometric condition on the gradients $\nabla f_1(x), \ldots, \nabla f_p(x)$ guarantees this?

$\mathcal{D}(x) \neq \emptyset$ if and only if 0 **does not belong to the convex hull of the gradients**, i.e.,

$$0 \notin \operatorname{conv}\{\nabla f_1(x), \ldots, \nabla f_p(x)\}.$$

Common descent directions

Definition (Common descent direction)

Given $f : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable, a direction $d \in \mathbb{R}^n$, $d \neq 0$, is a **common descent direction** *for f at x if*

$$\nabla f_i(x)^\top d < 0, \qquad i = 1, \dots, p.$$

Geometric interpretation. d must lie in the open half-space $\{d : \nabla f_i(x)^\top d < 0\}$ for *every* i simultaneously, i.e., in the intersection:

$$\mathcal{D}(x) = \bigcap_{i=1}^{p} \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0\}.$$

Question: When is $\mathcal{D}(x)$ non-empty? What geometric condition on the gradients $\nabla f_1(x), \dots, \nabla f_p(x)$ guarantees this?

$\mathcal{D}(x) \neq \emptyset$ if and only if 0 **does not belong to the convex hull of the gradients**, i.e.,

$$0 \notin \operatorname{conv}\{\nabla f_1(x), \dots, \nabla f_p(x)\}.$$

Common descent directions

Definition (Common descent direction)

Given $f : \mathbb{R}^n \to \mathbb{R}^p$ continuously differentiable, a direction $d \in \mathbb{R}^n$, $d \neq 0$, is a **common descent direction** *for f at x if*

$$\nabla f_i(x)^\top d < 0, \qquad i = 1, \dots, p.$$

Geometric interpretation. d must lie in the open half-space $\{d : \nabla f_i(x)^\top d < 0\}$ for *every i* simultaneously, i.e., in the intersection:

$$\mathcal{D}(x) = \bigcap_{i=1}^{p} \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0\}.$$

Question: When is $\mathcal{D}(x)$ non-empty? What geometric condition on the gradients $\nabla f_1(x), \dots, \nabla f_p(x)$ guarantees this?

$\mathcal{D}(x) \neq \emptyset$ if and only if 0 **does not belong to the convex hull of the gradients**, i.e.,

$$0 \notin \operatorname{conv}\{\nabla f_1(x), \dots, \nabla f_p(x)\}.$$

Unconstrained multiobjective problem

We now focus on the **unconstrained** case:

$$\min_{x \in \mathbb{R}^n} f(x) = (f_1(x), \ldots, f_p(x))^\top$$

with $f_i : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable for all $i = 1, \ldots, p$.

Definition (Descent direction set)

Given $x \in \mathbb{R}^n$, define:

$$\mathcal{D}_P(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x + \beta d) \preceq f(x)\ \forall \beta \in (0, \alpha],\ \alpha > 0\}.$$

$\mathcal{D}_P(x)$ is the set of directions along which f **does not increase** *(in the Pareto sense) for small steps.*

The set $\mathcal{D}(x)$ defined earlier satisfies:

$$\mathcal{D}(x) = \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0,\ i = 1, \ldots, p\} \subseteq \mathcal{D}_P(x).$$

Unconstrained multiobjective problem

We now focus on the **unconstrained** case:

$$\min_{x \in \mathbb{R}^n} f(x) = (f_1(x), \dots, f_p(x))^\top$$

with $f_i : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable for all $i = 1, \dots, p$.

Definition (Descent direction set)

Given $x \in \mathbb{R}^n$, define:

$$\mathcal{D}_P(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x + \beta d) \preceq f(x)\ \forall \beta \in (0, \alpha],\ \alpha > 0\}.$$

$\mathcal{D}_P(x)$ is the set of directions along which f **does not increase** *(in the Pareto sense) for small steps.*

The set $\mathcal{D}(x)$ defined earlier satisfies:

$$\mathcal{D}(x) = \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0,\ i = 1, \dots, p\} \subseteq \mathcal{D}_P(x).$$

Unconstrained multiobjective problem

We now focus on the **unconstrained** case:

$$\min_{x \in \mathbb{R}^n} f(x) = (f_1(x), \dots, f_p(x))^\top$$

with $f_i : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable for all $i = 1, \dots, p$.

Definition (Descent direction set)

Given $x \in \mathbb{R}^n$, define:

$$\mathcal{D}_P(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x + \beta d) \preceq f(x)\ \forall \beta \in (0, \alpha],\ \alpha > 0\}.$$

$\mathcal{D}_P(x)$ is the set of directions along which f **does not increase** *(in the Pareto sense) for small steps.*

The set $\mathcal{D}(x)$ defined earlier satisfies:

$$\mathcal{D}(x) = \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0,\ i = 1, \dots, p\} \subseteq \mathcal{D}_P(x).$$

Necessary condition for Pareto optimality

Proposition (Necessary condition)

If $x^* \in \mathbb{R}^n$ is a (local) Pareto optimum of the unconstrained problem, then

$$\mathcal{D}_P(x^*) = \emptyset,$$

i.e., no common descent direction exists at x^*.

Since $\mathcal{D}(x) \subseteq \mathcal{D}_P(x)$, a necessary condition is also:

$$\mathcal{D}(x^*) = \emptyset, \qquad \text{i.e.,} \quad 0 \in \operatorname{conv}\{\nabla f_1(x^*), \ldots, \nabla f_p(x^*)\}.$$

This means: at a Pareto optimum, the gradients must **surround the origin** — it is impossible to find a direction that simultaneously decreases all objectives.

Question: How does this reduce to the familiar condition $\nabla f(x^*) = 0$ when $p = 1$?

Necessary condition for Pareto optimality

Proposition (Necessary condition)

If $x^* \in \mathbb{R}^n$ is a (local) Pareto optimum of the unconstrained problem, then

$$\mathcal{D}_P(x^*) = \emptyset,$$

i.e., no common descent direction exists at x^*.

Since $\mathcal{D}(x) \subseteq \mathcal{D}_P(x)$, a necessary condition is also:

$$\mathcal{D}(x^*) = \emptyset, \qquad \text{i.e.,} \quad 0 \in \mathrm{conv}\{\nabla f_1(x^*), \dots, \nabla f_p(x^*)\}.$$

This means: at a Pareto optimum, the gradients must **surround the origin** — it is impossible to find a direction that simultaneously decreases all objectives.

Question: How does this reduce to the familiar condition $\nabla f(x^*) = 0$ when $p = 1$?

Necessary condition for Pareto optimality

Proposition (Necessary condition)

If $x^* \in \mathbb{R}^n$ is a (local) Pareto optimum of the unconstrained problem, then

$$\mathcal{D}_P(x^*) = \emptyset,$$

i.e., no common descent direction exists at x^*.

Since $\mathcal{D}(x) \subseteq \mathcal{D}_P(x)$, a necessary condition is also:

$$\mathcal{D}(x^*) = \emptyset, \qquad \text{i.e.,} \quad 0 \in \operatorname{conv}\{\nabla f_1(x^*), \ldots, \nabla f_p(x^*)\}.$$

This means: at a Pareto optimum, the gradients must **surround the origin** — it is impossible to find a direction that simultaneously decreases all objectives.

Question: How does this reduce to the familiar condition $\nabla f(x^*) = 0$ when $p = 1$?

Necessary condition for Pareto optimality

Proposition (Necessary condition)

If $x^* \in \mathbb{R}^n$ is a (local) Pareto optimum of the unconstrained problem, then

$$\mathcal{D}_P(x^*) = \emptyset,$$

i.e., no common descent direction exists at x^*.

Since $\mathcal{D}(x) \subseteq \mathcal{D}_P(x)$, a necessary condition is also:

$$\mathcal{D}(x^*) = \emptyset, \qquad \text{i.e.,} \quad 0 \in \operatorname{conv}\{\nabla f_1(x^*), \ldots, \nabla f_p(x^*)\}.$$

This means: at a Pareto optimum, the gradients must **surround the origin** — it is impossible to find a direction that simultaneously decreases all objectives.

Question: How does this reduce to the familiar condition $\nabla f(x^*) = 0$ when $p = 1$?

The stationarity measure $\theta(x)$

Definition (Stationarity measure — min-max form)

$$\theta(x) = -\min_{\|d\|\leq 1} \max_{i=1,\dots,p} \nabla f_i(x)^\top d \geq 0.$$

Key properties:

- $\theta(x) \geq 0$ for all x
- $\theta(x) = 0 \iff \mathcal{D}(x) = \emptyset \iff x$ is **Pareto stationary**
- $\theta(x) > 0 \implies x$ is **not** a local Pareto optimum, and the minimizer

$$d^* = \arg\min_{\|d\|\leq 1} \max_i \nabla f_i(x)^\top d$$

is a common descent direction with guaranteed worst-case descent rate $\theta(x)$

Connection to single-objective case ($p = 1$):

$$\theta(x) = -\min_{\|d\|\leq 1} \nabla f(x)^\top d = \|\nabla f(x)\|, \quad \text{with} \quad d^* = -\nabla f(x)/\|\nabla f(x)\|.$$

The stationarity measure $\theta(x)$

Definition (Stationarity measure — min-max form)

$$\theta(x) = -\min_{\|d\|\leq 1} \max_{i=1,\ldots,p} \nabla f_i(x)^\top d \geq 0.$$

Key properties:

- $\theta(x) \geq 0$ for all x
- $\theta(x) = 0 \iff \mathcal{D}(x) = \emptyset \iff x$ is **Pareto stationary**
- $\theta(x) > 0 \implies x$ is **not** a local Pareto optimum, and the minimizer

$$d^* = \arg\min_{\|d\|\leq 1} \max_i \nabla f_i(x)^\top d$$

is a common descent direction with guaranteed worst-case descent rate $\theta(x)$

Connection to single-objective case ($p = 1$):

$$\theta(x) = -\min_{\|d\|\leq 1} \nabla f(x)^\top d = \|\nabla f(x)\|, \quad \text{with} \quad d^* = -\nabla f(x)/\|\nabla f(x)\|.$$

The stationarity measure $\theta(x)$

Definition (Stationarity measure — min-max form)

$$\theta(x) = -\min_{\|d\|\leq 1} \max_{i=1,\dots,p} \nabla f_i(x)^\top d \geq 0.$$

Key properties:

- $\theta(x) \geq 0$ for all x
- $\theta(x) = 0 \iff \mathcal{D}(x) = \emptyset \iff x$ is **Pareto stationary**
- $\theta(x) > 0 \implies x$ is **not** a local Pareto optimum, and the minimizer

$$d^* = \arg\min_{\|d\|\leq 1} \max_i \nabla f_i(x)^\top d$$

is a common descent direction with guaranteed worst-case descent rate $\theta(x)$

Connection to single-objective case ($p = 1$):

$$\theta(x) = -\min_{\|d\|\leq 1} \nabla f(x)^\top d = \|\nabla f(x)\|, \quad \text{with} \quad d^* = -\nabla f(x)/\|\nabla f(x)\|.$$

$\theta(x)$ and the steepest descent direction

Question: how do we actually compute d^* and $\theta(x)$?

Convex hull characterization: The stationarity measure admits the equivalent form

$$\theta(x) = \min_{\lambda \in \Delta^p} \left\| \sum_{i=1}^{p} \lambda_i \nabla f_i(x) \right\|, \qquad \Delta^p = \left\{ \lambda \geq 0 : \textstyle\sum_{i=1}^{p} \lambda_i = 1 \right\},$$

and the steepest common descent direction is

$$d^* = -\frac{1}{\theta(x)} \sum_{i=1}^{p} \lambda_i^* \nabla f_i(x), \quad \text{where} \quad \lambda^* = \arg\min_{\lambda \in \Delta^p} \left\| \sum_i \lambda_i \nabla f_i(x) \right\|.$$

Geometric interpretation: $\theta(x)$ is the **distance from the origin to the convex hull** of the gradients $\{\nabla f_1(x), \ldots, \nabla f_p(x)\}$:

- $\theta(x) > 0$: origin lies *outside* the convex hull $\rightarrow$ a common descent direction exists
- $\theta(x) = 0$: origin lies *inside* (or on the boundary of) the convex hull $\rightarrow$ Pareto stationary

$\theta(x)$ and the steepest descent direction

Question: how do we actually compute d^* and $\theta(x)$?

Convex hull characterization: The stationarity measure admits the equivalent form

$$\theta(x) = \min_{\lambda \in \Delta^p} \left\| \sum_{i=1}^{p} \lambda_i \nabla f_i(x) \right\|, \qquad \Delta^p = \left\{ \lambda \geq 0 : \sum_{i=1}^{p} \lambda_i = 1 \right\},$$

and the steepest common descent direction is

$$d^* = -\frac{1}{\theta(x)} \sum_{i=1}^{p} \lambda_i^* \nabla f_i(x), \quad \text{where} \quad \lambda^* = \arg\min_{\lambda \in \Delta^p} \left\| \sum_{i} \lambda_i \nabla f_i(x) \right\|.$$

Geometric interpretation: $\theta(x)$ is the **distance from the origin to the convex hull** of the gradients $\{\nabla f_1(x), \ldots, \nabla f_p(x)\}$:

- $\theta(x) > 0$: origin lies *outside* the convex hull $\rightarrow$ a common descent direction exists
- $\theta(x) = 0$: origin lies *inside* (or on the boundary of) the convex hull $\rightarrow$ Pareto stationary

$\theta(x)$ and the steepest descent direction

Question: how do we actually compute d^* and $\theta(x)$?

Convex hull characterization: The stationarity measure admits the equivalent form

$$\theta(x) = \min_{\lambda \in \Delta^p} \left\| \sum_{i=1}^{p} \lambda_i \nabla f_i(x) \right\|, \qquad \Delta^p = \left\{ \lambda \geq 0 : \textstyle\sum_{i=1}^{p} \lambda_i = 1 \right\},$$

and the steepest common descent direction is

$$d^* = -\frac{1}{\theta(x)} \sum_{i=1}^{p} \lambda_i^* \nabla f_i(x), \quad \text{where} \quad \lambda^* = \arg\min_{\lambda \in \Delta^p} \left\| \sum_i \lambda_i \nabla f_i(x) \right\|.$$

Geometric interpretation: $\theta(x)$ is the **distance from the origin to the convex hull** of the gradients $\{\nabla f_1(x), \ldots, \nabla f_p(x)\}$:

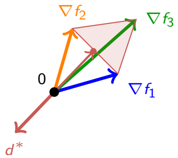
- $\theta(x) > 0$: origin lies *outside* the convex hull $\rightarrow$ a common descent direction exists
- $\theta(x) = 0$: origin lies *inside* (or on the boundary of) the convex hull $\rightarrow$ Pareto stationary

$\theta(x)$: geometric interpretation

$\theta(x)$ is the **distance from the origin to the convex hull** of the gradients $\{\nabla f_1(x), \nabla f_2(x), \nabla f_3(x)\}$:

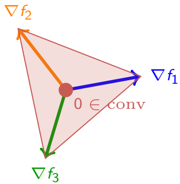
- $\theta(x) > 0$: origin lies *outside* the convex hull $\rightarrow$ common descent direction exists
- $\theta(x) = 0$: origin lies *inside or on the boundary* of the convex hull $\rightarrow$ Pareto stationary

Case 1: $\theta(x) > 0$



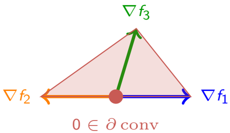
All gradients in same half-space. Descent direction d^* exists.

Case 2: $\theta(x) = 0$



Origin inside convex hull. Positive spanning set. x is Pareto stationary.

Case 3: $\theta(x) = 0$ (edge)



$\nabla f_1 = -\nabla f_2$: origin on boundary of convex hull. x is Pareto stationary.

Necessary optimality condition (Pareto)

Theorem (Necessary condition for Pareto optimality)

Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable. If x^* is a local Pareto optimum, then x^* is a **Pareto stationary point**:

$$\theta(x^*) = 0,$$

equivalently, $0 \in \operatorname{conv}\{\nabla f_1(x^*), \dots, \nabla f_p(x^*)\}$, *i.e., there exist* $\lambda_1, \dots, \lambda_p \geq 0$ *with* $\sum_{i=1}^{p} \lambda_i = 1$ *such that*

$$\sum_{i=1}^{p} \lambda_i \nabla f_i(x^*) = 0.$$

Remark: when $p = 1$, this reduces to $\lambda_1 \nabla f(x^*) = 0$ with $\lambda_1 = 1$, i.e., $\nabla f(x^*) = 0$ — the standard first-order condition.

Necessary condition for weak Pareto optimality

Definition (Weak descent direction set)

$$\mathcal{D}_w(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x+\beta d) \prec f(x)\ \forall \beta \in (0,\alpha],\ \alpha > 0\}.$$

Since $\mathcal{D}(x) \subseteq \mathcal{D}_w(x)$, the same argument applies:

Proposition

If x^* is a (local) weakly Pareto optimal point, then $\mathcal{D}_w(x^*) = \emptyset$, i.e.,

$$\theta(x^*) = 0.$$

Remark: Pareto optimality $\Rightarrow$ weak Pareto optimality $\Rightarrow$ Pareto stationarity. The converse implications do **not** hold in general.

$$\Omega_P \subseteq \Omega_D \subseteq \{x : \theta(x) = 0\}.$$

Necessary condition for weak Pareto optimality

Definition (Weak descent direction set)

$$\mathcal{D}_w(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x+\beta d) \prec f(x)\ \forall \beta \in (0,\alpha],\ \alpha > 0\}.$$

Since $\mathcal{D}(x) \subseteq \mathcal{D}_w(x)$, the same argument applies:

Proposition

If x^* is a (local) weakly Pareto optimal point, then $\mathcal{D}_w(x^*) = \emptyset$, i.e.,

$$\theta(x^*) = 0.$$

Remark: Pareto optimality $\Rightarrow$ weak Pareto optimality $\Rightarrow$ Pareto stationarity. The converse implications do **not** hold in general.

$$\Omega_P \subseteq \Omega_D \subseteq \{x : \theta(x) = 0\}.$$

Summary: descent direction sets

Let $f : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable and let $x \in \mathbb{R}^n$.

Pareto: $\mathcal{D}_P(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x+\beta d) \preceq f(x)\ \forall \beta \in (0,\alpha],\ \exists\, \alpha > 0\}$

Weak: $\mathcal{D}_w(x) = \{d \in \mathbb{R}^n : d \neq 0,\ f(x+\beta d) \prec f(x)\ \forall \beta \in (0,\alpha],\ \exists\, \alpha > 0\}$

First-order: $\mathcal{D}(x) = \{d \in \mathbb{R}^n : \nabla f_i(x)^\top d < 0,\ i = 1, \ldots, k\}$

Inclusion structure:

$$\mathcal{D}(x)\ \subseteq\ \mathcal{D}_w(x)\ \subseteq\ \mathcal{D}_P(x)$$

Optimality via empty cones:

Optimality type	**Necessary condition**	**Equivalent to**
Pareto optimum	$\mathcal{D}_P(x^*) = \emptyset$	$\Rightarrow \mathcal{D}(x^*) = \emptyset$
Weak Pareto optimum	$\mathcal{D}_w(x^*) = \emptyset$	$\Rightarrow \mathcal{D}(x^*) = \emptyset$
Pareto stationary	$\mathcal{D}(x^*) = \emptyset$	$\theta(x^*) = 0$