MULTICRITERIA OPTIMISATION, A PRIMER

PRELIMINARY VERSION

MARKUS GRASMAIR

Up to now, we have discussed optimisation problems of the form $\min_{x\in\Omega} f(x)$, where $\Omega \subset \mathbb{R}^d$ was some feasible set and we were given a cost function $f\colon \mathbb{R}^d \to \mathbb{R}$. Our goal is to find the (an) optimal point $x^* \in \Omega$, that is, a point x^* such that $f(x^*) \le f(x)$ for all other points $x \in \Omega$.

In some situations, however, it can happen that we have different competing interests at play at the same time. For instance, if we want to optimise the transport of goods, we have the two objectives of transport time and transport cost, and we would like to transport the goods as fast and as cheaply as possible. Since faster transportation usually is more expensive, this is in general not possible, though. The goal of this note is to develop a notion of solutions of these multi-criteria optimisation problems and also discuss a possible solution approach. More information about this topic can, for instance, be found in the books [2, 3].

1. Partial Orders and Cones

We start by looking at the multi-criteria optimisation problem in a more abstract setting.

Definition 1.1. Assume that S is a set. A *partial order* on S is a subset $P \subset S\times S$ with the following properties:

- *Reflexivity:* $(s,s) \in P$ for all $s \in S$.
- *Anti-symmetry:* If $(s,t) \in P$ and $(t,s) \in P$, then $s = t$.
- *Transitivity:* If $(s,t) \in P$ and $(t,u) \in P$, then $(s,u) \in P$.

A pair (S,P), where P is a partial order on S is called a *partially ordered set* (or *poset*).

Typically, we indicate a partial order P on S by the symbol $\preceq$, saying that s is smaller than t or

$$s \preceq t \qquad \Longleftrightarrow \qquad (s,t) \in P.$$

Then the conditions for a partial order can be restated in the following, more familiar form:

- *Reflexivity:* $s \preceq s$ for all $s \in S$.
- *Anti-symmetry:* If $s \preceq t$ and $t \preceq s$, then $s = t$.
- *Transitivity:* If $s \preceq t$ and $t \preceq u$, then $s \preceq u$.

Given a partially ordered set $(S,\preceq)$ and elements $s,\ t \in S$, we say that s is *strictly* smaller than t, denoted $s \prec t$, if $s \preceq t$ and $s \ne t$.

Remark 1.2. A partially ordered set $(S,\preceq)$ is called totally ordered, if all elements are comparable, that is, for each $s \ne t \in S$ we either have $s \prec t$ or $t \prec s$. For instance, the set $\mathbb{R}$ of real numbers is totally ordered.

If s and t are elements of a totally ordered set and s is not smaller than t, in symbols $s \not\preceq t$, then we can immediately conclude that $t \preceq s$. For a general partially ordered set, however, this is not the case.

Date: 4th April 2023.

Example 1.3. On $\mathbb{R}^n$ we can consider the *standard*, or *componentwise*, partial order $x \preceq y$, if and only if $x_i \le y_i$ for all $1 \le i \le n$. This is indeed a partial order, as the following considerations show:

- If $x \in \mathbb{R}^n$, then, trivially, $x_i \le x_i$ for each i, which implies that $x \preceq x$.
- If $x \preceq y$ and $y \preceq x$, then $x_i \le y_i$ and $y_i \le x_i$ for each i, which implies that $x_i = y_i$ for each i. Thus $x = y$.
- If $x \preceq y$ and $y \preceq z$, then we have for each i that $x_i \le y_i$ and $y_i \le z_i$, which implies that $x_i \le z_i$ for each i. Thus we also have that $x \preceq z$.

In dimensions $n \ge 2$, this is not a total order. For instance, the vectors $e_1 = (1, 0, 0, \ldots)$ and $e_2 = (0, 1, 0, \ldots)$ are not comparable; neither of the inequalities $e_1 \preceq e_2$ or $e_2 \preceq e_1$ holds.

When nothing else is specified, we will always consider the space $\mathbb{R}^n$ with this partial order.

Example 1.4. Let E be any set and denote by $\mathbb{P}(E)$ the *power set* of E, that is, the set of all sub-sets of E. Then the inclusion $\subseteq$ defines a partial order on $\mathbb{P}(E)$:

- For every $U \in \mathbb{P}(E)$ we have $U \subseteq U$.
- If U, $V \in \mathbb{P}(E)$ satisfy $U \subseteq V$ and $V \subseteq U$, then $U = V$.
- IF U, V, $W \in \mathbb{P}(E)$ satisfy $U \subseteq V$ and $V \subseteq W$, then also $U \subseteq W$.

Unless E is empty or contains only one element, the relation $\subseteq$ does not define a total order: If $u \ne v \in E$, then neither of the inclusions $\{u\} \subseteq \{v\}$ or $\{v\} \subseteq \{u\}$ holds, and thus the sets $\{u\}$ and $\{v\}$ are not comparable.

Example 1.5. Let $d \ge 2$ and $S := \bigl\{A \in \mathbb{R}^{d\times d} : A = A^T\bigr\}$ the set of (real) symmetric matrices of dimension $d \times d$. The relation $A \preceq B$ if and only if $B - A$ is positive semi-definite defines a partial order on S.

If the space S in addition is a vector space, we normally would expect a partial order to be compatible with the vector space structure. This is captured in the following definition.

Definition 1.6. Assume that U is a real vector space and that $\preceq$ is a partial order on U. We say that $(U, \preceq)$ is an *ordered vector space*, if (in addition to the requirements for a partial order) the following conditions hold:

- Assume that u, $v \in U$ satisfy $u \preceq v$ and that $w \in U$. Then also $u + w \preceq v + w$.
- Assume that u, $v \in U$ satisfy $u \preceq v$ and that $\lambda \in \mathbb{R}$ with $\lambda \ge 0$. Then also $\lambda u \preceq \lambda v$.

It turns out that there is a close connection between partial orders on ordered vector spaces and convex cones: In an ordered vector space $(U, \preceq)$ we can consider the set of non-negative elements $C := \bigl\{u \in U : 0 \preceq u\bigr\}$, which turns out to be a convex cone if the order structure is compatible with the vector space structure on U. Conversely, given a convex cone we can (under a small additional condition) always interpret it is the non-negative cone for a partial order on U. The precise relationship between partial orders and convex cones is given in the following result.

Lemma 1.7. *Assume that U is an ordered vector space and denote*

$$C := \bigl\{u \in U : 0 \preceq u\bigr\}.$$

Then C is a convex cone, that is, C has the following properties:

- *If u, $v \in C$, then also $\lambda u + (1-\lambda)v \in C$ for all $\lambda \in \mathbb{R}$ with $0 < \lambda < 1$.*
- *If $u \in C$, and if $\lambda \in \mathbb{R}$ satisfies $\lambda \ge 0$, then $\lambda u \in C$.*

In addition, C satisfies $C \cap (-C) = \{0\}$. Moreover, for u, $v \in U$ we have that $u \preceq v$ if and only if $v - u \in C$.

Conversely, if $C \subset U$ is a non-empty convex cone satisfying $C \cap (-C) = \{0\}$, then the relation $u \preceq v$ if and only if $v - u \in C$ makes $(U, \preceq)$ an ordered vector space.

Proof. Assume that $u, v \in C$ and that $\lambda \in \mathbb{R}$ with $0 < \lambda < 1$. Then $0 \preceq u$ and $0 \preceq v$. Because $\lambda > 0$ and $1 - \lambda > 0$, we therefore also have that $0 \preceq \lambda u$ and $0 \preceq (1 - \lambda)v$. As a consequence, we have that $(1-\lambda)v = 0+(1-\lambda)v \preceq \lambda u+(1-\lambda)v$, and therefore $0 \preceq (1 - \lambda)v \preceq \lambda u + (1 - \lambda)v$.

Next assume that $u \in C$ and that $\lambda \in \mathbb{R}$ satisfies $\lambda \geq 0$. Then $0 \preceq u$ and thus also $0 = \lambda\, 0 \preceq \lambda u$, which shows that $\lambda u \in C$.

Next, assume that $u, v \in U$. Then the inequality $u \preceq v$ implies that $0 = u - u \preceq v - u$ and thus $v - u \in C$. Conversely, if $v - u \in C$, then $0 \preceq v - u$ and thus $u = 0 + u \preceq v - u + u = v$.

Finally, let $v \in C \cap \{-C\}$ or, equivalently, $v \in C$ and $-v \in C$. Then $0 \preceq v$ and $0 \preceq -v$, the latter implying that $v \preceq 0$. Therefore $v = 0$ and thus $C \cap \{-C\} = \{0\}$.

Now assume that $C \subset U$ is a non-empty convex cone and the the relation $\preceq$ is defined by $u \preceq v \iff v - u \in C$. Since by assumption $0 \in C$, we have that $u - u \in C$ for all $u \in U$, which shows that $u \preceq u$.

Next, assume that $u \preceq v$ and $v \preceq u$. Then $v - u \in C$ and $u - v \in C$ and thus $v - u \in C \cap (-C) = \{0\}$. In other words, $u = v$.

Finally, assume that $u \preceq v$ and $v \preceq w$, that is $v - u \in C$ and $w - v \in C$. Since C is a convex cone, we also have that

$$w - u = 2\Big(\frac{1}{2}(w - v) + \frac{1}{2}(v - u)\Big) \in C,$$

showing that $u \preceq w$.

We have thus shown that $\preceq$ defines a partial order on U. It remains to show that this order is compatible with the vector space structure. To that end, assume that $u \preceq v$ and $w \in U$. Then $(w+v)-(w+u) = v-u \in C$ since C is a convex cone , and thus $u+w \preceq v+w$. Moreover, if $u \preceq v$ and $\lambda \geq 0$, then $\lambda v - \lambda u = \lambda(v-u) \in C$ since C is a cone, and thus $\lambda u \preceq \lambda v$. $\square$

Example 1.8. The standard order on $\mathbb{R}^n$ is given by the cone $C := \mathbb{R}^n_{\geq 0}$ (see Figure 1 for an illustration).

The order on the space of symmetric matrices discussed in Example 1.5 is defined by the cone of positive semi-definite matrices.

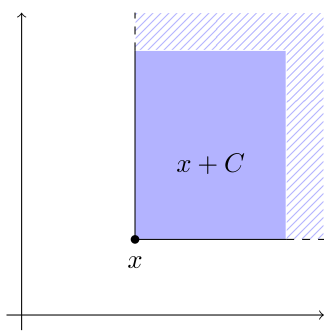

FIGURE 1. Illustration of the order relation in $\mathbb{R}^2$. The blue area indicates all points $y \in \mathbb{R}^2$ that satisfy $x \preceq y$. This is the same as the set $x + C$, where C is the non-negative cone $C = \{y \in \mathbb{R}^2 : 0 \preceq y\}$.

2. Pareto-optimality

Definition 2.1. Assume that X is some set, $(Y, \preceq)$ is a partially ordered set and that $f\colon X \to Y$ is some function. A point $x^* \in X$ is called *efficient* or *Pareto-optimal* for f, if there does not exist any $x \in X$ such that $f(x) \prec f(x^*)$.

The *Pareto-front* is the set of all Pareto-optimal points for f.

We consider now the case where $f\colon \Omega \subset \mathbb{R}^d \to \mathbb{R}^n$ is a vector function, and we consider the problem

$$\min_{x\in\Omega} f(x), \tag{1}$$

where we assume the standard order on $\mathbb{R}^n$. In this case, the definition of Pareto-optimality can be reformulated as the following:

Definition 2.2. Assume that $f\colon \Omega \subset \mathbb{R}^d \to \mathbb{R}^n$, where $\mathbb{R}^n$ is equipped with the standard order. A point $x^* \in \Omega$ is called *efficient* or *Pareto-optimal* for the problem (1), if there does not exist any $x \in \Omega$ such that $f_i(x) \le f_i(x^*)$ for all $1 \le i \le n$, and $f_k(x) < f_k(x^*)$ for some $1 \le k \le n$.

Put differently, it is not possible to improve the value of one of the component functions f_i without making another one worse. In other words, x^* is Pareto-optimal if and only it solves for each $1 \le k \le n$ the optimisation problem

$$\min_x f_k(x) \qquad \text{s.t. } f_i(x) \le f_i(x^*) \text{ for all } i \ne k.$$

Remark 2.3. Given a partially ordered set $(S, \preceq)$ and a subset $U \subset S$, we say that $s \in U$ is a minimal element of U if there does not exist any $t \in U$ with $t \prec s$. Using this notion of minimality, we can also say that the Pareto-optimal solutions of the problem $\min_{x\in\Omega} f(x)$ for a function $f\colon \Omega \to S$ are precisely those points $x^* \in \Omega$ for which $f(x^*)$ is a minimal element of the image $f(\Omega) = \big\{ f(x) \in S : x \in \Omega \big\}$ of Ω.

In the literature on multicriteria optimisation, this image of the Pareto-front is often identified with the Pareto-front itself. In this lecture, however, we will not follow this convention and consider the Pareto-front only as subset of the definition space of the function f, not of its image.

Example 2.4. Consider the functions f_1, $f_2\colon [0,1]^2 \to \mathbb{R}$,

$$\begin{aligned} f_1(x,y) &= x + y, \\ f_2(x,y) &= 2 - x^2 - y^2. \end{aligned}$$

Since we can write

$$f_2(x,y) = 2 - x^2 - y^2 = 2 - \frac{1}{2}(x+y)^2 - \frac{1}{2}(x-y)^2,$$

we see that for a fixed value of $f_1(x,y) = x + y$, the function $f_2(x,y)$ is minimal when the difference $|x - y|$ is the largest. For $0 \le x + y \le 1$, this happens precisely when $x = 0$ or $y = 0$; for $1 \le x + y \le 2$, this happens precisely when $x = 1$ or $y = 1$. Put differently, the Pareto-front of this problem consists of the points $(0,y)$, $(x,0)$, $(1,y)$, and $(x,1)$, with $0 \le x, y \le 1$.

Alternatively, we can consider the image

$$f([0,1]^2) = \big\{ (x+y, 2 - x^2 - y^2) : 0 \le x, y \le 1 \big\} \subset \mathbb{R}^2$$

of the function $f = (f_1, f_2)$, which is depicted in Figure 2. The image of the Pareto-front consists of the minimal elements of $f([0,1]^2)$, that is, the "lower left boundary" of the set $f([0,1]^2)$.

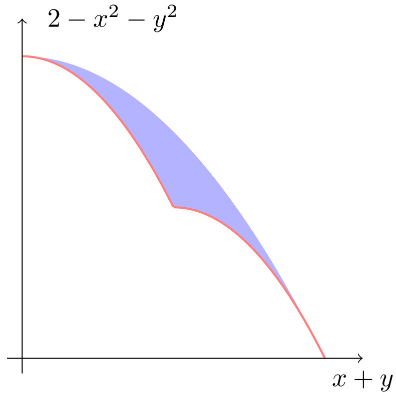

FIGURE 2. Illustration of the Pareto-front for the problem $\min_{0 \le x,y \le 1} f(x,y)$ with $f_1(x,y) = x+y$ and $f_2(x,y) = 2 - x^2 - y^2$. The blue area is the image of f, that is, the set $\{f(x,y) : 0 \le x, y \le 1\}$, and the red line shows the minimal elements in the image of f.

3. The Weighted Sum Method

We consider again the problem (1) with $f\colon \mathbb{R}^d \to \mathbb{R}^n$ and the standard order on $\mathbb{R}^n$. Our goal is to devise a method for actually computing Pareto-optimal solutions of multi-criteria optimisation problems. The approach we are chosing is that of *scalarisation*, where one transforms the multi-criteria problem into a *family* of standard optimisation problems, the solutions of which should coincide with Pareto-optimal solutions of (1). More specifically, we consider the idea of *weighted sum scalarisation*, where these optimisation problems are just a weighted sum of the component functions f_i. We will show in the following first that the solutions of these weighted sums are indeed Pareto-optimal. Then we will turn to the question whether it is possible to obtain the whole Pareto-front with this approach, or if we are missing some Pareto-optimal solutions.

Theorem 3.1. *Let $\lambda_i > 0$, $1 \le i \le n$, and let x_λ^* be a solution of the optimisation problem*

$$\min_{x \in \Omega} \sum_{i=1}^{n} \lambda_i f_i(x). \tag{2}$$

Then x_λ^* is a Pareto-optimal solution of (1).

Proof. Assume to the contrary that x_λ^* is not Pareto-optimal. Then there exists $y \in \Omega$ such that $f(y) \prec f(x_\lambda^*)$. That is, $f_i(y) \le f_i(x_\lambda^*)$ for all $1 \le i \le n$ and there exists some k such that $f_k(y) < f_k(x_\lambda^*)$. As a consequence, since $\lambda_i > 0$ for all $1 \le i \le n$,

$$\sum_{i=1}^{n} \lambda_i f_i(y) = \lambda_k f_k(y) + \sum_{i \ne k} \lambda_i f_i(x_\lambda^*) < \lambda_k f_k(x_k^*) + \sum_{i \ne k} \lambda_i f_i(x_\lambda^*) = \sum_{i=1}^{n} \lambda_i f_i(x_\lambda^*), \tag{3}$$

which shows that x_λ^* is no solution of (2). This shows that, in fact, x_λ^* must be Pareto-optimal. □

This result shows that we can construct at least part of the Pareto-front by taking all possible combinations of weights $\lambda_i > 0$ and computing the solutions x_λ^* of the corresponding weighted sum problem (2). In practice, of course, one would rather take a sufficiently fine discretisation of the set of weights. Also, it is sufficient to take weights λ_i such that $\sum_i \lambda_i = 1$, as multiplying all weights with the same (positive) number does not change the solution.

Remark 3.2. In Theorem 3.1, we have assumed that all the weights λ_i are strictly larger than zero. This is necessary in order to conclude that the inequality in (3) is strict: In the case $\lambda_k = 0$, we would only have that the left hand side is smaller or equal to the right hand side, and therefore would not obtain a contradiction to the assumption that x_λ^* solves (2).

As a (trivial) example where the positivity of the parameters λ_i is necessary, consider the case of two functions f_1, f_2 with $f_1(x) = 0$ for all x. Then the Pareto-optimal solutions of $\min_x(f_1(x), f_2(x))$ are precisely the (standard) minimisers of the function f_2, because f_1 does not add any restrictions. These same points are obtained by solving (2) for any weights $\lambda_1 \geq 0$ and $\lambda_2 > 0$. With weights $\lambda_1 > 0$ and $\lambda_2 = 0$, however, we simply obtain the function $\lambda_1 f_1(x) + 0 = 0$, which has as minimisers the whole set Ω. If we were to allow a weight to be equal to 0, we thus would obtain solutions of (2) that are not Pareto-optimal.

The next natural question is, whether the weighted sum method allows us to obtain *all* Pareto-optimal solution of the multi-criteria problem. Unfortunately, the answer to this question is in general *no*, as the following example shows.

Example 3.3. We consider again the functions f_1, $f_2 \colon [0,1]^2 \to \mathbb{R}$ from Example 2.4, that is,

$$\begin{aligned} f_1(x,y) &= x + y, \\ f_2(x,y) &= 2 - x^2 - y^2. \end{aligned}$$

Let now λ_1, $\lambda_2 > 0$, and consider the weighted sum problem

$$\min_{x,\, y \in [0,1]} \bigl(\lambda_1 f_1(x,y) + \lambda_2 f_2(x,y)\bigr). \tag{4}$$

Since f_1 is affine, f_2 is strictly concave, and $\lambda_2 > 0$, it follows that the objective function of this problem is strictly concave as well. Thus the only potential solutions of the problem (4) are the vertices of the polyhedron $[0,1]^2$, that is, the points $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$.

In fact, if we compute the function values of the weighted sum $\lambda_1 f_1 + \lambda_2 f_2$, we see that the solution of (4) is equal to $(1,1)$ for $\lambda_2 > \lambda_1$ and equal to $(0,0)$ for $\lambda_1 < \lambda_2$, and for $\lambda_1 = \lambda_2 > 0$ all the points $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ are global solutions of (4). However, it is not possible to obtain any other Pareto-optimum by minimising a function of the form $\lambda_1 f_1 + \lambda_2 f_2$ with λ_1, $\lambda_2 > 0$.

We will now show that the situation is better in the case where all the involved functions are convex. For that, we will need a result that is related to Farkas' lemma, which we have seen earlier when deriving the KKT conditions.

Theorem 3.4 (Gordan's Theorem)**.** *Assume that $g_i \in \mathbb{R}^d$, $1 \leq i \leq n$. Then precisely one of the two following alternatives is true:*

(1) There exist $\lambda_i \geq 0$, $1 \leq i \leq n$, with $\sum_i \lambda_i = 1$ such that

$$\sum_{i=1}^n \lambda_i g_i = 0. \tag{5}$$

(2) There exists $p \in \mathbb{R}^d$ such that

$$\langle g_i, p \rangle < 0 \qquad \textit{for all } 1 \leq i \leq n.$$

Proof. We follow the proof in [1, Thm. 2.2.1].

First assume that (1) holds and let $\lambda_i \ge 0$, $1 \le i \le n$, with $\sum_i \lambda_i = 1$ be such that $\sum_i \lambda_i g_i = 0$. Then we have for every $p \in \mathbb{R}^d$ that

$$\sum_{i=1}^{n} \lambda_i \langle g_i, p\rangle = 0,$$

and thus it is impossible that $\langle g_i, p\rangle < 0$ for all i. In other words, if (1) holds, then (2) cannot hold at the same time.

Now let us assume that (2) does not hold. Our goal is to show that, in this case, (1) necessarily holds. To that end, we define the function $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$f(p) = \ln\Big(\sum_{i=1}^{n} \exp\big(\langle g_i, p\rangle\big)\Big).$$

Since (2) does not hold, there exists for all $p \in \mathbb{R}^n$ some k such that $\langle g_k, p\rangle \ge 0$. Then, since ln is monotonically increasing and $\exp(t) > 0$ for all t,

$$f(p) = \ln\Big(\sum_{i=1}^{n} \exp\big(\langle g_i, p\rangle\big)\Big) \ge \ln\big(\exp\big(\langle g_k, p\rangle\big)\big) \ge 0.$$

For $\varepsilon > 0$ we now define the function $f^{(\varepsilon)}\colon \mathbb{R}^d \to \mathbb{R}$,

$$f^{(\varepsilon)}(p) := f(p) + \frac{\varepsilon}{2}\|p\|_2^2,$$

and consider the problem

$$\min_{p\in\mathbb{R}^d} f^{(\varepsilon)}(p).$$

Since $f(p) \ge 0$ for all p, it follows that $f^{(\varepsilon)}(p) \ge \varepsilon\|p\|_2^2/2$ for all p, and thus $f^{(\varepsilon)}$ is coercive. Therefore, this problem attains a minimiser $p^{(\varepsilon)} \in \mathbb{R}^d$. Moreover we can estimate

$$\frac{\varepsilon}{2}\|p^{(\varepsilon)}\|_2^2 \le f^{(\varepsilon)}(p^{(\varepsilon)}) \le f^{(\varepsilon)}(0) = f(0) = \ln(n)$$

for all $\varepsilon > 0$, which in particular implies that

$$\lim_{\varepsilon\to 0} \varepsilon\|p^{(\varepsilon)}\|_2 = \lim_{\varepsilon\to 0} \sqrt{\varepsilon}\sqrt{\varepsilon\|p^{(\varepsilon)}\|_2^2} \le \liminf_{\varepsilon\to 0} \sqrt{\varepsilon}\sqrt{\ln(n)} = 0.$$

Moreover, we have $\nabla f(p^{(\varepsilon)}) = 0$, that is,

$$0 = \nabla f^{(\varepsilon)}(p^{(\varepsilon)}) = \sum_{i=1}^{n} \frac{\exp\big(\langle g_i, p^{(\varepsilon)}\rangle\big)}{\sum_{j=1}^{n} \exp\big(\langle g_j, p^{(\varepsilon)}\rangle\big)} g_i + \varepsilon p^{(\varepsilon)}$$

Now define

$$\lambda_i^{(\varepsilon)} := \frac{\exp\big(\langle g_i, p^{(\varepsilon)}\rangle\big)}{\sum_{j=1}^{n} \exp\big(\langle g_j, p^{(\varepsilon)}\rangle\big)}.$$

Then $\lambda_i^{(\varepsilon)} \ge 0$ for all i and ε, and $\sum_i \lambda_i^{(\varepsilon)} = 1$ for all $\varepsilon > 0$. As a consequence, there exist a sequence $\varepsilon_k \to 0$ and $\lambda_i \ge 0$ with $\sum_i \lambda_i = 1$ such that

$$\lambda_i^{(\varepsilon_k)} \to \lambda_i$$

for all i. Moreover,

$$\Big\|\sum_i \lambda_i g_i\Big\|_2 = \lim_{k\to\infty} \Big\|\sum_i \lambda_i^{(\varepsilon_k)} g_i\Big\|_2 = \lim_{k\to\infty} \big\|\varepsilon_k p^{(\varepsilon_k)}\big\|_2 = 0.$$

Thus (5) holds.

Altogether, we have now shown that (1) holds, if and only if (2) does not hold, which is precisely the claim of the theorem. □

Now we return to our investigation of the weighted sum method for the computation of the Pareto-front, but we restrict ourselves to the case where all of the functions f_i are convex and differentiable. Moreover, for simplicity, we consider only the unrestricted case $\Omega = \mathbb{R}^d$. It is possible, though, to extend these results to the case where the functions f_i are not necessarily differentiable (but still convex) and the set Ω is convex, although the proof for that general setting is somewhat more technical.

The first result states that every Pareto-optimal solution of the multi-criteria problem can be obtained by minimising some weighted sum problem, provided that the functions f_i are convex and we are allowed to set some of the weights to 0.

Theorem 3.5. *Assume that the functions $f_i \colon \mathbb{R}^d \to \mathbb{R}$, $1 \le i \le n$, are convex and differentiable, and assume that x^* is a Pareto-optimal solution of the problem*

$$\min_{x \in \mathbb{R}^d} f(x). \tag{6}$$

Then there exist weights $\lambda_i \ge 0$, $1 \le i \le n$, with $\sum_i \lambda_i = 1$ such that x^* solves

$$\min_{x \in \mathbb{R}^d} \sum_{i=1}^n \lambda_i f_i(x). \tag{7}$$

Proof. Since x^* is a Pareto-optimal solution of (6), there exists no $\hat{x} \in \mathbb{R}^d$ such that $f_i(\hat{x}) \le f_i(x^*)$ for all i and $f_k(\hat{x}) < f_k(x^*)$ for at least one k. *A forteriori*, there exists no $\hat{x} \in \mathbb{R}^d$ such that $f_i(\hat{x}) < f_i(x^*)$ for *all* $1 \le i \le n$.

Next we will show that this implies that there does not exist any $p \in \mathbb{R}^d$ such that $\langle \nabla f_i(x^*), p \rangle < 0$ for all $1 \le i \le n$. Indeed, were such a p to exist, it would be a descent direction for all functions f_i at x^*, and thus we could find for each i some $\varepsilon_i > 0$ such that $f_i(x^* + tp) < f_i(x^*)$ for all $0 < t < \varepsilon_i$. Setting $t := \frac{1}{2} \min_i \varepsilon_i$ and $\hat{x} = x^* + tp$, we would thus obtain that $f_i(\hat{x}) = f_i(x^* + tp) < f_i(x^*)$ for all i, which would contradict our previous statement.

Since there does not exist any p such that $\langle \nabla f_i(x^*), p \rangle < 0$ for all $1 \le i \le n$, the first alternative in Gordan's Theorem has to be true. That is, there exist $\lambda_i \ge 0$, $1 \le i \le n$, with $\sum_i \lambda_i = 1$ such that

$$\sum_i \lambda_i \nabla f_i(x^*) = 0.$$

This, however, implies that x^* is a critical point of the function

$$x \mapsto \sum_i \lambda_i f_i(x).$$

Since this function is convex, every critical point is already a global solution. This shows that x^* solves (7) for these weights. ☐

Finally, if we replace convexity with *strict* convexity of the functions f_i, we obtain a one-to-one correspondance between the Pareto-optimal solutions of the multi-criteria problem and the solutions of the weighted sum problem.

Theorem 3.6. *Assume that the functions $f_i \colon \mathbb{R}^d \to \mathbb{R}$ are strictly convex and differentiable. Then $x^* \in \mathbb{R}^d$ is a Pareto-optimal solution of*

$$\min_{x \in \mathbb{R}^d} f(x)$$

if and only if there exist weights $\lambda_i \ge 0$, $1 \le i \le n$, with $\sum_i \lambda_i = 1$ such that x^* solves

$$\min_{x \in \mathbb{R}^d} \sum_{i=1}^n \lambda_i f_i(x). \tag{8}$$

Proof. The "only if" part has already been shown in Theorem 3.5. Assume therefore that x^* solves (8) for some $\lambda_i \geq 0$ with $\sum_i \lambda_i = 1$ and that $x \neq x^*$. Because of the strict convexity of the functions f_i, it follows that $\sum_i \lambda_i f_i$ is strictly convex as well and thus the minimiser of this function is unique. Therefore

$$\sum_{i=1}^{n} \lambda_i f_i(x^*) < \sum_{i=1}^{n} \lambda_i f_i(x).$$

Thus there exists at least one index i such that $\lambda_i f_i(x^*) < \lambda_i f_i(x)$, which in turn implies that $f_i(x^*) < f_i(x)$. Thus we cannot have that $x \prec x^*$. Since this holds for every $x \neq x^*$, it follows that x^* is Pareto-optimal. □

To summarise, we have obtained the following results:

- If x^* is a solution of some weighted sum problem where all the weights are positive, then it is a Pareto-optimum.
- If the functions f_i are convex and x^* is a Pareto-optimum then there exist non-negative parameters λ_i, with at least one of them being strictly positive, such that x^* solves the corresponding weighted sum problem.
- If the functions f_i are strictly convex, then x^* is a Pareto-optimum if and only if there exist non-negative parameters λ_i, with at least one of them being strictly positive, such that x^* solves the corresponding weighted sum problem.

References

[1] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization*, volume 3 of *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC*. Springer-Verlag, New York, 2000. Theory and examples.

[2] Matthias Ehrgott. *Multicriteria optimization*. Springer-Verlag, Berlin, second edition, 2005.

[3] Johannes Jahn. *Vector optimization*. Springer-Verlag, Berlin, 2004. Theory, applications, and extensions.

Trondheim, Norway
Email address: markus.grasmair@gmail.com