Here we consider multiobjective optimization problem $f_1(x) \rightarrow \min, \ldots, f_m(x) \rightarrow \min, x \in \mathbb{R}^d$. We will later focus on the bicriteria case $m = 2$, though some methods that will be discussed are more general.

In previous work ([1], [2], and [3]) the hypervolume gradient has been suggested as a direction for steering a set of points in the direction of sets with larger or maximal hypervolume indicator values. The set of points was interpreted as a vector of dimension $\mu \cdot d$, where $\mu$ is the number of points in the set and $d$ the dimension of any of these points. A search point, representing a set, is thus a vector $p \in (\mathbb{R}^d)^{\mu}$. We will refer to its components by $p^{(1)}, \ldots, p^{(\mu)}$, i.e. the concatenation $p^{(1)}, \ldots, p^{(\mu)}$ equals $p$. Given a sufficiently large reference point $r \in \mathbb{R}^m$ the hypervolume indicator is defined by:

$$HI(p) = \lambda(\cup_{i=1}^{\mu}[f(p^{(i)}), r])$$

The hypervolume gradient $\nabla HI(p)$ is defined for non-degenerate cases (see [1]). Practically, the degenerate cases do not occur, i.e. they form a measure zero subset of $\mathbb{R}^{d\mu}$. However, two other problem occurs when using the gradient direction as a search direction. In order to discuss these problems we will introduce subgradients

$$\nabla^{(i)} HI(p) := \left(\frac{\partial HI}{\partial p^{(i)}_1}(p), \ldots, \frac{\partial HI}{\partial p^{(i)}_d}(p)\right)^\top, i = 1, \ldots, \mu.$$

A component $p^{(i)}, i = 1, \ldots, \mu$ is called dominated w.r.t. $p$, if and only if there exists $p^{(j)}, j \in \{1, \ldots, \mu\}$ with $p^{(j)}$ (Pareto) dominates $p^{(i)}$. In this case the $i$-th subgradient is zero, i.e. $\nabla^{(i)} HI(p) = (0, \ldots, 0)^\top \in \mathbb{R}^d$.

In the hypervolume gradient ascent method a zero subgradient leads to the unwanted effect that part of the points in the set – the dominated points – stay in the same position [3].

The goal of this work is to propose and test methods that replace zero subgradients by vectors that lead in the direction of the Pareto front approximation. Once the dominated components get non-dominated (relative to the other components) they can contribute to the hypervolume indicator and their subgradients will be non-zero again.

Besides, a known problem in hypervolume gradient ascent is that the length of the subgradients can differ widely, which leads to a slow convergence in some parts of the objective space. This phenomenon was analysed in [3] and termed 'creepiness'. To counteract creepiness, it was proposed to normalize all subgradients before applying them in the gradient based search [5]. This strategy is also adopted in this work.

Five methods are proposed to replace subgradients. These are:

M1 Use the gradient of $f_1(p^{(i)}) + f_2(p^{(i)})$. In regular cases this method will lead the dominated points to the Pareto front albeit the methods tends to steer the points towards the same point on the Pareto front (tangent of $-45^\circ$).

M2 To avoid diversity loss as in M1, in M2 the proposed method is to use instead of the gradient of a randomly weighted aggregated sum $w_1f_1(p^{(i)}) + (1-w_1)f_2(p^{(i)})$, with $w_1 \sim U(0, 1)$ being a uniformly distributed random number.
M3 For this method a direction that points into the dominance cone of a single point was proposed in López as

\[ \nabla f_1(x)/||\nabla f_1(x)|| + \nabla f_2(x)/||\nabla f_2(x)||. \]

M4 This methods seeks to find a point that fills the nearest gap of the current Parteto front approximation with respect to the dominated point. For this it computes the secant of the two neighboring points of the closest gap and the angle of this secant. The weights of the linear aggregation \( w_1 f_1 + w_2 f_2 \) are chosen in such a way that a tangential point on the Pareto front with this angle is approximated. In other words, if the secant has the slope \( m \), the weights are chosen such that \( w_1/w_2 = m \).

M5 This method seeks to get closer to the midpoint of the secant described in M4. This midpoint we will term \( c \) and the subgradient is replaced by the subgradient of the squared Euclidean distance to this point. That is for a dominated component with index \( i \) we replace the subgradient by:

\[ \nabla \left( (f_1(p^{(i)}) - c_1)^2 + (f_2(p^{(i)}) - c_2)^2 \right) \]

The five methods are tested on both convex and concave Pareto fronts and their convergence behavior will be compared. The study will reveal interesting patterns in the convergence behavior and provide explanations for these.

References


