

# 1 Motivations

- Genetic Drift – [Wikipedia](#)
- Speciation – [Wikipedia](#)
- Punctuated Equilibria – [Wikipedia](#)

# 2 Implications for Evolutionary Optimisation

## Implicit approaches

Impose an equivalent of geographical separation. Impose an equivalent of speciation

## Explicit approaches

Make similar individuals compete for resources (fitness). Make similar individuals compete with each other for survival.

### 2.1 Explicit 1: Fitness Sharing

Restricts the number of individuals within a given niche by “sharing” their fitness, so as to allocate individuals to niches in proportion to the niche fitness. Need to set the size of the niche  $\sigma_{\text{share}}$ . In either genotype or phenotype space run EA as normal but after each gen set the shared fitness of each candidates with the equation of

$$f'(i) = \frac{f(i)}{\sum_{j=1}^{\mu} sh(d(i, j))} \quad sh(d) = \begin{cases} 1 - \frac{d}{\sigma}, & d < \sigma, \\ 0, & \text{otherwise.} \end{cases}$$

### 2.2 Explicit 2: Crowding

Attempts to distribute individuals **evenly** amongst niches. Relies on the assumption that offspring will tend to be close to parents. Uses a distance metric in phenotype/genotype space. Randomly shuffle and pair parents, then produce two offspring. Two parent/offspring tournaments - pair so that

$$d(p_1, o_1) + d(p_2, o_2) < d(p_1, o_2) + d(p_2, o_1)$$

# 3 Prerequisites

## 3.1 Ordered Set

### Partial Ordered Set

Formally, a partial order is a **homogeneous binary relation** that is **reflexive**, **antisymmetric**, and **transitive**. A **partially ordered set** (*poset* for short) is an *ordered pair*  $P = (X, \leq)$  consisting of a set  $X$  (called the *ground set* of  $P$ ) and a partial order  $\leq$  on  $X$ . When the meaning is clear from context and there is no ambiguity about the partial order, the set  $X$  itself is sometimes called a poset.

A **reflexive**, **weak**, or **non-strict partial order**, commonly referred to simply as a **partial order**, is a **homogeneous relation**  $\leq$  on a set  $P$  that is **reflexive**, **antisymmetric**, and **transitive**. That is, for all  $a, b, c \in P$ , it must satisfy:

1. **Reflexivity:**  $a \leq a$ , i.e. every element is related to itself.

2. **Antisymmetry:** if  $a \leq b$  and  $b \leq a$  then  $a = b$ , i.e. no two distinct elements precede each other.
3. **Transitivity:** if  $a \leq b$  and  $b \leq c$  then  $a \leq c$ .

A non-strict partial order is also known as an **antisymmetric preorder**.

### Total Order

In mathematics, a total order or linear order is a partial order *in which any two elements are comparable*. That is, a total order is a binary relation  $\leq$  on some set  $X$ , which satisfies the following for all  $a, b$  and  $c$  in  $X$ :

1.  $a \leq a$  (*reflexive*).
2. If  $a \leq b$  and  $b \leq c$  then  $a \leq c$  (*transitive*).
3. If  $a \leq b$  and  $b \leq a$  then  $a = b$  (*antisymmetric*).
4.  $a \leq b$  or  $b \leq a$  (*strongly connected*, formerly called *totality*).

### 3.2 Cartesian Product

In **mathematics**, specifically **set theory**, the **Cartesian product** of two sets  $A$  and  $B$ , denoted  $A \times B$ , is the set of all **ordered pairs**  $(a, b)$  where  $a$  is an element of  $A$  and  $b$  is an element of  $B$ . In terms of **set-builder notation**, that is

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

#### Orders on the Cartesian product of totally ordered sets

There are several ways to take two totally ordered sets and extend to an order on the **Cartesian product**, though the resulting order may only be **partial**. Here are three of these possible orders, listed such that each order is stronger than the next:

- **Lexicographical order:**  $(a, b) \leq (c, d)$  if and only if  $a < c$  or  $(a = c \text{ and } b \leq d)$ . This is a total order.
- $(a, b) \leq (c, d)$  if and only if  $a \leq c$  and  $b \leq d$  (the **product order**). This is a partial order.
- $(a, b) \leq (c, d)$  if and only if  $(a < c \text{ and } b < d)$  or  $(a = c \text{ and } b = d)$  (the reflexive closure of the **direct product** of the corresponding strict total orders). This is also a partial order.

#### Relations as Subsets of Cartesian Products [Mathematics Exchange](#)

The set of ordered pairs drawn from the Cartesian product can be defined as

$$R = \{(x, y) \in A \times B \mid xRy\}$$

Thus we can say

$$R \subseteq A \times B$$

## 4 Multiobjective Optimisation

Many problems have more than one goal function thus there is no single optimal function value. There is a general assumption that MO is as hard as Single-objective optimisation. We have the decision space of  $S^n$  with  $s \in S^n$  and with our objective function  $F(f_1(s), f_2(s), \dots, f_m(s))$  will yield the objective space of  $\mathbb{R}^m$ . Therefore, for a general MOP we have that

$$F : S^n \rightarrow \mathbb{R}^m, \quad \text{where } n < m \vee n > m \vee n = m$$

We say  $x \in \mathbb{R}^m$  dominates  $y \in \mathbb{R}^m$  if it is at least as good on all criteria and better on at least one.

## Dominance in the Objective Space

Assuming maximisation

$x$  weakly dominates  $y$  ( $x \succeq y$ ) iff  $x_i \geq y_i \forall i \in \{1, \dots, m\}$   
 $x$  dominates  $y$  ( $x \succ y$ ) iff  $x \succeq y$  and  $x \neq y$

This can be translated to the search space as

$$\begin{aligned} s \succeq s' &\iff F(s) \succeq F(s') \\ s \succ s' &\iff F(s) \succ F(s') \end{aligned}$$

## Pareto Front – Wikipedia

The Pareto Front consists of all the non-dominated objective vectors. Let  $\pi \subseteq \mathbb{R}^m$  be the set of feasible objective vectors such that

$$\pi = \{\pi \in \mathbb{R}^m : \pi = F(s) \text{ where } s \in S^n\}$$

For any objective vectors  $\pi', \pi'' \in \pi$ , let  $P(\pi)$  be the Pareto set thus

$$P(\pi) = \{\pi' : \{\pi'' \succ \pi'\} = \emptyset\}$$

## Pareto Optimal

A solution  $x^* \in X$  is then called *Pareto Optimal* iff there is no other solution in  $X$  that dominates  $x^*$  [Brockhoff 2009].

## Pareto set approximations

Specific sets of solutions are the so-called *Pareto set approximations*, which are solution sets of pairwise non-dominated solutions [Brockhoff 2009].

## EC for Pareto Front

Population-based nature of search means you can simultaneously search for set of points approximating Pareto front. Don't have to make guesses about which combinations of weights might be useful. Don't have to make guesses about which combinations of weights might be useful.

### 4.1 NSGA-II

The non-dominated sorting GA-II is an elitist approach which rank and preserve the best solution on top. In order to sort a population of  $N$  size each solution must be compared with every other solution in the population to find if it is dominated. This requires  $\mathcal{O}(MN)$  where  $M$  is the number of objective vectors. When this process is continued to find the members of the first non-dominated class for all population members the total complexity is  $\mathcal{O}(MN^2)$ . At this stage, all individuals in the first non-dominated front are found. In order to find the next front repeat the above procedure. As can be seen the worst case (when there exists only one solution in each front) complexity of this algorithm is  $\mathcal{O}(MN^3)$ . The fast NSGA-II will require at most  $\mathcal{O}(MN^2)$  computations. In the following algorithm, it's worth mentioning that the space complexity is  $\mathcal{O}(N^2)$ . [Deb et al. 2000]. Please refer to the pseudo-code in the *lecture slide*.

## 4.2 Crowding Distance

To get an estimate of the density of solutions surrounding a particular point in the population we take the average distance of the two points on either side of this point along each of the objective vectors. This quantity  $i_{\text{distance}}$  serves as an estimate of the size of the largest cuboid enclosing the point  $i$  without including any other point in the population. The time complexity for the sorting is  $\mathcal{O}(MN \log N)$  [Deb et al. 2000]. Please refer to the *lecture slide* for the pseudo-code and illustration.

## 4.3 Crowding Comparison Operator

The crowded comparison operator ( $\succ_n$ ) guides the selection process at the various stages of the algorithm towards a uniformly spread out Pareto-optimal front. This helps to have a more diverse solution set [Deb et al. 2000].

1. Non-domination rank –  $i_{\text{rank}}$
2. Local crowding distance –  $i_{\text{distance}}$

$$i \succ_n j \text{ if } (i_{\text{rank}} < j_{\text{rank}}) \text{ or } ((i_{\text{rank}} = j_{\text{rank}}) \text{ and } (i_{\text{distance}} > j_{\text{distance}}))$$

## 4.4 Iteration for NSGA-II

Please refer to the *lecture slide* for the pseudo-code.

Initially, a random parent population  $P$  is created. The population is sorted based on the non-domination. Each solution is assigned a fitness equal to its non-domination level (1 is the best level). Thus, minimization of fitness is assumed. Binary tournament selection, recombination, and mutation operators are used to create a child population  $Q_0$  of size  $N$ . Let us now look at the complexity of one iteration of the entire algorithm. The basic operations being performed and the worst-case complexities associated with them are as follows:

1. Non-dominated sort is  $\mathcal{O}(MN^2)$ ,
2. Crowding distance assignment is  $\mathcal{O}(MN \log N)$ , and
3. Sort on  $\geq_n$  is  $\mathcal{O}(2N \log(2N))$ .

As can be seen, the overall complexity of the above algorithm is  $\mathcal{O}(MN^2)$  [Deb et al. 2000].

## 4.5 Deteriorative Cycles

In single-objective optimization, every solution is mapped to a real value and solutions can always be pairwise compared via the less or equal relation  $\leq$  on  $\mathbb{R}$ . In another words, the total order (any elements are comparable)  $\leq \subseteq \mathbb{R} \times \mathbb{R}$  induces via  $f$  order on the search space  $X$  that is a total preorder. In a multiobjective scenario, the  $\leq$  relation is generalized to objective vectors, i.e.,  $\leq$  is a subset of  $\mathbb{R}^k \times \mathbb{R}^k$ . Here, the totality is not given due to vectors  $a, b \in \mathbb{R}^k$  where  $f_1(a) < f_1(b)$  but  $f_2(a) > f_2(b)$ —the relation  $\leq$  on the set of objective vectors is only a partial order, i.e., reflexive, antisymmetric, and transitive. This means that *not any elements are comparable*. [Brockhoff 2009]

### 4.5.1 Downfall of The Evolutionary Algorithm Multiobjective Optimisation

Among the well-established ones, NSGA-II [Deb et al., 2002] and SPEA2 [Zitzler et al., 2002] have to be mentioned here. Both use the Pareto dominance concept as the main selection criterion in an elitist manner where non-dominated solutions are favored over dominated ones. In addition, a second selection criterion establishes diversity among the solutions. However, experimental studies have shown that both algorithms do not scale

well if the number of objectives increases and that a cyclic behavior can be observed. This means that—although non-dominated solutions are preferred over dominated ones—over time, previously dominated solutions enter the population again, resulting in an oscillating distance to the Pareto front. [Brockhoff 2009]

## 4.6 Hypervolume Indicator

One of the latest areas of evolutionary multiobjective optimization where theoretical investigations have been made is the area of hypervolume-based search. The hypervolume indicator, initially proposed for performance assessment, has nowadays been used to guide the search in several multiobjective evolutionary algorithms such as the SMS-EMOA of Beume et al. [2007b], the multiobjective version of CMA-ES [Igel et al., 2007], or HypE [Bader and Zitzler, 2008, 2009]. As mentioned above, the hypervolume indicator is a unary quality measure mapping a set of solutions to the hypervolume in objective space that is dominated by the corresponding objective vectors of the solutions but not by a predefined set of *reference points*. More formally, for a given set  $A \subseteq S^n$  of a solution set and set of reference points in the objective space  $R \subseteq \mathbb{R}^m$ . We define the set of objective vectors that are dominated by  $A$  but not by  $R$  as [Brockhoff 2009]

$$H(A, R) = \{h : \exists a \in A, \exists r \in R : F(a) \leq h \leq r\}$$

And the corresponding hypervolume indicator with reference  $R$  is  $I_H^R$  as

$$I_H^R := \text{VOL} \left( \bigcup_{s \in A} [f_1(s), r_1] \times [f_2(s), r_2] \times \dots \times [f_m(s), r_m] \right), \quad \text{where} \quad \begin{cases} r_i \in R \subseteq \mathbb{R}^m \\ s \in A \end{cases}$$

This means the objective value with the corresponding reference point.

## References

- Brockhoff, Domo (Sept. 2009). *Theoretical Aspects of Evolutionary Multiobjective Optimization—A Review*. Research Report 7030. Inria Research Report RR-7030. France: Inria.
- Deb, Kalyanmoy et al. (2000). *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II*. Tech. rep. KanGAL Report No. 200001. Available at: {deb,samira,apratap,m Kanpur, India: Kanpur Genetic Algorithms Laboratory (KanGAL), Indian Institute of Technology Kanpur. URL: <http://www.iitk.ac.in/kangal>.