On The Dynamics Of Evolutionary Multi-objective Optimization

Tatsuya Okabe, Yaochu Jin, Bernhard Sendhoff

2002

Preprint:

This is an accepted article published in GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference. The final authenticated version is available online at: https://doi.org/[DOI not available]

Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2002), pp. 247-255, 2002

On the Dynamics of Evolutionary Multi-Objective Optimisation

Tatsuya Okabe, Yaochu Jin, Bernhard Sendhoff
HONDA R&D Europe (Deutschland) GmbH
Future Technology Research
Carl-Legien-Strasse 30, 63073 Offenbach/M, Germany
{tatsuya.okabe, yaochu.jin, bernhard.sendhoff}@de.hrdeu.com

Abstract

In evolutionary multi-objective optimisation (EMOO) using dynamic weighted aggregation (DWA), very interesting dynamic behaviours of the individuals have been observed [7] [8]. In this paper, the dynamics of the individuals on fitness space (FS) during multi-objective optimisation (MOO) using evolution strategies (ES) is studied by investigating the mapping of a normal distribution in the parameter space (PS) onto the FS. It is found that the movement of the individuals on the FS is strongly dependent on the characteristics of the projected distribution. Results on three test functions are given to show the good agreement of dynamics predicted from theoretical calculation with that observed in MOO using DWA.

1 INTRODUCTION

Evolutionary algorithms (EA) have been shown to be very successful for multi-objective optimisation (MOO) problems. Up to now a variety of methods for MOO have been proposed [2, 3]. In addition, theoretical studies on the accuracy of the approximation of the Pareto front, on the convergence properties and on the diversity of individuals in a population have also been reported. However, the dynamics of individuals during optimisation, that is, the characteristics of the movement of the individuals on the fitness space (FS) has not yet been investigated to the best of our knowledge. This might be attributed to the fact that in our notion the fitness space[1] of single objective optimisation problems is one dimensional and it can be argued that we might not learn much from observing the dynamics of the population on one axis.

However, in MOO the situation is different, the space is at least two-dimensional and it is believed that the investigation of the dynamics of individuals on the FS will lead to a deeper insight into the properties of Pareto fronts and to a better understanding of the working mechanism of MOO algorithms, which will ultimately enable us to improve the performance of MOO algorithms.

In this paper, we will study the dynamics of the individuals on the FS empirically and analytically by concentrating on the mapping of the mutation distribution from PS to the FS. This approach is motivated by the way evolution strategies produce offsprings, i.e., by adding normally distributed random vectors to the parent parameter vector. Therefore, by understanding the changes the fitness function induces on the normal distribution, we can understand some of the properties of multi-objective evolutionary algorithms based on evolution strategies. At the same time, the notion of a search distribution is not restricted to evolution strategies and has been proposed as a unified approach to the analysis of evolutionary algorithms, see e.g. [10]. Thus, we are confident that the general approach presented in this paper is not restricted to evolution strategies.

The work in this paper is partly motivated by the behaviours observed in MOO using the dynamically weighted aggregation (DWA) algorithm [7, 8]. The basic idea is to combine the optimisation objectives with different weights, which are changed dynamically during optimisation so that a set of Pareto-optimal solutions instead of one single solution will be obtained. It has been shown that the method is not only effective for problems with a convex Pareto front, but also for those with a concave Pareto front. In the opti-

[1] This should not be confused with the notion of fitness landscape, where the fitness values are mapped over the parameter or more general genotype/phenotype configurations.

misation, it is found that when the weights change, the individuals move along the Pareto front once they reach one point on it, even if the Pareto front is discontinuous. To understand why individuals follow the Pareto front was the initial target of this work. Note, that this type of movement is even observed for sudden large weight changes and *after* mutation but *before* selection. Therefore, the trivial explanation that the individuals simply follow the Pareto front since it is the path of "highest fitness" is not sufficient.

Of course, we had to choose objective functions for which we carry out our analysis, we chose three functions (concave, convex, discontinuous). In the course of our work, it turned out that the presented analysis can also be used to determine whether and when (e.g. with respect to the initialisation) a test problem is difficult or not for a specific algorithm; we will comment on these findings in Section 5.

The remainder of this paper is structured as follows: In the next section, we will briefly recall some of the theoretical work on MOO. In Section 3, we will concisely outline Evolution Strategy, the DWA method, the test functions and present some of the empirical observations. In Section 4, we will analyse the transformation of the mutation distribution and relate it to the results from Section 3. Further implications of Section 4, e.g. for the difficulty of test functions are discussed in Section 5.

2 THEORY FOR EMOO

Results on the convergence of evolutionary multiobjective optimisation have been presented by Rudolph [12, 13] based on the Markov chain approach which has been successfully used for the analysis of single objective evolutionary algorithms, see e.g. [11] among others. The work by Hanne [6] is also mainly concerned with the convergence of evolutionary multiobjective algorithms. Complexity issues have been addressed for example by van Veldhuizen [17]. Very recently an interesting approach has been suggested by Thiele et al. [16] to define a simple problem class for multi-objective optimisation to facilitate the theoretical analysis of evolutionary algorithms for this domain.

Teich presented some theoretical investigations for uncertain objectives for MOO [15], based on which he developed the Estimate Strength Pareto Evolutionary Algorithm (ESPEA).

Since the comparison of different algorithms for multiobjective optimisation is much harder than for single objective ones, it has also been the focus of some theoretical investigations, which are mainly based on a statistical approach using an appropriate metric, see e.g. the work by Fonseca et al. [5] and by Zitzler et al. [19].

Even though the above list is likely to be incomplete, compared to the overall number of publications in MOO, theoretical approaches have been sparse in particular for the analysis of the dynamics of individuals during the evolutionary search and of the main factors which determine the characteristics of this movement. Surely the approach in this work can only be seen as a starting point, however, we believe it can be beneficial to explain some of the empirical observations, which we will outline in the next section.

3 DYNAMICALLY WEIGHTED AGGREGATION

3.1 EVOLUTION STRATEGIES

In evolution strategies (ES), mutation plays the major role in search. The mutation is performed by adding a random number generated from a normal distribution $N(0, \sigma_i^2)$, where σ_i is the standard deviation. In the standard ES, new individuals are generated in the following way [1]:

$$\vec{x}(t) = \vec{x}(t-1) + \vec{z} \quad (1)$$

$$\sigma_i(t) = \sigma_i(t-1)\exp(\tau' z)\exp(\tau z_i), \quad (2)$$

where, $\vec{x}$ is an n-dimensional parameter vector, $\vec{z}$ is an n-dimensional random number vector with $\vec{z} \sim N(0, \sigma(t)^2)$, z and z_i are normally distributed random numbers with $z, z_i \sim N(0, 1)$. In ES, the σ_i are also called stepsizes, and are evolved together with the object parameters. This is known as self-adaptation, which is an important feature of the ES.

The parameters τ and τ' in equation (2) are constants that are given as follows:

$$\tau = \frac{1}{\sqrt{2\sqrt{n}}} \quad (3)$$

$$\tau' = \frac{1}{\sqrt{2n}} \quad (4)$$

In the ES usually a deterministic selection method is used. In the non-elitist (μ, λ) method, the best μ individuals from the λ offspring are chosen as the parents of the next generation.

3.2 BASIC IDEA OF DWA

Jin et al. [7, 8] proposed dynamically weighted aggregation as an efficient method to easily apply any evolu-

tionary algorithm (and evolution strategies in particular) to multi-objective optimisation problems. Since our empirical observations are based on the DWA method, we will explain it briefly in the following.

The basic idea is to linearly combine all objectives like in the conventional aggregation method: $f = \sum_{i=1}^{m} w_i f_i$. Here m, w_i and f_i are the number of objective functions, the weights for the f_i and the objective functions, with $\sum_{i=1}^{m} w_i = 1$. In order to obtain the whole Pareto front, the weights w_i are changed dynamically in each generation using a periodical function between $[0, 1]$, for example, the sine function. To achieve the whole Pareto front, it is necessary to maintain an archive of non-dominated solutions.

Whereas conventional weighted aggregation methods with restart cannot obtain concave Pareto fronts, it has been argued and empirically demonstrated in [8] that the DWA methods indeed can (at least if some mild assumptions about the changing functions for the $w_i(t)$ are made).

3.3 TEST FUNCTIONS

Several test functions for multi-objective optimisation have been proposed in the literature, see e.g. [9, 18, 4, 14]. Here we chose three functions whose Pareto front can be calculated analytically and which represent the three important cases of convex, concave and discontinuous Pareto fronts.

3.3.1 Function T_1 (Convex Case)

The first test function T_1 is defined as follows [9]:

$$f_1 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad (5)$$

$$f_2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - 2.0)^2 \quad (6)$$

The convex Pareto front can be calculated analytically with the following result:

$$f_2 = f_1 - 4\sqrt{f_1} + 4 \quad (7)$$
$$0 \le f_1 \le 4, 0 \le f_2 \le 4$$

3.3.2 Function T_2 (Discontinuous Case)

The second test function T_2 is defined as follows [18]:

$$f_1 = x_1 \quad (8)$$

$$f_2 = g \times \left(1 - \sqrt{\frac{f_1}{g}} - \frac{f_1}{g}\sin(10\pi\, f_1)\right) \quad (9)$$

$$g(x_2, \ldots, x_n) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i$$
$$x_i \in [0, 1]$$

The discontinuous Pareto front is:

$$f_2 = 1.0 - \sqrt{f_1} - f_1 \sin(10\pi f_1), \quad (10)$$

where f_1 can be from the following intervals: $f_1 \in [0.0000, 0.0830]$, $(0.1823, 0.2579]$, $(0.4095, 0.4541]$, $(0.6187, 0.6528]$, $(0.8237, 0.8523]$. These constraints on f_1 were not obtained analytically, but from simulations. As it is evident, the Pareto front is discontinuous.

3.3.3 Function T_3 (Concave Case)

The third test function T_3 is defined as follows [4, 14]:

$$f_1 = 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) \quad (11)$$

$$f_2 = 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2\right) \quad (12)$$

This original test function ($n = 8$) was proposed by Fonseca and Fleming in 1993. Here we generalised it to the n-dimensional case. The concave Pareto front is given by:

$$f_2 = 1 - \frac{1 - f_1}{e^4}\exp\left(4\sqrt{-\log(1 - f_1)}\right) \quad (13)$$
$$f_1 = [0, 1 - e^{-4}]$$

3.4 DYNAMICS OF THE DWA ALGORITHM

When we apply the DWA algorithm to the three test functions described in the previous section and observe the dynamics of the individuals in particular during the movement along the Pareto front, some interesting phenomena can be observed. Unfortunately, we cannot *show* the dynamics directly, therefore, we have to present snapshots for different generations and describe the behaviour in between.

For all experiments, the following weight change function was used:

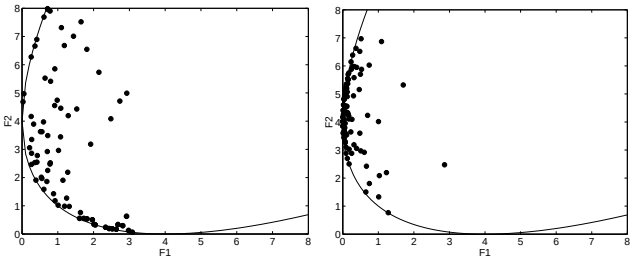
$$w_1 = \frac{1}{2}\text{sign}\left(-\sin(0.1\pi\, t)\right) + \frac{1}{2}, \quad (14)$$

where t is the number of generations.

In this paper, we use standard ES with $(\mu, \lambda) = (15, 100)$, and the archive size is 200. The initial standard deviation is 0.1, 0.01 and 0.1 for test functions

T_1, T_2 and T_3 respectively. Two dimensional cases, i.e. $n = 2$, are shown.

In Figure 1 the distributions of individuals (circles) *after* mutation and *before* selection[2] are shown for test function T_1. Whereas in Figure 1(a) the distribution is fairly widespread, in Figure 1(b) is concentrated on the Pareto front. In both cases selection has little influence on the shape of the individual's distribution which is mainly determined by the shape of the mutation distribution of the parents. Even though, one can argue that the diversity is considerably decreased in later parent generations, this does not account for the non-isotropic nature of the distribution, which is particularly evident from Figure 1 (b), where nearly all offspring - before selection - are located on the Pareto front. Indeed if we observe the continuous dynamics, we see that the individuals move nearly perfectly along the Pareto front, which makes the search very efficient.

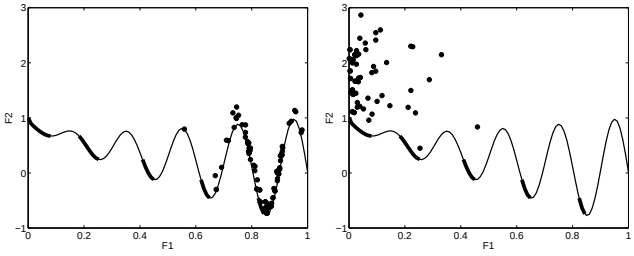


(a) (b)

Figure 1: Distribution of the individuals *after* mutation and *before* selection for test function T_1 for the DWA algorithm after generation 1 in figure (a) and generation 18 in figure (b).

Similar dynamics can be observed for the second test function T_2, for which snapshots of the distribution of the individuals before selection are shown in Figure 2. The solution to equation (10) is given by the thin curve and the Pareto front (all non-dominated solutions), which consists of parts of this curve, is given by the thicker curve elements. Again we see that when the parents are located near $(f_1, f_2) = (1, 0)$ at generation 18, (Figure 2(a)), the individuals are restricted to the set of non-dominated solutions, whereas near $(f_1, f_2) = (0, 1)$ at generation 29, (Figure 2(b)), the distribution of individuals is rather wide spread. Indeed, dynamically one can nicely observe, how the individuals move from $(f_1, f_2) = (0, 1)$ to $(f_1, f_2) = (1, 0)$ and back along the thin curve being wider distributed near the left end and strongly concentrated near the right end of the curve in Figure 2. This movement can be observed several times during the periodic change of the weights, according to equation (14).

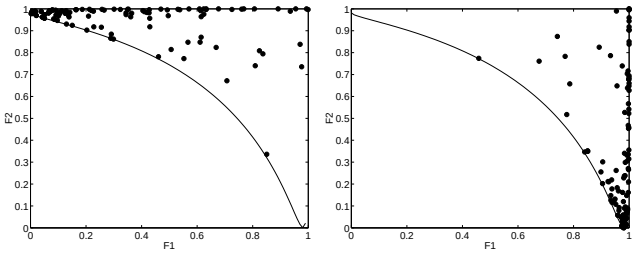
[2]In each generation the complete evolution cycle (mutation *and* selection) is carried out. However, our snapshots are shown for one generation, say t_1, and the distribution is shown after mutation$_{t_1}$ and before selection$_{t_1}$ in order to highlight that the non-isotropic distribution of the individuals is to a large degree a result of mutation and not just of selection.



(a) (b)

Figure 2: Distribution of the individuals *after* mutation and *before* selection for test function T_2 for the DWA algorithm after generation 18 in figure (a) and generation 29 in figure (b).

For the third test function the results for generation 16 and 26 are shown in Figure 3(a) and (b). We observe that individuals are clustered near three points: $(f_1, f_2) = (0, 1)$, $(f_1, f_2) = (1, 0)$ and $(f_1, f_2) = (1, 1)$. The interior of this "triangle" bounded from below by the Pareto front, is very sparsely represented. During the dynamical observation, the peculiarity of the three points becomes even more evident, since in some generations nearly all individuals are concentrated in these points irrespective of the weight changes.



(a) (b)

Figure 3: Distribution of the individuals *after* mutation and *before* selection for test function T_3 for the DWA algorithm after generation 16 in figure (a) and generation 26 in figure (b).

We can summarise our observations as follows:

- The distribution of individuals strongly depends

on the position in search space nearly irrespective of selection.

- The movement of individuals between points on the Pareto front follows a very distinctive pattern which is also not only controlled by selection.

4 THE SHAPE OF THE MUTATION DISTRIBUTION IN FITNESS SPACE

We have argued in the previous section that the characteristics of our empirical observations for all three test functions is to a large extent independent of selection. Therefore, it has to depend on the shape of the mutation distribution which is for evolution strategies in PS the normal distribution. Of course, when we observe the movement of individuals in FS, we must consider what the shape of the normal distribution looks like in FS. We will see that the transformed distribution can have very distinctive features which help to explain our empirical observations. The fitness values at generation t after mutation are given by

$$f(x(t)) = f(x(t-1)+z),\ z \sim N(0,\sigma_t^2) \tag{15}$$

Here f, $x(t)$ and σ_t are the objective function, the design variable at generation t, and the standard deviation at generation t. In this section, we neglect self-adaptation, which implies $\sigma_t = \sigma$. We will take it into consideration in the future work.

Without loss of generality, we restrict the following analytical investigation to the case $n = 2$.

In the two-dimensional case, the normal distribution on PS is given by:

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \tag{16}$$

Here, $f(x_1,x_2)$, σ and μ are the probability density function (pdf), the standard deviation and the mean, respectively.

Now, let us firstly assume that f_1 and f_2 are one-to-one functions, i.e. $x = f^{-1}(f(x))$. The other case, e.g. test function T_1, will be dealt with later. From this assumption, we can get the following equation:

$$\int_{f_1}^{f_1+\Delta f_1}\int_{f_2}^{f_2+\Delta f_2} g(f_1',f_2')df_2'df_1' = \int_{x_1}^{x_1+\Delta x_1}\int_{x_2}^{x_2+\Delta x_2} f(x_1',x_2')dx_2'dx_1' \tag{17}$$

Here $g(f_1,f_2)$ is the pdf in the FS. The area (x_1,x_2) - $(x_1+\Delta x_1, x_2+\Delta x_2)$ corresponds to (f_1,f_2) - $(f_1+\Delta f_1, f_2+\Delta f_2)$. From equation (17) we obtain

$$g(f_1,f_2) = \frac{1}{|J|} f(x_1,x_2), \tag{18}$$

where J is the Jacobian matrix.

Now we can calculate the mutation distribution in FS for each of the three test functions and compare the results with the observations of Section 3.

4.1 FUNCTION T_1 (CONVEX CASE)

$$g(f_1,f_2) = \frac{1}{2|x_2-x_1|}\left\{f(x_1,x_2)+f(x_2,x_1)\right\} \tag{19}$$

$$(x_1,x_2) = \frac{1}{4}(f_1-f_2+4)(1,1) + (\pm c, \mp c), \tag{20}$$

with c given by

$$c = \frac{1}{4}\sqrt{-f_1^2-f_2^2-16+2f_1f_2+8f_1+8f_2} \tag{21}$$

In order to obtain equation (19), we had to slightly modify equation (17) to take the fact into account that T_1 is not one-to-one:

$$\int_{f_1}^{f_1+\Delta f_1}\int_{f_2}^{f_2+\Delta f_2} g(f_1',f_2')df_2'df_1' = \int_{x_1}^{x_1+\Delta x_1}\int_{x_2}^{x_2+\Delta x_2} f(x_1',x_2')dx_2'dx_1' + \int_{-x_1-\Delta x_1}^{-x_1}\int_{-x_2-\Delta x_2}^{-x_2} f(x_1',x_2')dx_2'dx_1' \tag{22}$$

The results for some points in the FS are shown in Figure 4, the standard deviations are given by $(\sigma_1,\sigma_2) = (1,1)$.

Figures 4 demonstrate that once one solution on the Pareto front is found the individuals will move along the Pareto front with a high probability - independent of selection - solely because the shape of the normal distribution which defines mutation in evolution strategies is mapped to the shape shown in figures 4 (a)-(c). However, when the Pareto front has not been reached yet, i.e. the individual is concentrated in the interior, the search distribution is nearly isotropic and the success solely depends on selection, as shown in Figure 4(d).

4.2 FUNCTION T_2 (DISCONTINUOUS CASE)

$$g(f_1,f_2) = \frac{1}{9-\frac{9\sqrt{x_1}}{2\sqrt{1+9x_2}}} f(x_1,x_2) \tag{23}$$

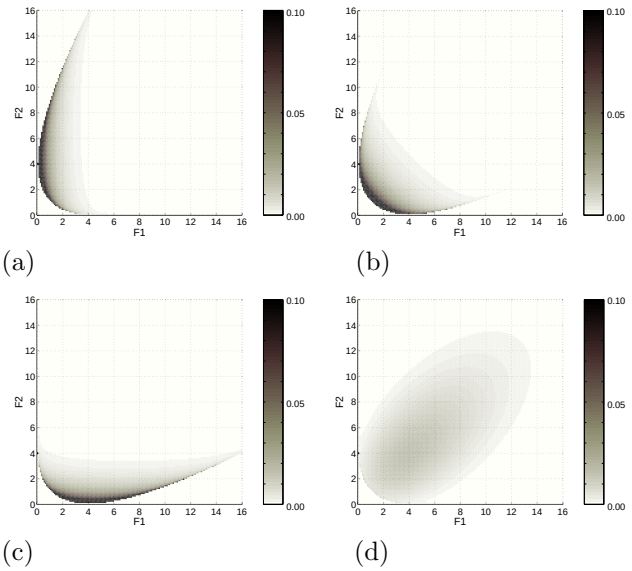


Figure 4: The shape of the normal distribution in FS for test function T_1. (a) $(x_1, x_2) = (0, 0)$, $(f_1, f_2) = (0, 4)$; (b) $(x_1, x_2) = (1, 1)$, $(f_1, f_2) = (1, 1)$; (c) $(x_1, x_2) = (2, 2)$, $(f_1, f_2) = (4, 0)$; (d) $(x_1, x_2) = (-1, 3)$, $(f_1, f_2) = (5, 5)$.

$$
\begin{aligned}
(x_1, x_2) &= \left(f_1, \frac{1}{9}(f_2 + \frac{f_1}{2} + f_1 \sin(10\pi f_1) + c) \right) \\
c &= \sqrt{f_1 f_2 + \frac{f_1^2}{4} + f_1^2 \sin(10\pi f_1)} - 1
\end{aligned}
$$

From the shape of the transformed mutation distributions in Figure 5 (a) and (b) we conjecture that individuals located at $(f_1, f_2) = (1, 0)$ are very likely to move along the curve of non-dominated solutions and discover the discontinuous Pareto front. Whereas, the opposite direction, i.e. the movement of individuals starting at $(f_1, f_2) = (0, 1)$ along this curve, is much less likely since the distribution is more isotropic.

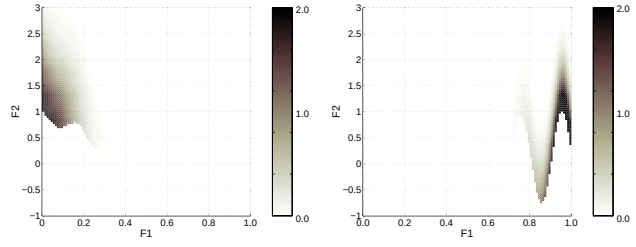


Figure 5: The shape of the normal distribution in FS for test function T_2. (a) $(x_1, x_2) = (0, 0)$, $(f_1, f_2) = (0, 1)$; (b) $(x_1, x_2) = (1, 0)$, $(f_1, f_2) = (1, 0)$.

4.3 FUNCTION T_3 (CONCAVE CASE)

$$
\begin{aligned}
g(f_1, f_2) &= \frac{1}{|J|} \{ f(x_1, x_2) + f(x_2, x_1) \} && (24) \\
J &= 4\sqrt{2} \exp\left(-2(x_1^2 + x_2^2 + 1)\right)(x_1 - x_2) \\
(x_1, x_2) &= \frac{\sqrt{2}}{8}(h_1 - h_2)(1, 1) + (\pm c, \mp c) && (25)
\end{aligned}
$$

Here, c, h_1 and h_2 are:

$$
\begin{aligned}
c &= \sqrt{-h_1^2 - h_2^2 - 16 + 2h_1 h_2 + 8h_1 + 8h_2} \\
(h_1, h_2) &= (-\log(1 - f_2), -\log(1 - f_1))
\end{aligned}
$$

Figure 6 shows the logarithm of the transformed mutation distribution in FS for different individuals positioned on the Pareto front as well as on the interior. We observe that in all cases the "boundaries" have a very high probability whereas the interior of the shown part of the fitness space has a very small probability. Thus, even without selective pressure individuals are very likely to move along the concave Pareto front, simply due to the shape of the transformed mutation distribution.

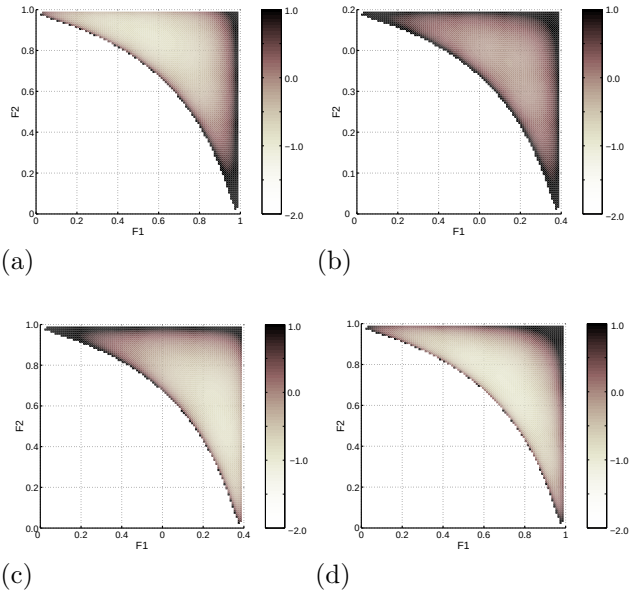


Figure 6: The shape of the normal distribution in FS for test function T_3, logarithmic values are shown. (a) $(x_1, x_2) = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, $(f_1, f_2) = (1 - e^{-4}, 0)$; (b) $(x_1, x_2) = (0, 0)$, $(f_1, f_2) = (1 - e^{-1}, 1 - e^{-1})$; (c) $(x_1, x_2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $(f_1, f_2) = (0, 1 - e^{-4})$; (d) $(x_1, x_2) = (-1, 1)$, $(f_1, f_2) = (1 - e^{-3}, 1 - e^{-3})$;

5 FURTHER STUDIES ON FUNCTION T_1

In order to better understand the constraints which lead to the "compression" of the mutation distribution onto the Pareto front in some cases, we look at Function T_1 again in a bit more detail.

The Pareto front in the parameter and the fitness space for function T_1 is shown in Figure 7 by the thick curve, which is a straight line in PS. The parallel lines $x_2 = x_1 \pm c\sqrt{2}, (0 \leq x_1 + c\sqrt{2}/2 \leq 2)$, above and below the PS-Pareto line in a distance c are projected to the following curves:

$$f_2 = f_1 - 4\sqrt{f_1 - \frac{1}{2}c^2} + 4. \tag{26}$$

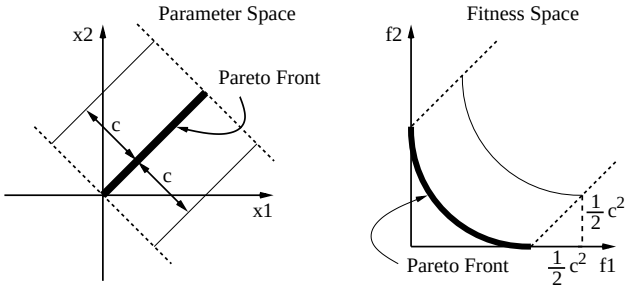
Equation (26) can be written as follows:

$$f_2 - \frac{1}{2}c^2 = f_1 - \frac{1}{2}c^2 - 4\sqrt{f_1 - \frac{1}{2}c^2} + 4, \tag{27}$$

with the constraints

$$0 \leq f_1 - \frac{1}{2}c^2 \leq 4 \tag{28}$$

$$0 \leq f_2 - \frac{1}{2}c^2 \leq 4. \tag{29}$$



(a) (b)

Figure 7: (a) The Pareto front in PS and parallel lines with distance c; (b) the Pareto front in FS and the images of the parallel lines in FS.

From the above considerations and from Figure 7, we can better understand in which way the mutation distribution is changed. The distance to the Pareto front in PS for individuals which lie on one of the parallel lines is c, in FS it is $\frac{1}{\sqrt{2}}c^2$. Therefore, for $c = \sqrt{2}$ this distance remains unchanged. Whereas for $c < \sqrt{2}$ the distance is decreased or "compressed" under the mapping of function T_1, it is increased for $c > \sqrt{2}$.

In the context of the probability distribution, it means that if the individual is located below the thin curve $\frac{1}{\sqrt{2}}c^2$ for $c = \sqrt{2}$ in Figure 7 (b), the probability of the area closer to the Pareto front is increased. The opposite holds for individuals above this curve for which it becomes more unlikely to move towards the Pareto front.

For function T_1 the recommended initialisation [9] of the parameters is $-2 \leq x_1 \leq 2$ and $-2 \leq x_2 \leq 2$. Therefore, the uniform probability density is given by:

$$p(x_1, x_2) = \begin{cases} 0.0625 & , -2 \leq x_i \leq 2 \quad (i = 1, 2) \\ 0 & , \text{else} \end{cases} \tag{30}$$

In fitness space this probability density is projected as follows:

$$\begin{aligned} p(f_1, f_2) &= \begin{cases} \frac{1}{8b} & , -2 \leq \frac{1}{4}\{a \pm b\} \leq 2 \\ 0 & , \text{else} \end{cases} \\ a &= f_1 - f_2 + 4 \\ b &= \sqrt{-f_1^2 - f_2^2 - 16 + 2f_1f_2 + 8f_1 + 8f_2} \end{aligned} \tag{31}$$

Equation (31) is shown in Figure 8 with logarithmic scale, it agrees well with simulations which we carried out.

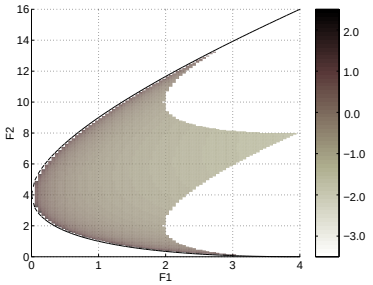
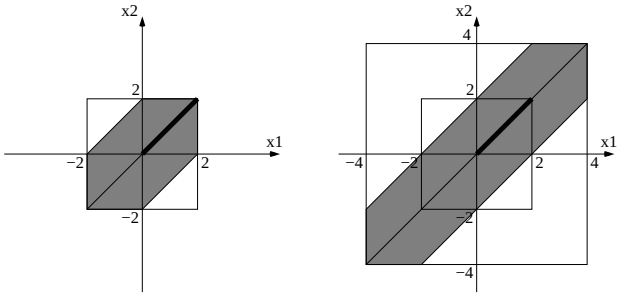


Figure 8: The shape of the probability density in fitness space for individuals, which have uniform distribution on $[-2, 2]$ in parameter space for function T_1.

We observe that the probability is very high for points on or close to the boundary including Pareto front and that it decreases with increased distance from the boundary. From our considerations above, this can be easily understood. For all points, which lie within the corridor show in Figure 7 for $c = \sqrt{2}$, their distance to the boundary (the Pareto front is the section of the boundary between the two coordinates) is reduced under the mapping. For the square $-2 \leq x_i \leq 2$, $(i = 1, 2)$, these are 3/4 of all points, thus the probability density is increased near the boundary, see Figure 9.

In Figure 9 the thick lines represent the Pareto front on PS and the gray areas are "compressed" regions.



(a) (b)

Figure 9: The area of initialisation on T_1. (a) shows the initialisation $[-2, 2]$. The gray ("compressed") area covers 3/4 of the total region. (b) shows $[-4, 4]$. The gray area covers 7/16 of the total region.

We conclude that the initialisation of the individuals, even if it is uniform in parameter space, might result in an important bias in fitness space. In particular for problem T_1 we can conclude that the proposed initialisation will lead to an initial population where it is rather likely that some individuals are already located on or very near the Pareto front. If this initialisation is e.g. used together with the DWA method the performance of the algorithm will be very high, however, mostly because of the specific relations between PS and FS. Generally, we would advise to use an initialisation $-4 \leq x_i \leq 4$, $(i = 1, 2)$, at least, where the number of points whose distance is decreased is roughly equal to the ones whose distance is increased (relation 7/9) to reduce the probability that initial individuals are already located near the Pareto front. The shape of the probability density in the fitness space is shown in Figure 10 when the parameters are initialised on $[-4, 4]$ in parameter space.

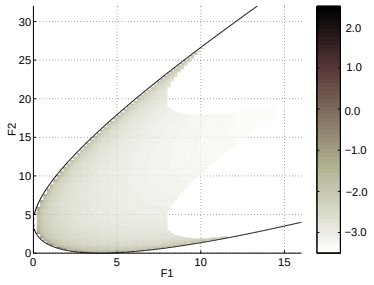


Figure 10: The shape of the probability density in fitness space for individuals, which have uniform distribution on $[-4, 4]$ in parameter space for function T_1.

6 CONCLUSION

In this paper, we investigated the dependence of the dynamics of the individuals in fitness space on the properties of the mapping of the probability density function for mutation or more general for the population of the next generation from parameter space to fitness space. The analysis which we presented here is restricted to evolution strategies or at least to those evolutionary algorithms where a normally distributed mutation is the main operator, e.g. evolutionary programming. Although we did not explicitly test this, we are very confident that the results of this paper are valid for any selection method, indeed the analytical investigations in Section 4 and 5 are completely independent from selection.

We believe our approach can be a starting point for a more general investigation of the influence of the PS-FS mapping on the search distribution which is usually only discussed in the PS. In particular for multi-objective optimisation where FS is at least two-dimensional, the movement of the individuals on this space can show a much richer dynamics. Although the importance of the PS-FS mapping on all aspects of evolutionary algorithms is widely accepted, the analysis of its influence on the search without selection has not received much attention so far. The fact that in some cases the Pareto front is a local attractor for the population (in a probabilistic sense, see Figure 4) without the influence of selection seems to be worth noticing.

We started our analysis with the observation of some dynamic behaviour of the DWA method under some conditions. However, as we have shown in the former section, the results are not restricted to the DWA or related methods, but bear strong implications on such important questions, like "When is a test function difficult?". In the literature this question is usually discussed in the context of such properties like deceptiveness, ruggedness, etc., however, the much simpler notion of the relation of distances in PS and in FS is hardly addressed. Nevertheless, as Figure 8 and 10 shows, it might have a very direct influence on the algorithm's performance, without telling much about the true *strength* of the algorithm.

Acknowledgements

The authors would like to thank E. Körner for his support and T. Arima and M. Olhofer for discussions on the subject.

References

[1] T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, 1996.

[2] C.A. Coello Coello. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. *Knowledge and Information Systems*, 1(3):269–308, 1999.

[3] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, 2001.

[4] C.M. Fonseca and P.J. Fleming. Genetic algorithm for multiobjective optimization, formulation, discussion and generalization. In *Genetic Algorithms: Proceedings of the Fifth International Conference*, pages 416–423, 1993.

[5] C.M. Fonseca and P.J. Fleming. On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers. In *Lecture Notes in Computer Science 1141 Parallel Problem Solving from Nature - PPSN IV*, pages 584–593, 1996.

[6] T. Hanne. On the convergence of multiobjective evolutionary algorithms. *European Journal of Operational Research*, 117(3):553–564, 1999.

[7] Y. Jin, T. Okabe, and B. Sendhoff. Adapting Weighted Aggregation for Multiobjective Evolution Strategies. In *Lecture Notes in Computer Science 1993 Evolutionary Multi-Criterion Optimization*, pages 96–110, 2001.

[8] Y. Jin, M. Olhofer, and B. Sendhoff. Dynamic Weighted Aggregation for Evolutionary Multi-Objective Optimization: Why Does It Work and How? In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1042–1049, 2001.

[9] J.D. Knowles and D.W. Corne. Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation*, 8(2):149–172, 2000.

[10] H. Mühlenbein and T. Mahnig. Evolutionary Algorithms: From Recombination to Search Distributions. In *Theoretical Aspects of Evolutionary Computing*, pages 137–176. Springer, 2000.

[11] G. Rudolph. Convergence Analysis of Canonical Genetic Algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.

[12] G. Rudolph. On a Multi-Objective Evolutionary Algorithm and Its Convergence to the Pareto Set. *Proceedings of the 5th IEEE Conference on Evolutionary Computation*, pages 511–516, 1998.

[13] G. Rudolph and A. Agapie. Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Congress on Evolutionary Computation 2000*, volume 2, pages 1010–1016, 2000.

[14] K.C. Tan, E.F. Khor, C.M. Heng, and T.H. Lee. Exploratory Multi-objective Evolutionary Algorithm: Performance Study and Comparisons. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 647–654, 2001.

[15] J. Teich. Pareto-Front Exploration with Uncertain Objectives. In *Lecture Notes in Computer Science 1993 Evolutionary Multi-Criterion Optimization*, pages 314–328, 2001.

[16] L. Thiele, K. Deb, and E. Zitzler. Talk presented at the Dagstuhl Seminar "Theory of Evolutionary Computation", January 2002.

[17] D.A. van Veldhuizen and G.B. Lamont. Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art. *Evolutionary Computation*, 8(2):125–147, 2000.

[18] E. Zitzler, K. Deb, and L. Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, 2000.

[19] E. Zitzler and L. Thiele. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.