Multidimensional Optimization with Genetic Algorithms using Fuzzy Fitness Functions

S. Voget and M. Kolonko
Institut für Mathematik, Universität Hildesheim
Marienburger Platz 22, D-31141 Hildesheim
Tel.: ++49 5121 883713
e-mail kolonko@informatik.uni-hildesheim.de

Abstract: We present a new method to approximate the set of Pareto-optimal solutions in multi-criteria optimization with genetic algorithms. The direction of the selection pressure is adapted to the actual state of the population forcing it to explore a broad range of so far undominated solutions. The adaptation is done by a fuzzy-type control of the fitness function and the selection mechanism. Experiments with a timetable optimization problem show that this fuzzy adaptive approach avoids most of the shortcomings of other approaches.

KEYWORDS: OPTIMIZATION WITH MULTIPLE CRITERIA, GENETIC ALGORITHMS, ADAPTIVE FITNESS FUNCTION, PARETO-OPTIMAL SOLUTIONS

1 Introduction

We consider the multidimensional cost function $f(x) = (f_1(x), \dots, f_K(x))$ where the solutions x vary over a set S and $f_i: S \to \mathrm{IR}_+,\ \ 1 \leq i \leq K$. We say $x \in S$ *dominates* $x' \in S\ \ (x \preceq x')$ if $f_i(x) \leq f_i(x')$ for all $i = 1, \dots, K$ and $f_j(x) < f_j(x')$ for at least one $j \in \{1, \dots, K\}$. Let $P := \{\, x \in S \mid \neg\exists x' \in S\ \ x' \preceq x \,\}$ be the set of *Pareto-optimal* solutions ([7]). Our aim is to find an approximation of P.
Genetic Algorithms (GA) (cp. [2]) provide a general optimization framework particularly suited for treating multidimensional optimization problems (cp. [1, 3, 6]). The lower envelope of $F_n = \{(f_1(x), ..., f_K(x)) | x \in G_n\}$, where G_n is the population of solutions in the n-th step, provides a natural approximation of P if the population G_n is forced into the 'right' direction of the solution space. As usual G_n is transformed into G_{n+1} by production (crossover, mutation etc.) and reduction (selection according to a fitness function $g(x)$).
The VEGA algorithm ([6]) selects for each i, $1 \leq i \leq K$, a fixed number $\frac{N}{K}$ of individuals according to the fitness $g(x) := f_i(x)$ where N denotes the population size. In our experiments, this procedure led to a splitting of the population as indicated in Fig. 1 (b) (see also [3]). The system MOGA ([1]) selects an individual x according to its *Pareto-rank*

$$r(x) := 1 + |\{\, x' \in G_n^+ \mid x' \preceq x \,\}| \tag{1}$$

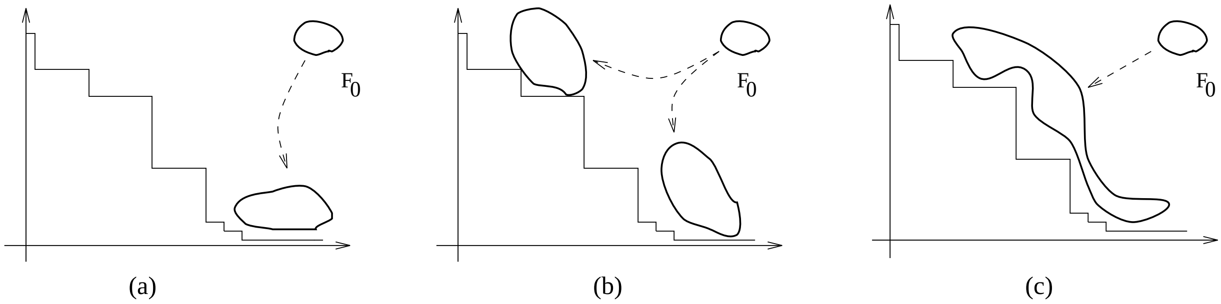


Figure 1: Starting from F_0 the population may deviate as in (a) or split as in (b). (c) would be an ideal evolution. The step function marks the (unknown) set of Pareto-optimal values.

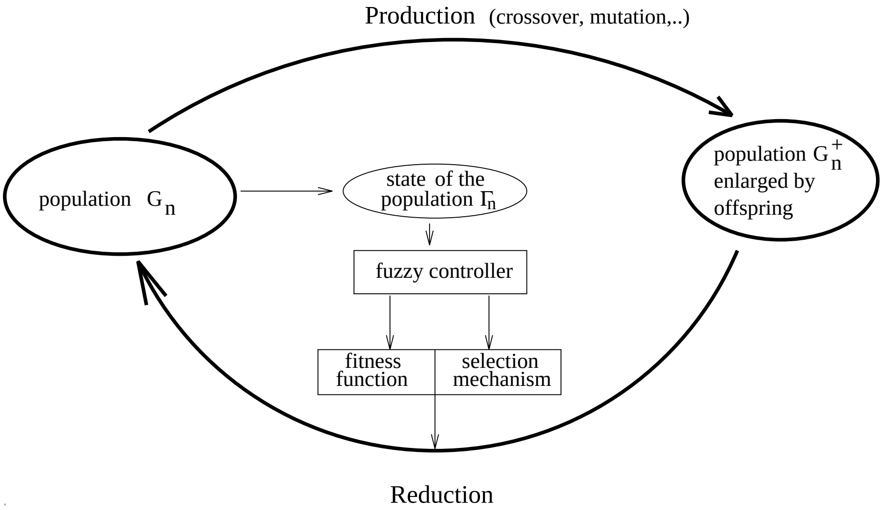


Figure 2: The production/reduction cycle of a genetic algorithm uses a fitness function and a selection mechanism that are adapted to the state of the population by a fuzzy controller.

This approach tends to a behaviour as indicated in Fig. 1 (a), see also Fig. 5.
Our approach observes a 'state' Γ_n of the population G_n at each step n. It then adapts the fitness function $g_n(x)$ and the selection scheme to be used to the actual state Γ_n. A fuzzy controller decides on how to choose the selection for the next reduction $G_n^+ \to G_{n+1}$, cp. Fig. 2 for an overview.

2 Defining Desirable Goals

To define what a 'desirable' behaviour of F_n is, in order to approximate the Pareto-optimal points, one needs overall knowledge of the cost landscape which is not available. Instead, we take a random sample $x_1, \ldots, x_N$ of solutions and examine their cost function values. We take the statistical properties of this sample as an estimate of basic properties of the problem.

In the context of genetic algorithms, a suitable chosen starting population G_0 provides a sample of the solution space. We compare the statistics of G_0 (or rather F_0) with those of G_n and collect the results in the state Γ_n. Γ_n serves as input for the fuzzy controller. We use the following basic data of a population G_n, $n \geq 0$:

$$\begin{aligned}
\lambda_i(n) &:= \min_{x\in G_n} f_i(x) \\
\mu_i(n) &:= \lambda_i(n) + \left(\max_{x\in G_n} f_i(x) - \min_{x\in G_n} f_i(x)\right)/2 \\
\sigma_i(n) &:= \max_{x\in G_n} f_i(x) - \min_{x\in G_n} f_i(x)
\end{aligned}$$

for $i = 1, \ldots, K$. Then $\lambda(n) := (\lambda_1(n), \ldots, \lambda_K(n))$ describes the lower half of the enclosing rectangle of F_n, more precisely, $\lambda(n)$ is the Pareto-optimal point of this rectangle. $\mu(n) := (\mu_1(n), \ldots, \mu_K(n))$ is the center of the rectangle and $\sigma(n) := (\sigma_1(n), \ldots, \sigma_K(n))$ represents its volume. $\lambda(n)$ and $\sigma(n)$ completely determine the enclosing rectangle. Note that $\mu(n)$ and $\sigma(n)$ could be replaced by the mean and variance of F_n, but these terms seem to be less appropriate for our purpose.

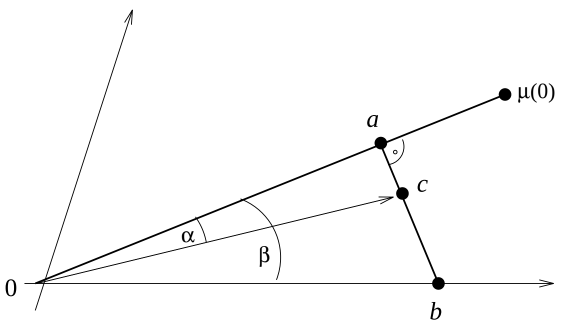
Our overall goal is to push F_n towards $0 = (0, \ldots, 0)$ in $\mathbb{R}_+^K$ with a broad but connected front (as indicated in Fig. 1 (c)). If lower bounds $L_i > 0$ for f_i $(1 \leq i \leq K)$ are known then $(0, \ldots, 0)$ should be replaced by $(L_1, \ldots, L_K)$ in the sequel. Similarly, instead of the positive quadrant $\mathbb{R}_+^K = [0, \infty)^K$ we would then consider $[L_1, \infty) \times \cdots \times [L_K, \infty)$ as the image space of f. From this overall goal we may derive the following more detailed goals.
Goal 1: $\mu(n)$ should stay on the line $\overline{0, \mu(0)}$ connecting 0 and $\mu(0)$.
Goal 2: $\sigma(n)$ should not become smaller than $\sigma(0)$.
Goal 3: F_n should form a connected cloud, i.e. G_n should not be split into separate subpopulations.
Goal 4: The lower half of the enclosing rectangle of F_n should move towards $(0, \ldots, 0)$ without deviating into one direction, i.e. $\lambda(n)$ should move along $\overline{0, \lambda(0)}$.

Goal 1 tries to avoid situations as in Fig. 1. It also reflects our experience that it is particularly difficult to obtain good solutions in the 'center', i.e. with all $f_i(x)$, $1 \leq i \leq K$, small. Goal 2 avoids loosing the variation within a population which easily occurs if the selection is too strict. Goal 3 is to avoid situations as in Fig. 1 (b) observed e.g. in the system VEGA, see Fig. 5. Goal 4 enforces Goal 1 in that it requires not only the center $\mu(n)$ but also the lower half of the enclosing rectangle to move towards the minimal value $(0, \ldots, 0)$.

Note that one could easily add other goals in a similar spirit. The above collection turned out to be quite successful in our experiments.

3 The State of the Population

The state Γ_n of population G_n contains the deviation $\Delta_i = \Delta_i(G_n)$ from the ith goal and, if appropriate, the direction(s) $\Theta_i = \Theta_i(G_n)$ of the deviation.
3.1 Deviation of the center $\mu(n)$: To measure the deviation of the point $\mu(n)$ from the line $\overline{0, \mu(0)}$, we use a normed perpendicular distance. Let c be a point in $\mathbb{R}_+^K$ and let a be the point on $\overline{0, \mu(0)}$ nearest to c and b the point on $\overline{a, c}$ with the largest distance



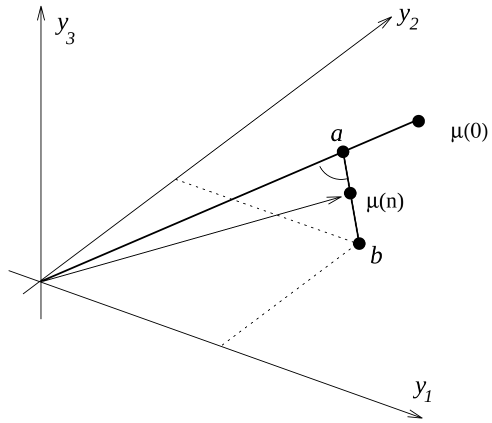


Figure 3: The distance of a point c from $\overline{0, \mu(0)}$ is calculated in the plane determined by c, $\mu(0)$ and 0. Note that the intersections of this plane with the coordinate planes of IR_+^K need not be perpendicular to each other.

to $\overline{0, \mu(0)}$ and still in IR_+^K, cp. Fig. 3. Then b lies on one of the coordinate planes. Put $\delta(c) := \frac{\|c-a\|}{\|b-a\|}$. It follows from simple geometric considerations that $0 \leq \delta(c) \leq 1$ and

$$\delta(c) = 1 - \frac{\| \mu(0) \|^2}{\langle \mu(0), c \rangle} \min_{1 \leq k \leq K} \frac{c_k}{\mu_k(0)},$$

where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product. We put

$$\Delta_1(G_n) := \delta(\mu(n)) = 1 - \frac{\| \mu(0) \|^2}{\langle \mu_k(0), \mu_k(n) \rangle} \min_{1 \leq k \leq K} \frac{\mu_k(n)}{\mu_k(0)} \tag{2}$$

$\Delta_1(G_n)$ gives a quantitative measure for the distance of $\mu(n)$ from $\overline{0, \mu(0)}$. We shall now define the direction of the deviation. Let

$$\Theta_1(G_n) := \left\{ k \mid 1 \leq k \leq K, \frac{\mu_k(n)}{\mu_k(0)} = \min_{1 \leq k' \leq K} \frac{\mu_{k'}(n)}{\mu_{k'}(0)} \right\}$$

be the set of all minimizing coordinates in (2). Then the coordinates of IR_+^K not contained in $\Theta_1(G_n)$ determine the coordinate hyperspace which is first met following the line from a over $\mu(n)$ (at the point b in Figure 3). $\Theta_1(G_n)$ gives the coordinates k', i.e. the objectives $f_{k'}$ in whose direction $\mu(n)$ deviates most from $\overline{0, \mu(0)}$.

3.2 Deviation of the Variation $\sigma(n)$: To measure the violation of goal 2 we compare the total 'span' $\sum_{i=1}^K \sigma_i(n)$ of all values in F_n with those in the starting population. In order to obtain values $\Delta_2(G_n) \in [0, 1]$ we define

$$\Delta_2(G_n) := \left[1 - \frac{\sum_{i=1}^K \sigma_i(n)}{\sum_{i=1}^K \sigma_i(0)} \right]^+$$

Values near 1 indicate that the population tends to loose variation. This is only a very rough indicator in conjunction with the other components of Γ_n it turned out to be sufficient. An analogon to Θ_1 seems not reasonable for goal 2.

3.3 Deviation from Goal 3: To detect separation into subpopulations we compare the rectangle that contains the innermost third of the points $f(x)$, $x \in G_n$ with the enclosing rectangle of F_n. For $x \in G_n$, let $d(x) := 2 * \max_{1\leq i\leq K} \frac{|f_i(x)-\mu_i(n)|}{\sigma_i(n)}$ be a normed distance of $f(x)$ from $\mu(n)$. Note that for $r > 0$, $d(x) \leq r$ iff $f_k(x) \in [\mu_k(n) - r\sigma_k(n)/2, \mu_k(n) + r\sigma_k(n)/2]$ for all $1 \leq k \leq K$, i.e. iff $f(x)$ lies in the enclosing rectangle of F_n scaled by the factor r. $d(x)$ can take on values in $[0, 1]$. Now let

$$\varrho_\beta(G_n) := \min\{\, r > 0 \mid |\{\, x \in G_n \mid d(x) \leq r \,\}| \geq \beta \cdot N\},$$

then $\varrho_\beta(G_n)$ determines the smallest rectangle that contains $\beta \cdot 100\%$ of the points of F_n and is centered at $\mu(n)$. $\varrho_\beta(G_n)$ is a kind of empirical β quantile with $\varrho_1(G_n) = 1$. If the points of F_n are uniformly distributed around $\mu(n)$ then it should approximately hold that $\varrho_\beta(G_n) = \beta^{1/K}$. For $\beta = 1/3$ the deviation is given by the ratio $\varrho_{1/3}(G_n)/(\frac{1}{3})^{1/K}$. In order to obtain values $\Delta_3(G_n) \in [0, 1]$ we define

$$\Delta_3(G_n) := \left[\frac{\varrho_{1/3}(G_n) * 3^{1/K} - 1}{3^{1/K} - 1}\right]^+$$

Large values (near 1) indicate that relatively many points are far away from the center, i.e. that the population tends to drift apart. A definition of a direction $\Theta_3(G_n)$ does not seem reasonable.

3.4 Deviation from Goal 4: Goal 4 requires the lower left corner $\lambda(n)$ of F_n to move along the line $\overline{0, \lambda(0)}$. Hence we may use the distance measure from 3.1 with $\mu(0)$ replaced by $\lambda(0)$ and obtain

$$\Delta_4(G_n) = 1 - \frac{\|\ \lambda(0)\ \|^2}{\langle\lambda(0), \lambda(n)\rangle} \min_{1\leq k\leq K} \frac{\lambda_k(n)}{\lambda_k(0)}.$$

Let

$$\Theta_4(G_n) := \left\{ p(k) \mid p(k) := \frac{\|\ \lambda(0)\ \|^2}{\langle\lambda(0), \lambda(n)\rangle} \frac{\lambda_k(n)}{\lambda_k(0)},\ \ 1 \leq k \leq K \right\}$$

be the set of inverse values $p(k)$ of the relative deviation in the k-th coordinate.

Now the complete (crisp) state of population G_n is given as

$$\Gamma_n = (\Delta_1(G_n), \Delta_2(G_n), \Delta_3(G_n), \Delta_4(G_n), \Theta_1(G_n), \Theta_4(G_n))$$

4 The 'Fuzzy Reduction Step'

As indicated in Fig. 1, we need a decision on what kind of reduction step "$G_n^+ \to G_{n+1}$" is to be used given the state $\Gamma_n = (\Delta_1, \ldots, \Delta_4, \Theta_1, \Theta_4)$ of the population G_n. We use a fuzzy controller, therefore we transform the crisp value components of Γ_n into fuzzy sets. These fuzzy sets assign membership values to the linguistic variables derived from the goals. Then the fuzzy sets are combined into several IF–THEN rules where the THEN

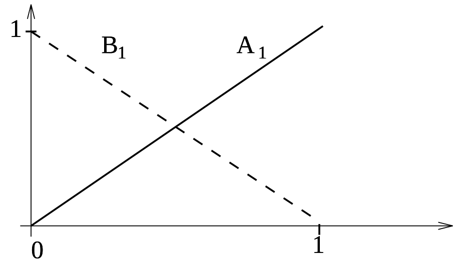


Figure 4: The membership functions used to describe the deviations from the two goals.

IF						
F_n is in state			A_1	B_1	B_1	B_1
AND	A_2	A_2	B_2	B_2	B_2	B_2
AND				A_3	B_3	B_3
AND	A_4	B_4			A_4	B_4
THEN						
use for reduction	V*	V	E	P	V*	V

Table 1: 6 rules for all combinations of the four properties. A blank space means A_i OR B_i

part contains the reduction to be taken. Using the so-called method of activation rates one rule will be activated. Although the idea of the method seems to be complicated, an implementation is very easy and needs only little computational effort during the optimization run.

With respect to goal 1 we can classify the population using the following two linguistic variables A_1 := "deviation of $\mu(n)$ too large" and B_1 := "deviation of $\mu(n)$ tolerable". We use the fuzzy sets of Fig. 4. That means that for a given value $\Delta_1(G_n)$ the population is in state A_1 with a probability given by the solid line (has that membership value) and in state B_1 with the complementary probability given by the dashed line in Fig. 4.

The same fuzzy sets can be used for the variables B_2 := "enough variation" and A_2 := "variation too small" derived from goal 2, for B_3 := "F_n connected" and A_3 := "F_n split" connected to goal 3 and A_4 := "deviation of $\lambda(n)$ too large" and B_4 := "deviation of $\lambda(n)$ tolerable" for goal 4. The combination of the sets related to the four goals leads to 16 possible rules. For each combination one has to fix a method of reduction which is regarded to be the best countermeasure against the undesirable behaviour stated in the IF part. As a natural choice and from our experience we recommend the rules as in Table 1.

We used the following types of reduction:

"E" (Elite): Use the fitness function

$$g(x) = \sum_{k \in \mathcal{K} \setminus \Theta_1(G_n)} f_k(x) \qquad x \in S$$

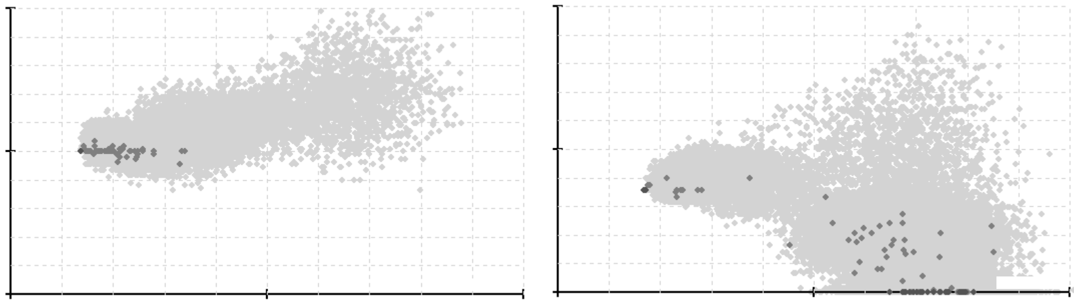


Figure 5: 250 generations constantly using "E" (left figure) and "V" (right figure)

(where $\mathcal{K} := \{1, \ldots, K\}$) and take the best individuals of G_n^+ for the next generation G_{n+1}. Here, the f_k with $k \in \Theta_1(G_n)$ are dropped from the fitness function. Hence the selection pressure during the reduction will concentrate on the coordinates $k \notin \Theta_1(G_n)$ allowing higher values $f_k(x)$ for $k \in \Theta_1(G_n)$.

"P" (Pareto-rank): Take the best individuals of G_n^+ for the next generation G_{n+1} with respect to $g(x) = r(x)$ (cp. (1)). From our obervation this reduction will concentrate the population.

"V" (VEGA): as described in the introduction. This will increase the overall variation in the population.

"V*" (VEGA*): Take the $\frac{p(k)}{\sum_{k'=1}^{K} p(k')}$ best individuals according to f_k, $1 \leq k \leq K$, where the $p(k)$ are obtained from $\Theta_4(G_n)$. This will increase the variation in the population in those dimensions k' where the deviation from $\overline{0, \lambda(0)}$ is large.

Note that each of these methods has several disadvantages if used constantly. The main point is the adaptation of the method to the actual state.

5 Cost benefit analysis for timetables

Railway companies in Europe are presently trying to minimize the waiting times that occur when passengers have to change the trains [4, 5]. To obtain so-called integrated timetables with no waiting times at the important stations the companies have to invest into their networks (tunnels, faster trains, larger stations). We implemented a system 'HiTT' that uses the german intercity network to demonstrate the effect of invested capital onto the minimal total waiting time W of the net. Here W is the sum of the waiting times of all connections in the net weighted by the average number of passengers using the particular connection.
Figure 5 shows the effects of constant reduction methods "E" and "V". Figure 6 shows the result of our adaptive fuzzy-reduction approach.

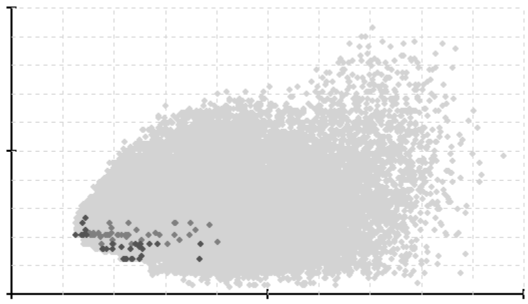


Figure 6: 250 generations using the fuzzy-reduction approach

References

[1] C.M.M.da Fonseca. *Multiobjective Genetic Algorithms with Application to Control Engineering Problems.* PhD thesis, University of Sheffield, September 1995.

[2] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Wesley Publishing Company, 1989.

[3] H. Ishibuchi and T. Murata. Moga: Multi-objective genetic algorithm. In *IEEE International Conference on Evolutionary Computation*, pages 289–294, 1995.

[4] M. Kolonko, K. Nachtigall, and S. Voget. Exponat der universitaet hildesheim auf der cebit 96: Optimierung von integralen taktfahrplaenen mit genetischen algorithmen. *Hildesheimer Informatik Berichte*, 8/96:1–39, Maerz 1996.

[5] M. Lichtenegger. Der integrierte taktfahrplan. *ETR*, 40:171–175, 1991.

[6] J.D. Schaffer and J.J. Grefenstette. Multi-objective learning via genetic algorithms. In *Proceedings of the ninth International Joint Conference on Artificial Intelligence*, pages 593–595, 1985.

[7] P.L. Yu. *Multiple Criteria Decision Making: Concepts Techniques and Extensions.* Plenum Press, New York, 1985.