THE JOURNAL OF COGNITIVE SYSTEMS
VOLUME 01 NUMBER 01 DECEMBER 2016

PROBABILISTIC CONSIDERATIONS UNDERLYING A NOVEL EVOLUTIONARY COMPUTATION

Tahir Cetin Akinci[1], Ozer Ciftcioglu[2], Jelena Dikun[3], Emine Ayaz[1]

[1] Istanbul Technical University, Istanbul, Turkey
[2] Delft University of Technology, Delft, The Netherlands
[3] Lithuanian Maritime Academy, Klaipeda, Lithuania

Penalty function approach used for optimization has a growing interest in the literature due to its effectiveness not only for multiobjective optimization but also for constraint optimization. Although there are several excellent papers on the penalty function approaches, up till now there is no clear method for the systematic selection of penalty parameters per constraint since the topic is quite elusive. The issues being well realized, there are several researches addressing these issues to some extent. However, still the robustness of these methods remains the main issue due to some newly added additional parameters subject to determination. This work endeavors to address this issue and first it makes a systematic analysis. Following the analysis it establishes a probabilistic approach as the issue is entirely in the domain of probability. According to the best knowledge of the authors the approach is unique as to probabilistic treatment of the issue. The approach models the probability density of the random population throughout the generations and based on this, penalty parameters are determined following the probabilistic derivations. The theoretical considerations are substantiated by computer experiments and a demonstrative example is presented showing the salient effectiveness of the approach..

Index Terms— Evolutionary algorithm, multiobjective optimization, constraint optimization, probabilistic modeling.

I. INTRODUCTION

EVOLUTIONARY multiobjective optimization is a popular approach in science and engineering. It is particularly important in cognitive science, because of the decision-making process and ensuing optimization process for action and communication. In this work, constraint optimization is the subject matter, which is to consider as multiobjective optimization due to the method of Penalty function approach. Its use for constraint optimization has a growing interest in the literature, due to its effectiveness not only for multiobjective optimization but also for constraint optimization. Although there are several excellent papers on the penalty function approaches, up till now there is no a clear method for the systematic selection of penalty parameters per constraint, since the topic is quite elusive. The issues of common penalty parameter pertinent to all constraints are well understood. Still the robustness of these methods remains the main issue due to variation of the parameters during the optimization process. The penalty function methods are widely used methods for evolutionary constraint optimization, which differ from each other due to some different strategies. In this respect some examples are static penalty, dynamic penalty, annealing penalty, adaptive penalty, co-evolutionary penalty, death penalty and their associated penalty parameters [1-10]. Strategies that did not require a penalty parameter were proposed in the literature, e.g. [11, 12], while the latter work was later superseded by the penalty function approach [13]. This variety of penalty-function oriented researches is the manifestation of the persisting issue of determining the penalty parameters with respect to each constraint.

In this work a new approach is proposed. Probabilistic considerations underlying the approach are described in detail. The approach is based on the evolutionary probabilistic modeling of the random solutions and the introduction of a probabilistic distance metric. The model is used for effective ranking of genetic population members and thereby yields efficient converging solutions.

The organization of the paper is as follows. In section two, formulation of the general multiobjective optimization problem and non-linear ranking are presented. In section three, important implications of the evolutionary probabilistic approach are described. In section four a demonstrative computer experiment is given. The section is followed by conclusions.

Manuscript received October 10, 2016; accepted November 24, 2016.
Corresponding author: Özer Ciftcioglu (E-mail: o.ciftcioglu@tudelft.nl).
Digital Object Identifier:

II. PROBLEM FORMULATION AND NON-LINEAR RANKING

A. WEIGHTING METHOD

A well-defined method for dealing with the multi-objective optimization is known as *weighting method* [14-16]. In this method each objective is associated with a weighting coefficient and minimizes the weighting sum of the objectives. In this way, the multiple objective functions are transformed into a single objective function. We assume that the weighting coefficients w_i are real numbers such that $0 \le w_i$ for all objectives $i=1,\ldots,k$ so that a weighting problem can be stated as

$$\min \sum_{i=1}^{k} w_i f_i(\boldsymbol{x}); \;\; \boldsymbol{x} \in S \tag{1}$$

In the constraint handling in this work a single objective is involved which is subject to minimization. Therefore the problem can be stated as

$$\begin{aligned} &\min \;\; f(\boldsymbol{x}) \\ &\textit{subject to} \;\; g(\boldsymbol{x}) = [g_1(x), g_2(x), \ldots, g_m(x)]^T \le 0 \end{aligned} \tag{2}$$

We assume that the feasible region is of the form

$$S = \{x \in R^n \mid \; g(\boldsymbol{x}) = [g_1(x), g_2(x), \ldots, g_m(x)]^T \le 0\} \tag{3}$$

If we assume that, the summation of the constraint violations is as another objective subject to minimization the problem formulation becomes a problem of two objective functions subject to minimization. The formulation of the problem in this case becomes

$$\min \;\; w_1 f(\boldsymbol{x}) + w_2 G(\boldsymbol{x}) \tag{4}$$

where

$$G(\boldsymbol{x}) = \sum_{i=1}^{k} \mu_i g_i(\boldsymbol{x}) \tag{5}$$

Hence, the problem definition becomes

$$\begin{aligned} &\min \;\; f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) \\ &S = \{x \in R^n \mid \; g(\boldsymbol{x}) = [g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x})]^T \le 0\} \end{aligned} \tag{6}$$

With this formulation, the weighting method becomes appropriate to employ where $w_1=1$, $w_{2i}=\mu_i$. In (6) the problem formulation becomes two objective functions subject to minimization or alternatively a single objective function with an objective vector subject to minimization. This formulation of the problem is equivalent to a single objective problem with constraints where the constraints are given by the vector $g(\boldsymbol{x})$ which is considered to be a penalty function and the parameters μ_i are the associated penalty parameters. We formulate the multiobjective optimization as two-objective optimization, which can be further treated as single objective optimization with constraints, without deviating from generality. This approach is known to be as *ε-Constraint* method [16, 17]. Among the objective functions one function is selected to be optimized, and, by setting an upper bound to each of them, all the other objective functions are converted into constraints. The problem now has the form

minimize $f_l(x)$

subject to $f_j(x) \le \varepsilon_j$ for all $j=1,2,\ldots,k$, $j \ne l$; $x \in S$; $l \in \{1,\ldots,k\}$

The inequalities can be transformed to equalities by considering $\varepsilon_j=0$ for all $j=1,2,\ldots,k$, $j \ne l$. Based on the above considerations, we assume the problem formulation as a constraint optimization with single objective, so that in a general constrained optimization problem the problem formulation is written as

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{J} \mu_j g_j(\boldsymbol{x}) \tag{7}$$

In (7), we make the following observation. Since in the weighting method the weights are positive, in (7) the penalty parameter μ_j is positive. This implies that in a general case we can surmise that the optimization problem is in the form as depicted in figure 1.

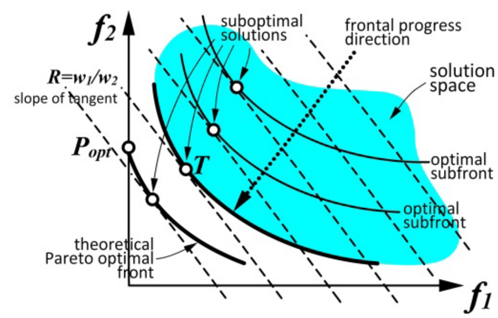


Fig. 1. Approach to the final optimal solution by means of constant penalty parameter R.

Referring to (6) in figure 1 $f_2(\boldsymbol{x})=f(\boldsymbol{x})$ and $f_1(x)=g(\boldsymbol{x})$ denoting violations; also in place of multiple μ_i each of which belong to one constraint, we can consider both a common and constant penalty parameter R which is the slope of the tangent of the Pareto optimal front during the progressive search of the front. In figure 1, the slope of the tangent being negative, the violations are represented as negative quantities so that μ_j and $g_j(x)$ become positive quantities. If we consider that the optimal front is a series of solutions determined by the tangent points of the tangent line and the optimal front, we conclude that the optimal front is simply the envelope of the tangents. This envelope is established as follows.

We assume that a theoretical optimal front compromises the solutions between the objectives $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ where objective $g(\boldsymbol{x})$ admits to be minimally zero. In this case each solution on the optimal front can individually be represented by a line that is tangent to the optimal front at that particular solution. The parametric representation of the tangent is given by

$$\frac{f(\boldsymbol{x})}{t} + \frac{g(\boldsymbol{x})}{P_{opt} - t} = 1 \tag{8}$$

where t is the parameter. In (8), P_{opt} is the optimum solution located at the point $f(\boldsymbol{x}) = P_{opt}$ and $g(\boldsymbol{x}) = 0$. From (8), we write

$$f(\boldsymbol{x}) = \frac{t}{t - P_{opt}(\boldsymbol{x})} g(\boldsymbol{x}) + t \tag{9}$$

where the slope of the tangent is given by

$$r = \frac{t}{t - P_{opt}(\boldsymbol{x})} \tag{10}$$

as a *new* penalty parameter r. The envelope of the tangent is shown in figure 2. The Pareto front is obtained by arranging (10) with respect to t and admitting a single solution for it; namely,

$$t^2 + [g(x) - f(x) - P_{opt}(x)]t + f(x)P_{opt}(x) = 0 \quad (11)$$

$$[f_1(\boldsymbol{x}) - f_2(\boldsymbol{x}) - P_{opt}(\boldsymbol{x})]^2 - 4 f_2(\boldsymbol{x}) P_{opt}(\boldsymbol{x}) = 0 \quad (12)$$

then the optimal front is obtained by equating the discriminant to zero that gives the envelope of the tangent as the optimal front. The new penalty parameter r is zero for $t=0$ and it monotonically increases as t increases. For $t=P_{opt}$ the penalty parameter r goes to infinity.

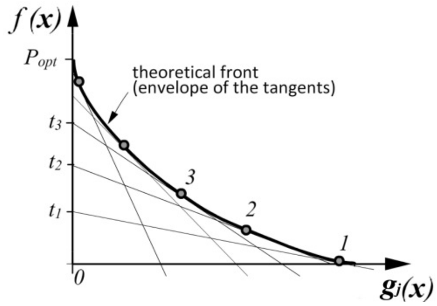


Fig. 2. The envelope of tangent and the *new* penalty parameter r. $r=(P_{opt}-T)/T$ where $T=P_{opt}-t$

If we consider the optimal front for each constraint separately, (10) can be written as

$$r_j = \frac{t}{t - P_{opt}(\boldsymbol{x})} \quad (13)$$

so that (7) becomes

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{J} r_j g_j(\boldsymbol{x}) \quad (14)$$

B. PROBABILISTIC DISTANCE METRIC

In (14) $g_j(\boldsymbol{x})$ at each generation continually is tried for its vanishing during the evolutionary optimization process. This is accomplished by the evolutionary algorithm, giving higher probability of reproduction to population members with small g_j values. Therefore, with respect to the population density of solutions, the probability density of $g_j(\boldsymbol{x})$ is highest about zero violations, and the density gradually diminishes proportional with the degree of violation. Based on the randomly generated population of the evolutionary algorithm, we can model the violations as a random variable, where the violations are independent due to random population formation by the random composition of chromosomes at each generation. The number of violations per unit violation gradually decreases with the degree of violation conforming to the commensurate number of chromosomes created by the elitism and sorting strategy in the genetic algorithm (GA). This probabilistic pattern continues in the same way without change throughout the generations. The probabilistic description of this process can be modeled by the exponential probability density (pdf), because of its memorylessness property, i.e. the form of the density remains the same being independent of the range it models and exponential pdf is a unique density having this property. With this information peculiar to the subject matter of this research, we can confidently apply the exponential probability density function (pdf), which is given by

$$f_\lambda(y) = \lambda e^{-\lambda y} \quad (15)$$

where λ is the decay parameter. If we define

$$y = g_j(x) \quad (16)$$

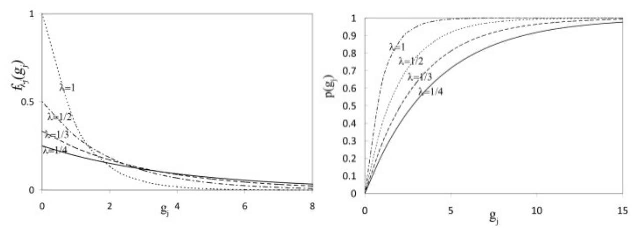
the pdf in (15) becomes

$$f_{g_j}(g_j) = \lambda_j e^{-\lambda_j g_j} \quad (17)$$

The mean value of the exponential pdf function is equal to λ_j^{-1}. During the evolutionary search $g_j(\boldsymbol{x})$ is a general form of violation which applies to any member s of the population although s is not explicitly denoted. However, in explicit form, we can write

$$f_{g_j}(g_{j,s}) = \lambda_j e^{-\lambda_j g_{j,s}} \quad (18)$$

The variation of the exponential pdf for different decay parameters is shown in figure 3a.



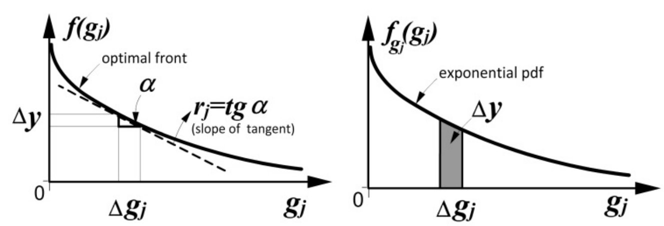
(a) (b)

Fig. 3. Plot of exponential pdf for different decay constants vs j-th violation g_j (a); Plot of $p(g_j)$ vs g_j for various mean values of $p(g_j)$ (b)

The mean value of the violations g_j is the characteristic of the constraint j and it defines the shape of the exponential distribution of the violations representing the decay constant

$$\lambda_j = 1/\bar{g}_j \quad (19)$$

The typical shape of the optimal front in figure 4, and the variation of the exponential distribution is shown together in figure 4, where 4a indicates the optimal front and 3b indicates the exponential probability density.



(a) (b)

Fig. 4. Forming the optimal front as an envelope of a slope by means of the probilistic modelling of rnadom solutions as exponential distribution.

In figure 4b, a small change in violation g_j causes small change in probability density and the probability of violations in this interval is given by

$$\Delta y_f = f_{g_j}(g_j)\Delta g_j \quad (20)$$

From figure 4a we note that a small change in the violation g_j causes a small change in the objective function along the optimal front, and it is given by

$$\Delta y_p = r_j(g_j)\Delta g_j \tag{21}$$

During the search evolution, at each generation the decay constant is newly estimated by the mean of the violations as given by(19), so that λ_j is assumed to be constant from one generation to another. Hence, (17) becomes

$$f_{g_j}(g_j) = \frac{1}{\bar{g}_j} e^{-g_j/\bar{g}_j} \tag{22}$$

In the same way, at each generation r_j is newly estimated, so that r_j is assumed to be constant from one generation to another. Taking infinitesimally small violation intervals, and equating (20) and (21), that is, equating to the objective function change to the probability in the interval dg_j we write

$$r_j(\bar{g}_j)dg_j = \frac{1}{\bar{g}_j} e^{-g_j/\bar{g}_j} dg_j \tag{23}$$

It is to note that by means of this equation above we are relating the objective function space to a probability domain in the form of a transformation. The important implications of this transformation are presented in section 3.

Defining

$$p(g_j) = \frac{1}{g_j}\int_0^{g_j} e^{-\lambda_j g_j}\, dg_j = 1 - e^{-\lambda_j g_j} \tag{24}$$

Integration of (24) from zero to g_j gives

$$\int_0^{g_j} r_j(\bar{g}_j)dg_j = \int_0^{g_j} \frac{1}{\bar{g}_j} e^{-g_j/\bar{g}_j} dg_j \tag{25}$$

That is,

$$r_j(\bar{g}_j)g_j = 1 - e^{-g_j/\bar{g}_j} = p(g_j) \quad , \tag{26}$$

or briefly

$$r_j g_j = p(g_j) \cdot \tag{27}$$

The variation of $p(g_j)$ with g_j is shown in figure 4b.

In (7), μ_j is replaced by Cr_jg_j, namely

$$\mu_j = Cr_j g_j \tag{28}$$

where C is a constant, and the substitution of (28) into (7) with the consideration of (27) yields

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + C\sum_{i=1}^{J} p(g_j) \tag{29}$$

where J denotes the number of constraints; C is a common constant for all constraints. The probability $p(g_j)$ controls the penalty parameter $\mu_i(g_j)$ in (7), which is absorbed in $p(g_j)$. The importance of this transformation, namely from $\mu_i g_j$ to $p(g_j)$ is mainly due to its use for ranking in the probabilistic domain during the genetic optimization process.

In view of (27), r_j is given by

$$r_j = f(g_j) = p_j(g_j)/g_j \tag{30}$$

The variation of the slope r_j versus g_j is plotted in figure 5, where the variation of the slope given by (13) is also plotted.

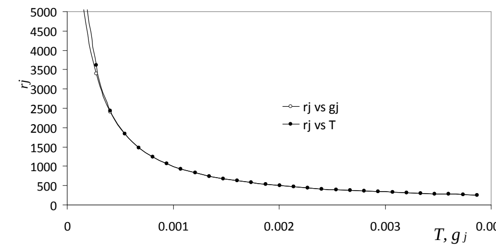


Fig. 5. Illustration of the *new* penalty parameter r as to probabilistic modeling: $r=(1-\exp(-\lambda g))/g$ and as to bi-objective formulation: $r=t/(P_{opt}-t)$

The two slopes, namely one obtained as the tangent, the envelope of which forms the Pareto front, and the other one obtained from a probabilistic model, introduced in this research, coincide satisfactorily, as seen in the figure.

The probability $p(g_j)$ is a *probabilistic distance function* or a *metric* measuring the distance from the zero violation, as it fulfils all the conditions to be a distance measure [18, 19]. The probability density of this distance metric given by (26) is computed by

$$f_p(p) = \frac{f_{g_j}(g_j)}{\left|\frac{dp(g_j)}{dg_j}\right|_{g_j=p^{-1}(p)}} \tag{31}$$

which gives

$$f_p(p) = 1 \tag{32}$$

as uniform pdf. The defined distance metric in the probability domain $p(g_j)$ is used for ranking the chromosomes for effective tournament selection and elitism, in place of remaining in the objective function space. The important implications of this transformation from objective function space to the probability domain are given in the next section.

III. Implications of the Probabilistic Distance Metric

A. Stiffness Handling

The stiffness is defined as the large numerical difference among several constraints subject to minimization. If there is stiffness among the constraints, the summation in (6) is dominated by the constraints, the pdfs of which have small decay constants. However, by using the probabilistic distance measure varying between zero and unity, this drawback is eliminated. The treatment is illustrated in figure 6, where the probabilistic distances for the constraints random variables $g_{\lambda 1j}$ and $g_{\lambda 2j}$ are the same, and the distance is between zero and unity. By giving the same priority or rank in the tournament selection process for the constraints $g_{\lambda 1j}$ and $g_{\lambda 2j}$ we consider purely their associated probabilities based on the probabilistic model without imposing any bias about the nature of the constraints.

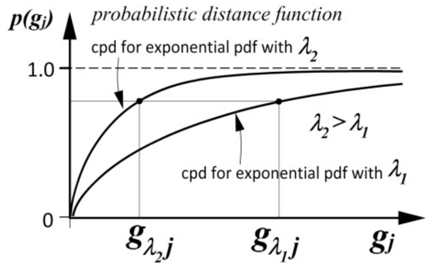


Fig. 6. Illustration of the stiffness handling

B. IMPARTIAL ELITISM SELECTION

During elitism we consider the population from the preceding generation. Therefore, below first we compute the probability of having smaller constraint violation. If we consider two exponential probability density functions with the random variables X_1 and X_2 and the associated decay parameters λ_1 and λ_2 respectively, the probability $P(X_2<X_1)$ is computed as follows. The probability of $g_2 \le X_1$, namely $P(g_2 \le X_1)$ is given by

$$P(g_2 \le X_1) = \int_{g_2}^{\infty} \lambda_1 e^{-\lambda_1 g_1}\, dg_1 = e^{-\lambda_1 g_2} \tag{33}$$

so that

$$P(X_2 < X_1) = P(g_2 \le X_1)\,P(g_2 \le X_2)$$
$$= \int_0^{\infty} e^{-\lambda_1 g_2} \lambda_2 e^{-\lambda_2 g_2}\, dg_2 = \lambda_2 \int_0^{\infty} e^{-(\lambda_1+\lambda_2) g_2}\, dg_2 \tag{34}$$

which gives

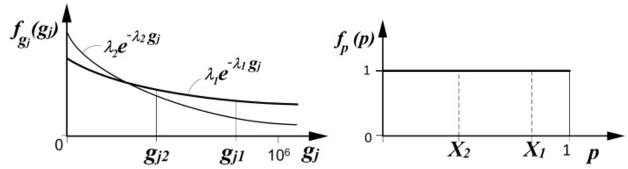
$$P(X_2 < X_1) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \tag{35}$$

Let us carry out the same calculations with respect to the random variable P, the pdf of which is given by (32).

$$P(g_2 \le X_1) = \int_{g_2}^{1} dg_1 = 1 - g_2 \tag{36}$$

$$P(X_2 < X_1) = P(g_2 \le X_1)\,P(g_2 \le X_2)$$
$$= \int_0^1 (1 - g_2) dg_2 = \frac{1}{2} \tag{37}$$

This result shows that, irrespective to the decay constants λ_1 and λ_2 of two exponential *pdf*s, the probability $P(X_2 \le X_1)$ is always 0.5, which means ranking the random solutions during the genetic search, probabilistic distance function $p(g_j)$ is fully impartial with respect to the decay constants. This means, although the decay constants vary and they are updated each iteration, this is not reflected to the elitism. With other words, the procedure is not biased by apparently less favorable population of the latest generation due to higher probabilistic distances caused by higher decay constant. Instead, the preceding generation and the following generation are treated in perspective without bias, eliminating the decay constant factor in the computation. The exponential pdf $f_{gj}(g_j)$ in (17) and uniform pdf in (32) are sketched in figures 7a and 7b, where the random variables X_1, X_2 and two corresponding violations g_{j1}, g_{j2} are also shown. It should be pointed out that the uniformity of the uniform pdf is not affected even if the modelling error in the surmised exponential pdf model exists.



(a) (b)

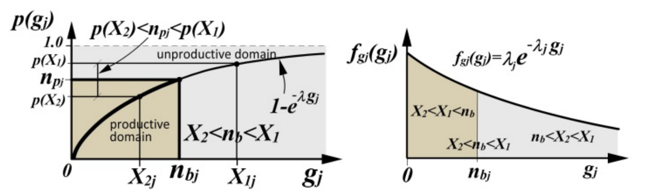
Fig. 7. Probability density $f_{gj}(g_j)$ vs g_j (a) Probability density $f_P(p)$ vs p (b)

Two important aspects in this work, beyond the basic elitism procedure, are the followings.

1. During the elitism, the combination of the present and the preceding populations is accomplished using their respective decay constants (λ). In this case the situation is depicted in figure 6, where the same rank is assigned to different violations depicted $g\lambda_{2j}$ as present violation and $g\lambda_{21j}$ as the preceding violation. By doing so, diversity in the genetic population is maintained although it slows down the convergence to some extent. However, the gain is reducing the risk of premature convergence.
2. Solutions during tournament selection will be evaluated depending on the condition given by

$$\sum_{j=1}^{J} p(g_j) < n_{pj}\, J \tag{38}$$

where J is equal to the number of constraints, and n_{pj} denotes a probability threshold, above which a solution is deemed *unproductive* among the infeasible solutions, and below which a solution is deemed *productive*. It has a counterpart in the objective space denoted by n_{bj}.



(a) (b)

Fig. 8. Illustration of the threshold assessment for the tournament selection in both NS and NR procedures.

In case one solution fulfills (38), while the other one does not, then the solution in the productive domain wins the tournament over the other one, without considering rank or crowding information. This case is shown in the same figure, where the violation in the productive domain is denoted by X_{2j} and its counterpart is X_{1j}. The counterpart of (38) in the objective space, and is given by

$$\overline{g}_T = n_{b_j} \sum_{j=1}^{J} \bar{g} = \sum_{j=1}^{J} \frac{n_{b_j}}{\lambda_j} \tag{39}$$

Referring to figure 8b, the probability P_j of the event relevant to the case described above is given by X_2<nb<X1, and

$$P_j = P(g_j) = P(X1_j)P(X2_j) = e^{-\lambda_j n_{bj}} - e^{-2\lambda_j n_{bj}} \tag{40}$$

However, since λ_j is evolving from generation to generation, g_T is not constant. In contrast with this, in the probabilistic non-linear ranking domain, the location of maximum probability of the event that two solutions appear on either side of the threshold n_{bj} is always at n_p=0.5, irrespective of λ_j. The case for the probabilistic raking domain is illustrated in figure 9**,** where the variation of $p(g_j)$ with respect to n_{bj} is illustrated also for the productive and unproductive domains.

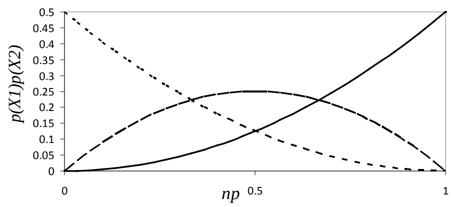


Fig. 9. Plot of the probabilities for different conditions that can arise during binary tournament selection

The case for the objective space is illustrated in figure 10, where the maximum occurs for $n_{bj}=ln2/\lambda_j$, which is the median of the exponential probability density shown in figure 8b. The single plot for each of the three possible conditions, during a binary tournament seen in figure 9, depending on the occurrence of solutions in productive or non-productive domains, correspond to a family of plots with respect to the parameter λ_j in figures 10-11.

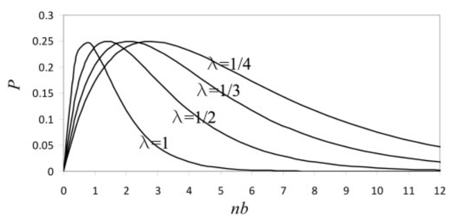


Fig. 10. Plot of the probability that two solutions occur on different sides of the threshold n_{bj} for λ=1, 1/2, 1/3, 1/4. The respective maximum occurs at n_b=0.693/λ

Explicitly, for $n_{bj}=ln2/\lambda_j$, its counterpart in terms of the probabilistic ranking domain is n_{pj}=0.5. Thus, the constant probabilistic distance measure provides an adaptive threshold for productive chromosomes throughout the generations, in any scale permitted by the machine or genotype precision. By means of this particular tournament selection procedure, the detrimental effect on the average violation by the stiff constraints, that is, by the members with high violations, is prevented; namely, during two consecutive generations the progressive diminishing of the average is augmented against for the contingent average increase that may occur especially during the advanced stages of the convergence. The smaller total mean of the constraint violations implies improved convergence to the optimum.

For the other cases, namely

$X_2<X_1$, $X_1,X_2 < n_b$

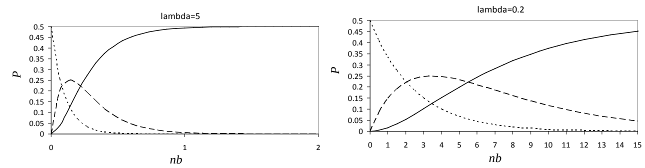
$$P(X_2 < X_1 < n_b) = 0.5(1-e^{-2\lambda_j n_b}) - e^{-\lambda_j n_b}(1-e^{-\lambda_j n_b}) \qquad (41)$$

and for $X_2<X_1$, $n_b<X1,X2$

$$P(n_b < X_2 < X_1) = 0.5e^{-2\lambda n_b} \qquad (42)$$

The variations of the different probabilities in (41)-(42) are plotted together in figure 11. It is to note that for any value of n_b, the summation of the probabilities is equal to 0.5, which conforms to (28) for $\lambda_1=\lambda_2$.

Figure 11 is the counterpart of figure 9 in the objective function space.



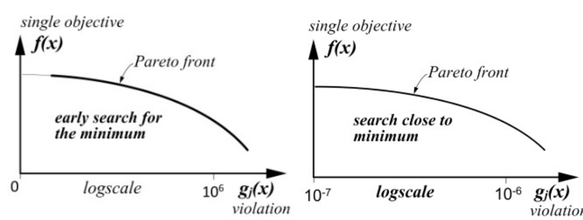
(a) (b)

Fig. 11. Plot of the probabilities for different conditions that can arise during binary tournament selection for λ=5 (a); λ=0.2 (b)

It is seen from figure 11 that the shape of the probability functions depends on λ_j, whereas in the probabilistic domain in figure 9, the shape remains constant, i.e. independent of λ_j.

C. ZOOMING FOR ROBUST RANKING

Zooming for robust ranking is accomplished by accurate ranking the favourable solutions between zero and unity as probabilistic distances, even though the actual constraint values may be close to the utopic optimal point as much as allowed by the computer precision that may be at the range of 10^{-10} or below. Illustration of the Pareto front at the early stage of the genetic search is given in figure 12a. Illustration of the Pareto front at the last stage of the genetic search is given in figure 12b.



(a) (b)

Fig. 12. Illustration of the formation of the Pareto front at the early stage (a) and last stage of the genetic search (b)

Considering figure 12b, the probabilistic distance to the minimum is illustrated as a typical example in figure 13a by the shaded area where computation of the shaded area is very precarious at the tournament selection process. This is due to the issue of both exact parameterization of the exponential pdf in the existing range and the finite machine precision. This issue is prevented in figure 13b by taking simply $p(g_j)$ as the probability distance to the minimum. The marked areas in figure 13a and 13b are the same, and they are equal to $p(g_j)$.

The marked area in figure 13a, is represented in figure 13b by the probabilistic distance function $p(g_j)$ which varies between zero and unity. This means if the penalty function to be minimized can be close to the optimal point in a micro scale, say in the range of 10^{-10}, the minimization process i.e., tournament selection and ranking of the random solutions takes place in a macro scale in the probabilistic space as shown in figure 13b. This treatment is equivalent to applying a matching 'magnifying glass' to the space formed by actual objective function and the constraint functions, in order to carry out the convergence process without being affected by any scale of convergence present in this very space. The Pareto front at this micro scale is illustrated in figure 12.

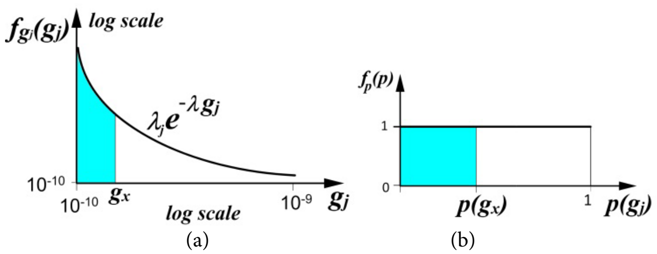


Fig. 13. An example illustration of the probability density of the constraint violations in the objective functions space (a) and the probabilistic distance space (b).

D. FAST AND ROBUST CONVERGENCE

With the probabilistic distance for nonlinear ranking we obtain an optimal step for convergence at each generation. To see this, from (27)

$$r_j = \frac{p(g_j)}{g_j} = \frac{1-\exp\left[-\lambda_j g_j\right]}{g_j} \approx \frac{1-\exp\left[-\frac{g_j}{\bar{g}_j}\right]}{g_j} \tag{43}$$

In the limiting case, i.e., convergence to the minimum, r_j becomes

$$\lim_{g_j \to 0} \; r_j = \frac{p(g_j)}{g_j} = \lim_{g_j \to 0} \lambda_j e^{-\lambda_j g_j} = \lambda_j \to \infty \tag{44}$$

As it is seen, the genetic search algorithm is extraordinarily stable, that is, the convergence is due, and due to monotonic increase of the slope r_j, the convergence is fast.

IV. COMPUTER EXPERIMENT

Computer experiments have been carried out using a standard optimization problem from the literature. To demonstrate the robust, fast and accurate computations the course of the convergence are given in detail.

The following problem is due to Himmelblau [20]. given by

$$f(\mathbf{x}) = 5.3578547x_3^2 + 0.8356891x_1x_5 + 37.293239x_1 - 40792.141 \tag{45}$$

where the ranges for the independent variables are given by

$$78 < x_1 < 102;\ 33 < x_2 < 45;\ \ 27 < x_i < 45 \ \ (i = 3,4,5) \tag{46}$$

subject to:

$$\begin{aligned}
&g_1(\mathbf{x}) = 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - \\
&0.0022053x_3x_5 - 92 \le 0 \\
&g_2(\mathbf{x}) = -85.334407 - 0.0056858x_2x_5 - 0.0006262x_1x_4 + \\
&0.0022053x_3x_5 \le 0 \\
&g_3(\mathbf{x}) = 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + \\
&0.0021813x_3^2 - 110 \le 0 \\
&g_4(\mathbf{x}) = -80.51249 - 0.0071317x_2x_5 - 0.0029955x_1x_2 - \\
&0.0021813x_3^2 + 90 \le 0 \\
&g_5(\mathbf{x}) = 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + \\
&0.0019085x_3x_4 - 25 \le 0 \\
&g_6(\mathbf{x}) = -9.300961 - 0.0047026x_3x_5 - 0.0012547x_1x_3 - \\
&0.0019085x_3x_4 + 20 \le 0
\end{aligned} \tag{47}$$

The problem consists of a single objective with 6 constraints, subject to minimization. The best known optimum is located at

$f(\mathbf{x}^*)=-30665.53867178332$

and the corresponding best variable values are

$x_1^*=78$; $x_2^*=33$; $x_3^*=29.9952560256815985$; $x_4^*=45$; $x_5^*=36.7758129057882073$.

The algorithm is executed with the following settings: population size=200; amount of generations=60; C=100000; crossover probability=0.9; mutation probability=0.05. The results are shown in figure 14-16 using a logarithmic scale for the horizontal axis, which shows the total violation G. It is noted that a single iteration of the algorithm consists of five generations.

After 5 iterations the population is seen in figure 12, where the best feasible solution is

$f(\mathbf{x})=-30569.5213239566$.

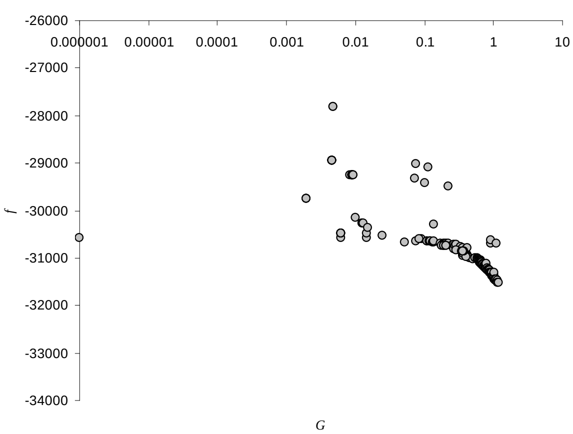


Fig. 14. Population after the 5-th iteration

The independent variables of this solution take:

$x_1=78.0265736760284$; $x_2=33.658770910086$; $x_3=30.470062623374$; $x_4=44.7895003744468$; $x_5=35.8616204529277$.

After 10 iterations the population is seen in figure 14, where the best feasible solution is found to be

$f(\mathbf{x})$=-30653.7876324169.

The independent variables of this solution take:

x_1=78.0261567922629; x_2=33.0152143977847; x_3=30.0228958188968; x_4=44.8232338738051; x_5=36.7924311001587.

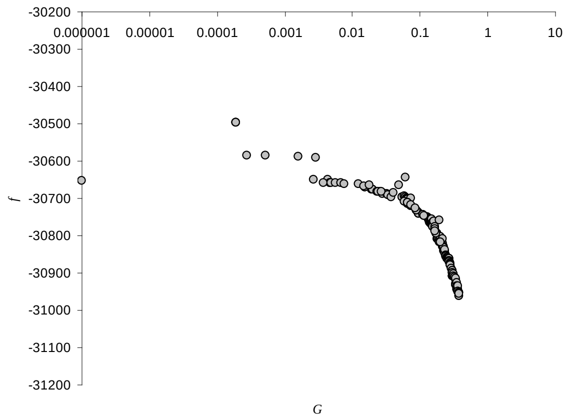


Fig. 15. Population after the 10-th iteration

After 30 iterations the population is seen in figure 15, where the best feasible solution is found to be

$f(\mathbf{x})$= -30665.4759429232.

The independent variables of this solution take:

x_1=78.0000867626641; x_2=33.000032984143; x_3=29.9955726882451; x_4=45; x_5=36.7751232308258.

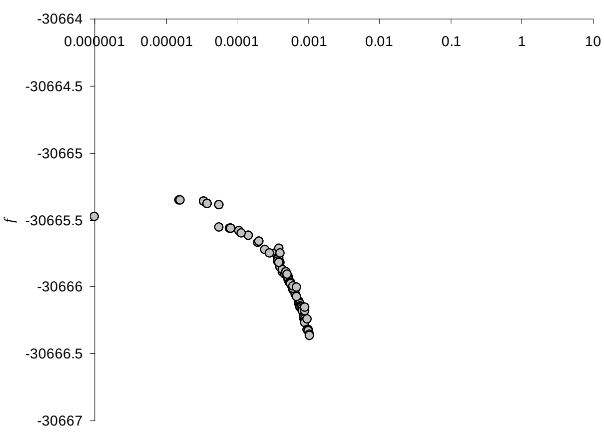


Fig. 16. Population after the 30-th iteration

It is noted that the process will continue to improve the solution more and more as the search continues, i.e. the population converges at the optimal solution, demonstrating the robustness of the approach. Namely after 60 iterations the population is seen in figure 16, where the best feasible solution is found to be

$f(\mathbf{x})$= -30665.5386683921.

The independent variables of this solution take

x_1=78.0000000039558; x_2=33.0000000083502; x_3=29.9952560418378; x_4=45; x_5=36.7758128740195.

V. Conclusions

Probabilistic considerations underlying a novel evolutionary computation are presented. In this work, multi-objective optimization is considered in the form of constraint optimization, the case conventionally being described in the literature, selecting appropriate penalty function parameters. However, since these parameters vary during the search process the determination of these parameters is very elusive and remained an issue to treat for researches. In contrast to this, in this work, a probabilistic model is introduced, by means of which the penalty parameters are embedded in the model, and they are inherently tuned, as the model is adaptively modified throughout the generations. The probabilistic model also has several favorable implications, which are treated in this research. These are stiffness handling, impartial elitism, zooming for robust ranking, as well as fast and robust convergence. The theory presented in this work is exemplified by an optimization problem for demonstration of the general effectiveness resulting from this analytical treatment of the constraint optimization methodology. However the method is not restricted to constraint optimization, but suitable for multi-objective optimization in general. The reported results include not only the final outcomes but also the progress of the convergence throughout the optimization process conforming exactly to the theoretical considerations presented.

References

[1] A. Homaifar, S. H. V. Lai, and X. Qi, "Constrained optimization via genetic algorithms," *Simulation,* vol. 9, p. 19, 1994.

[2] Z. Michalewicz and M. Schoenauer, "Evolutuionary algorithms for constrained parameter optimization problems," *Evolutionary Computation,* vol. 4, p. 33, 1996.

[3] Z. Michalewicz, *Genetic Algorithms+ Data Structures=Evolution Programs*. Berlin: Springer, 1999.

[4] S. Venkatraman and G. G. Yen, "A Generic Framework for Constrained Optimization Using Genetic Algorithms," *IEEE Trans. Evolutionary Computation,* vol. 9, p. 12, 2005.

[5] Z. Huang and H. Tian, "A genetic algorithm with constrained sorting method for constrained optimization problems," presented at the IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. , Shanghai 2009.

[6] A. C. A. Coello, "Use of a self adaptive penalty approach for engineering optimization problems," *Computers in Industry,* vol. 41, pp. 113–127, 2000.

[7] S. B. Hamida and M. Schoenauer, "An adaptive algorithm for constraint optimization problems," in *Lecture Notes on Computer Science PPSN 2000*. vol. 1917, K. Deb, Ed., ed Heidelberg: Springer, 2000, pp. 529-538.

[8] S. B. Hamida and M. Schoenauer, "ASHEA: New results using adaptive segregtional constraint handling," presented at the IEEE Congress on Evolutionary Computation (CEC 2002), Honolulu, Hawaii, 2002.

[9] F. Hoffmeister and J. Sprave, "Problem-independent handling of constraints by use of metric penalty functions," presented at the Fifth Annual conference on evolutionary programming EP 1996, San Diego, CA, 1996.

[10] H. Sarimveis and A. Nikolakopoulos, "A line up evolutionary algorithm for solving non-linear constraint optimization problems," *Computers & Operations Research* vol. 32, pp. 1499-1514, 2005.

[11] K. Deb, "An efficient constraint handling method for genetic algorithms.," *Computer Methods in Applied Mechanics and Engineering,* vol. 186, p. 28, 2000.

[12] T. Ray, H. K. Singh, A. Isaacs, and W. Smith, "Infeasibility driven evolutionary algorithm for constrained optimization," in *Constraint-Handling in Evolutionary Optimization*, ed Berlin: Springer, 2009, pp. 145-165.

[13] K. Deb and R. Datta, "A fast and accurate solution of constrained optimization problems using a hybrid bi-objective and penalty function approach," presented at the Evolutionary Computation (CEC), 2010 IEEE Congress on, Barcelona 2010.

[14] S. Gass and T. Saaty, "The computational algorithm for the parametric objective function," *Naval Research Logistics Quarterly,* vol. 2, p. 7, 1955.

[15] L. Zadeh, "Non-scalar-valued performance criteria," *IEEE Trans. Automatic Control,* vol. 8, p. 2, 1963.

[16] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston: Kluwer Academic, 1999.

[17] Y. Y. Haimes, L. S. Lasdon, and D. A. Wismer, "On a bicriterion formulation of the problems of integrated system identification and system optimization," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 1, p. 2, 1971.

[18] G. Bachman and L. Narici, *Functional Analysis*. New York: Dover, 2000.

[19] J. T. Oden and L. F. Demkowicz, *Applied Functional Analysis*: CRC Press, 1996.

[20] D. M. Himmelblau, *Applied Nonlinear Programming*: Mc-Graw-Hill, USA, 1972.

Özer Ciftcioglu was born in Turkey. He received the PhD degree in Electrical Engineering from Istanbul Technical University (ITU) in 1976. He became associate professor in 1982 and full professor in 1988 at the same university.

Since 2009 he has been working at TU Delft in the Netherlands as Emeritus professor. He had been working at the same university from 1998 to 2009. From 1967 to 1998 he was affiliated with ITU as an academic member. His current research interests are Signal Processing, Computational Intelligence, and Computational Cognition.

Prof. Ciftcioglu is a senior member of IEEE since 2001.

Tehir Cetin Akinci received the BSc degree in Electrical Engineering. His master's and PhD degrees, Institute of Pure and Applied Science from Marmara University, Istanbul, Turkey. He is an associate professor in the Department of Electrical Engineering at Istanbul Technical University (ITU), Istanbul, Turkey. His research interests are signal processing, data mining, intelligent systems, ferroresonance phenomenon, artificial neural networks, and renewable energy sources.

Jelena Dikun was born in Klaipeda, Lithuania. She received P.B. from Klaipeda Business and Technology College in Electrical Engineering Faculty (2005-2009) and BSc. from Klaipeda University in Electrical Engineering Faculty (2009 – 2011). She visited as a trainee Istanbul Technical University between October 21 and November 22, 2012 that provided by projects JUREIVIS and “Lithuanian Maritime Sectors' Technologies and Environmental Research Development”. Her areas of interest are the study of electric and magnetic fields.

Emine Ayaz received the BSc, MSc and PhD degrees from the Istanbul Technical University (ITU), Electrical Engineering Department, in 1993, 1997 and 2002 respectively. In 1999, she joined to the University of Tennessee, Nuclear Engineering Department and Maintenance & Reliability Centre to do research on accelerated aging studies of the electric motors. She is currently associate professor in Electrical Engineering Department of ITU. Her research interests are signal processing, soft computing, and condition monitoring in electric power systems.