THE JOURNAL OF COGNITIVE SYSTEMS
VOLUME 01 NUMBER 01 DECEMBER 2016

PROBABILISTIC SORTING FOR EFFECTIVE ELITISM IN MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS

S. Serhat Seker[1], Michael S. Bittermann[2], Ramazan Caglar[1], Rituparna Datta[3]

[1] Istanbul Technical University, Istanbul, Turkey
[2] Maltepe University, Istanbul, Turkey
[3] Graduate School of Knowledge Service Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, 305-701, South Korea

One of the essential points in the evolutionary algorithms is the rank determination for the genetic population members. In this respect a new approach is presented, which is a probabilistic sorting for effective elitism and ensuing improved and robust convergence. This is achieved by an adaptive probabilistic model representing the commensurate probability density of the random solutions throughout the generations that it yields a probabilistic distance measure which is nonlinear with respect to the range of solutions as to their location in the objectives space. The implementation of the theoretical results leads an effective evolutionary optimization algorithm accomplished in two stages. In the first stage linear non-dominated sorting, tournament selection and elitism is carried out in objective space. In the second stage the same is executed in a transformed objective space, where probabilistic distance measure for ranking prevails. The effectiveness of the method is exemplified by a demonstrative computer experiment. The problem treated is selected from the existing literature for comparison, while the experiment carried out and reported here demonstrates the marked performance of the approach. The experiment complies with the theoretical foundations, so that the robust and fast convergence with precision as well as with accuracy is accomplished throughout the search up to 10-10 range or beyond, limited exclusively by machine precision.

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

I. INTRODUCTION

COMPUTATIONAL cognition makes use of the evolutionary optimization algorithms due to the decision-making process in the cognition. This is especially important in the action and communication stage of cognition. This work describes a research, which provides an effective method to enhance the effectiveness of evolutionary optimization algorithms, and consequently improve the cognition process. Evolutionary algorithms are powerful heuristic computations for multiobjective optimization problems. Their various forms of utilization are ubiquitous and they are reported regularly in the literature, e.g. [1, 2]. Some text book are available e.g. [3-5] that one can approach to master the topic. During the last decades evolutionary algorithms received growing interest, since they proved to be important tools for optimization. Added to that, they also proved to be effective in constraint optimization problem solving as the modern technological application areas imposes limitations on the solutions. The conventional constrained optimization methods generally use methods based on various penalty functions. Penalty function methods are generic but care has to be exercised to use the penalty parameter in a measured way to keep the balance between the constraints and the objective to avoid false optima and infeasible solutions. A strategy that does not use penalty parameter in evolutionary constrained optimization was proposed by Deb in 2000 [6, 7]. Although the penalty parameter can be kept constant during the search process, a better approach is to use a variable penalty parameter, which is adapted to the progress of the convergence, providing an effective approach to the optimum in the decision variable space. In this respect Coello proposed a self-adaptive penalty approach [8]. By doing so, also the evolutionary concept is clearly demonstrated. Conventionally in the penalty function approach, the constrained optimization problem is a search of the best compromises of the objective value and constraint satisfaction. Due to this construction, net result is the unsatisfactory convergence properties which are deemed to be repaired by some additional methods borrowed from classical optimization methods which are collectively addressed as 'local search' methods. One of the essential components in evolutionary algorithms is the rank determination for the individuals. In this work, this issue is addressed by means of probabilistic distance measure which is used for probabilistic sorting and effective elitism by a nonlinear ranking. The method provides a kind of 'mathematical lens,' so that at any stage of convergence the level of rank resolution remains the same that it leads systematic, smooth convergence to the optimum without recourse to additional methods which are collectively regarded as 'local search' methods.

The present work addresses the conversion of a single objective constrained optimization problem into a multiobjective, unconstrained optimization together with a penalty function. In this form, it is a bi-objective optimization problem. Each of the constraints has its own penalty

Manuscript received September 9, 2016; accepted December 7, 2016.
Corresponding author: Michael S. Bittermann (E-mail: michaelsbittermann@maltepe.edu.tr).
Digital Object Identifier:

parameter. For each constraint a probabilistic model of the random solutions is used to derive a nonlinear distance measure. This measure is used for the genetic algorithm, to rank the population members for efficient, i.e. fast convergence. In this form it is a constraint optimization problem. This is the new approach proposed in this paper, for robust and stable solutions. The method is implemented by a computer program developed for this research working based on non-dominated sorting (NS) and non-linear ranking (NR).

The organization of the paper is as follows. In section two, formulation of general multiobjective optimization problem as constrained single objective problem and probabilistic constraint handling is presented. In section three, probabilistic modeling for nonlinear ranking is given. In section four the probabilistic nonlinear ranking for elitism is revealed. The important implications of the probabilistic modeling are highlighted in section five. In section six a demonstrative computer experiment is given and the section is followed by discussion and conclusions.

II. OPTIMIZATION METHOD FOR MULTI-OBJECTIVITY

Weighting method is a powerful instrument for the multi-objective optimization. Its formulation in this work is adapted according to the works reported in the literature [9-11]. The weighting method deals with the weighted summation of the objective functions. Each function is associated with a weighting coefficient and weighting sum of the objectives is minimized. Thus, the multiple objective functions are expressed via a single objective function. The weighting coefficients w_i are real numbers such that $0 \leq w_i$ for all objectives $i=1,\ldots,k$ so that a weighting problem can be stated as

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

In the constraint handling a single objective is used and is subject to minimization. It can be stated that

$$\min \quad f(\boldsymbol{x}) \;\; \textit{subject to} \;\; \boldsymbol{g}(\boldsymbol{x}) = [g_1(x), g_2(x), \ldots, g_m(x)]^T \leq \tag{2}$$

We assume that the form of the feasible region is given by

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

We consider that the summation of the constraint violations is another objective 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

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

Therefore, the problem definition becomes as below.

$$\min \;\; f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) = f(\boldsymbol{x}) + G(\boldsymbol{x}) \tag{6}$$

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

where $w_1=1$, $w_{2i}=\mu_i$. With this, the problem is equivalent to a single objective problem, where the objective is denoted by $f(\boldsymbol{x})$ and the constraints denoted by $g_j(\boldsymbol{x})$. The method known as *ε-Constraint* method is such an approach [11, 12]. In this method one of the objective functions is selected to be optimized, while all the other objective functions are converted into constraints. This is done by setting an upper bound to each of them. The problem to be solved is now of the form

$$\min \;\; f_1(\boldsymbol{x}); \text{ subject to } f_j(\boldsymbol{x}) \leq \varepsilon_j \quad \text{for all j=1,2,.,k, j} \neq \text{l; x} \in \text{S} \tag{7}$$

With the above considerations we minimize $f_l(\boldsymbol{x})$; subject to $f_j(x) \leq \varepsilon_j$ for all $j=1,2,\ldots,k$, $j \neq l$; $x \in S$
where $l \in \{1,\ldots,k\}$. Naturally, inequalities can be converted to equalities by taking $\varepsilon_j=0$ for all $j=1,2,\ldots,k$, $j \neq l$.
In the present case, the minimization of the function in (6) takes the form

$$\min \;\; P(\boldsymbol{x}, \boldsymbol{R}) = f(\boldsymbol{x}) + \sum_{i=1}^{J} R_j g_j(\boldsymbol{x}) \tag{8}$$

where J is the number of constraints; function $g_j(\boldsymbol{x})$ is considered to be a *penalty function* and the parameters R_j are the associated *penalty parameters*. The determination of the penalty parameters is an issue and although this issue addressed in the literature [6], the issue still persists and is subject to improvements. In this work this issue is addressed by a probabilistic approach which underlies also the probabilistic sorting for effective elitism, subject matter of this work.

III. NONLINEAR RANKING WITH PROBABILISTIC CONSIDERATIONS

In general a constrained optimization (8) is written in the form

$$\min \;\; P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{J} \mu_j g_j(\boldsymbol{x}) \tag{9}$$

where $f(\boldsymbol{x})$ denotes the single objective function to be minimized; $g(\boldsymbol{x})$ is the violation of the g_i-th constraint, namely penalty function, μ_i is the associated parameter of the penalty function given by

$$\mu_j(g_j) = C\, r_j(g_j) \tag{10}$$

In (10), r_j is a new penalty parameter; C is a constant common for all constraints. As $g_j(\boldsymbol{x})$ is at each generation continually tried to be vanishing during the evolutionary minimization process, with respect to the population density of solutions, the probability density of $g_j(\boldsymbol{x})$ is highest about zero violations, and its value gradually diminishes proportional with the degree of violation. In 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. 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, while the 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 pdf, which is given by

$$f_\lambda(y) = \lambda e^{-\lambda y} \tag{11}$$

where λ is the decay parameter. If we define

$$y = g_j(x) \tag{12}$$

then the pdf in (11) becomes

$$f_{g_j}(g_j) = \lambda_j e^{-\lambda_j g_j} \tag{13}$$

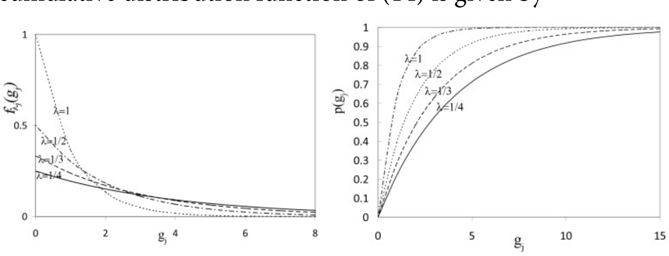
The mean value of the exponential pdf function is equal to λ_j^{-1}. During the evolutionary search $g_i(\boldsymbol{x})$ is a general form of violation, which applies to any member s of the population, and therefore, in explicit form, we can write

$$f_{g_j}(g_{j,s}) = \lambda_j e^{-\lambda_j g_{j,s}} \tag{14}$$

where s denotes a population member. We can characterize the exponential pdf function according to the constraint j simply by equating the mean value of the violations g_j to the mean of the exponential pdf, namely

$$\lambda_j = 1/\bar{g}_j \tag{15}$$

It is to note that the mean of the exponential probability density of g_j is equivalent to the mean of a uniform probability density applied to the violations g_j. Therefore the mean of the exponential density function is estimated by taking the mean of the violations which are from a uniform probability density and they are independent. Variation of the exponential pdf for different decay parameters is shown in figure 5a. The cumulative distribution function of (14) is given by



(a) (b)

Fig. 1. Plot of exponential pdf for different decay constants vs j-th violation g_j (a); $p(g_j)$ vs g_j (b)

$$p(g_j) = \frac{1}{\bar{g}_j}\int_0^{g_j} e^{-\frac{g_j}{\bar{g}_j}} dg_j = 1 - e^{-\frac{g_j}{\bar{g}_j}} \tag{16}$$

If we take $p(g_j)$ as a new random variable, the probability density $f_p(p)$ of the new random variable p is given by[13]

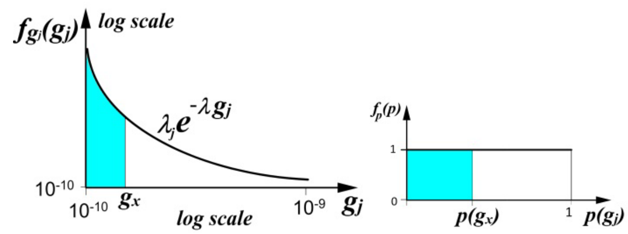
$$f_p(p) = \frac{f_{g_j}(g_j)}{\left|\frac{dH(g_j)}{dg_j}\right|_{g_j = H^{-1}(p)}} \tag{17}$$

that gives

$$f_p(p) = 1 \tag{18}$$

which is a uniform probability density shown in figure 2b together with the exponential distribution in figure 2a. In this figure the marked areas are equal having important implication in nonlinear ranking and elitism. The probability $p(g_j)$ measures the magnitude or effectiveness of a violation, so that it can be considered as a *probabilistic distance function* or a *metric* measuring the distance from the zero violation fulfilling all the conditions to be a distance measure [14, 15]. Substitution of (10) into (9) yields

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + C\sum_{i=1}^{J} r_j(g_j) g_j(\boldsymbol{x}) \tag{19}$$



(a) (b)

Fig. 2. Pdf of the constraint violations in the objective functions space (a); in the probabilistic distance space (b).

where the constant C is called as *convergence parameter* as it is related to the convergence properties of the search. The new penalty parameter r_j which is a function of g_j, in general. In (19), $r_j(g_j)g_j$ is replaced by $p(g_j)$, in the form

$$r_j(g_j) g_j = p_j(g_j) \tag{20}$$

so that (19) becomes

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + C\sum_{i=1}^{J} p_j(g_j(\boldsymbol{x})) \tag{21}$$

In view of (20), r_j is given by

$$r_j = f(g_j) = p_j(g_j)/g_j \tag{22}$$

The new formulation (21) yields favourable far reaching implications which are presented in the next section.

IV. PROBABILISTIC SORTING FOR EFFECTIVE ELITISM

A. STAGE ONE: NON-DOMINATED SORTING AND ELITISM

The implementation of the theoretical results yielding an evolutionary optimization algorithm is accomplished in two stages. In the first stage non-dominated sorting ***(NS)***, tournament selection and elitism is carried out in a way essentially based on that as described in [7]. This is schematically illustrated in figure 3, where subtle details are also indicated for clarity.

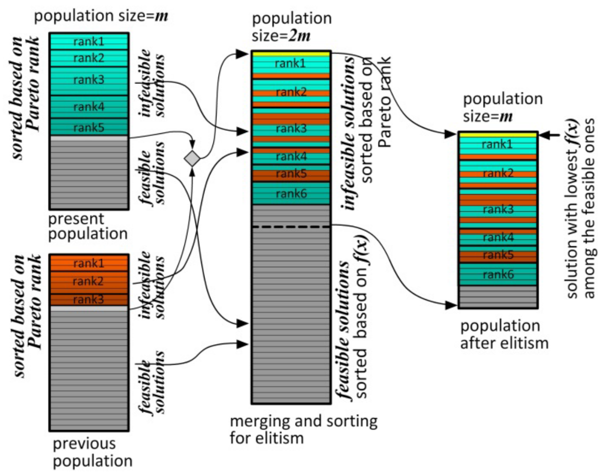


Fig. 3. NS based sorting and elitism

B. Stage two: Non-linear Ranking and Elitism

The NS algorithm described above is repeated for some number of generations so that the Pareto front sufficiently develops. Thereafter a non-linear ranking *(NR)* procedure based on the probabilistic considerations described above is employed as follows. During the tournament selection process, for two infeasible solutions from the productive domain, the value $P(g_j, \boldsymbol{x})$ in (2) is used to determine the winner of the tournament. In this procedure, clearly, a solution with lower $P(g_j, \boldsymbol{x})$ value is preferred over the solution with a larger $P(g_j, \boldsymbol{x})$ value. If a solution in the tournament belongs to the non-productive domain, then the same consequences apply as in the NS tournament. Namely, productive solutions win over non-productive solutions, and among non-productive solutions, the solution which is nearest to the productive domain wins. The possible outcomes during the non-linear ranking procedure are exemplified in figure 4. For instance, in this figure solution *B* represents the best solution among the feasible ones. When this solution is in a tournament with in infeasible solution from the productive domain, e.g. solution E, the winner of the tournament is obtained using $P(g_j, \boldsymbol{x})$. That is solution B is considered as if it were an infeasible one for

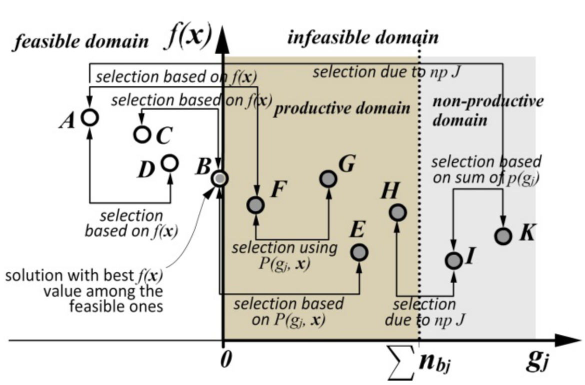


Fig. 4. Sketch for the tournament selection during NR

this comparison, so the chance *B* remains in the population is increased. For solutions from the productive domain, as $P(g_j, \boldsymbol{x})$ is a summation of function value $f(\boldsymbol{x})$ and summed up values of $p(g_j)$, population members that have a low function value and at the same time small sum of $p(g_j)$ are favored in the selection process. A solution having a low summation of $p(g_j)$ means that this solution has the unusual property that it violates several constraints with an extraordinarily low amount, when considered in perspective with the average violations of the respective constraints. In contrast to the Pareto-ranking based algorithm exercised before, the probabilistic selection mechanism will not permit solutions with low function value to remain in the population, provided the coefficient *C* is selected large enough.

The important implication of the *NR* tournament selection is assigning a commensurate right penalty parameter for every constraint, and even for each population member, where the penalty parameter is embedded in the non-linear distance function [16]. By means of this, the robustness and precision of the algorithm is guaranteed, together with the high stability of the search process. After the non-linear ranking based tournament selection, $P(g_j, \boldsymbol{x})$ is used during an elitism procedure, as seen in figure 10. From the figure it is noted that in the sorting step for the elitism the infeasible solutions are sorted based on their $P(g_j, \boldsymbol{x})$ values. Generally the mean values for the different constraints of two consecutive generations being merged for elitism differ, and it is generally expected that the mean values improve from generation to generation. In order to ensure accurate convergence, in this implementation for the sorting procedure during the NR elitism $P(g_j,\boldsymbol{x})$ is obtained using the mean value of the respective generation when the chromosome was created. This way the convergence is slowed down in order to ensure that the solutions from the past generation will also have significant influence in the ensuing generation. This is in order to maintain diversity during the search and carefully target the minimum being approached with the population. The nonlinear ranking based sorting and elitism is illustrated in figure 5.

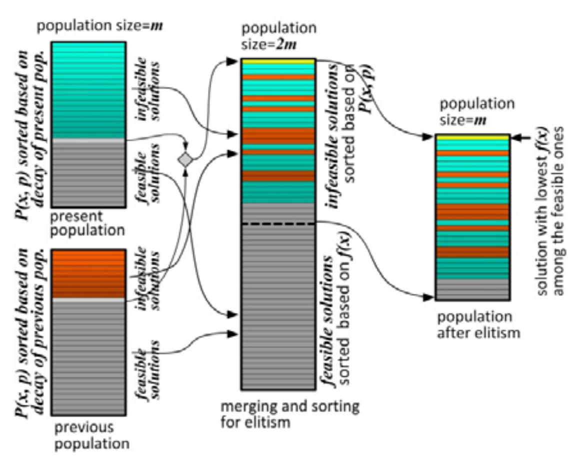
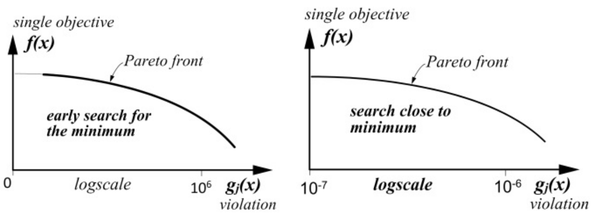


Fig. 5. NR based sorting and elitism

V. IMPLICATIONS OF THE PROBABILISTIC MODELING

A. Adaptive Zooming for Ranking and Effective Elitism

Adaptive zooming for ranking with precision is accomplished as follows. The favourable solutions are by accurately ranked in the range zero and unity as probabilistic distances, even though the actual constraint values may be close to the optimal point as much as the machine or genotype coding precision can allow, say at the range of 10^{-10}. A sketch of the Pareto front at the early stage of the genetic search is given in figure 6a. Illustration of the Pareto front at the last stage of the genetic search is given in figure 6b.



(a) (b)

Fig. 6. Sketch of formation of the Pareto front at the early stage (a); at the at the last stage of the GA search (b).

The probabilistic distance to the minimum is illustrated as a typical example in figure 2a by the indicated area. The computation of the colored area in the figure is very precarious at the tournament selection process, due to the issue of both exact parameterization of the exponential pdf in the existing range and the finite machine precision as well as the finite genotype coding. This situation is circumvented in figure 2b by taking simply $p(g_j)$ as the probability distance to the minimum. The indicated areas are the same and they are equal to $p(g_j)$. The indicated area in figure 2b defines 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 the transformed probabilistic space in a macro scale between zero and unity, always. This situation is equivalent to apply a commensurate mathematical 'lens' to the space formed by actual objective function and the constraint functions to carry out the convergence process without being effected by any scale of convergence happening in this space.

B. Effective Tournament Selection

Two important aspects in this work, beyond the straightforward tournament selection process, are the followings.

1. In the tournament of the non-linear ranking, the present and the preceding populations is accomplished using their respective decay constants (λ). In this case the situation is depicted in figure 7, 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.

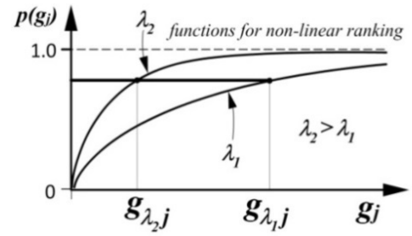


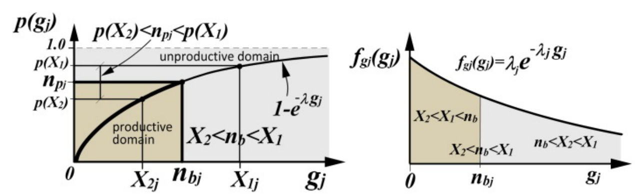
Fig. 7. Illustration of the gneration dependent ranking procedure during non-linear elitism

2. Solutions in NS as well as NR tournaments will be evaluated depending on the condition given by

$$\sum_{j=1}^{J} p(g_j) < n_{pj}\; J \qquad (23)$$

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}. This is seen in figure 8, where horizontal axis refers to NS (nondominated sorting) procedures and vertical axis refers to NR (nonlinear ranking) procedures.

In case one solution fulfills (23), while the other one does not, then the solution in the productive domain wins the tournament over the other one, without considering rank or



(a) (b)

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

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 (23) in the objective space is given by

$$g_T = n_{b_j} \sum_{j=1}^{J} \bar{g} = \sum_{j=1}^{J} \frac{n_{b_j}}{\lambda_j} \qquad (24)$$

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.

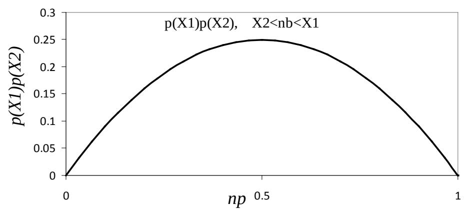


Fig. 9. Plot of the probability that two solutions occur on different sides of the thtreshold n_{pj}.

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. In figure 10 the single plot seen in figure 9 corresponds to a family of plots with respect to the parameter λ_j.

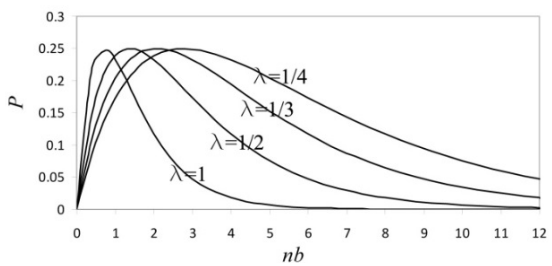


Fig. 10. Plot of the probability that two solutions occur on different sides of the threshold n_{bj}.

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 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.

Referring to figure 8b, the probability P_j of the event relevant to the case described above is given by

$$P_j = P(g_j) = P(X1_j)P(X2_j) = e^{-\lambda_j n_{bj}} - e^{-2\lambda_j n_{bj}} \qquad (25)$$

C. Fast and robust convergence

Thanks to the probabilistic distance providing nonlinear ranking, robust progress for convergence at each generation is obtained. To see this, from (22)

$$r_j = \frac{p(g_j)}{g_j} = \frac{1-\mathrm{e}^{-\lambda_j g_j}}{g_j} \qquad (26)$$

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

$$\lim_{g_j\to 0}\ r_j = \frac{p(g_j)}{g_j} = \lim_{g_j\to 0} \lambda_j e^{-\lambda_j g_j} = \lambda_j \qquad (27)$$

The variation of the penalty parameter r_j with g_j , based on (36) is shown in figure 11.

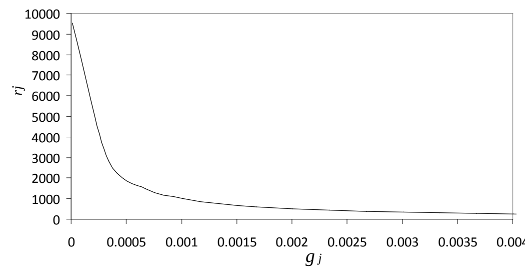


Fig. 11. Illustration of the *new* penalty parameter *r* as to probabilistic modeling: r=(1-exp(-λg))/g, where λ=10000

VI. Computer Experiment

Computer experiments have been carried out using a standard optimization problem from the literature. The following problem is due to Floundas and Pardalos [17]. The problem consists of a single objective with 9 constraints, subject to minimization, as given by (38)-(40).

$$\mathit{Minimize}\ f(\boldsymbol{x}) = 5\sum_{i=1}^{4} x_i - 5\sum_{i=1}^{4} x_i^2 - \sum_{i=5}^{13} x_i \qquad (28)$$

$$\begin{aligned}
g_1(\boldsymbol{x}) &= 2x_1 + 2x_2 + x_{10} + x_{11} - 10 \le 0\\
g_2(\boldsymbol{x}) &= 2x_1 + 2x_2 + x_{10} + x_{12} - 10 \le 0\\
g_3(\boldsymbol{x}) &= 2x_2 + 2x_3 + x_{11} + x_{12} - 10 \le 0\\
g_4(\boldsymbol{x}) &= -8x_1 + x_{10} \le 0\\
g_5(\boldsymbol{x}) &= -8x_2 + x_{11} \le 0\\
g_6(\boldsymbol{x}) &= -8x_3 + x_{12} \le 0\\
g_7(\boldsymbol{x}) &= -2x_4 - x_5 + x_{10} \le 0\\
g_8(\boldsymbol{x}) &= -2x_6 - x_7 + x_{11} \le 0\\
g_9(\boldsymbol{x}) &= -2x_8 - x_9 + x_{12} \le 0
\end{aligned} \qquad (29)$$

where the ranges for the independent variables are given by

$$0 \le x_i \le 1\ (i = 1,...,9);\ 0 \le x_i \le 100\ (i = 10,11,12);\ 0 \le x_i \le 1\ (i = 1 \qquad (30)$$

The best known optimum is

$f(\boldsymbol{x}^*)$=-15.0,

and the corresponding best variable values are $\boldsymbol{x}^*$=(1,1,1,1,1,1,1,1,1,3,3,3,1).

The algorithm is executed with the following settings: population size=200; amount of generations=100; *C*=1000; ratio of NS/NR procedures=4/1; crossover probability=0.9; mutation probability=0.05. The results are shown in figure 12-15 using a logarithmic scale for the horizontal axis, which shows the total violation *G*. From the figures it is observed how the initial population gradually approaches towards the optimal solution. It is emphasized that an iteration of the algorithm consists of 4 Pareto-ranking based generations, followed by one probabilistic selection based generation.

After 10 iterations the best feasible solution is found to be *f*(***x***)=-13.98583864.
The population is shown in figure 12.

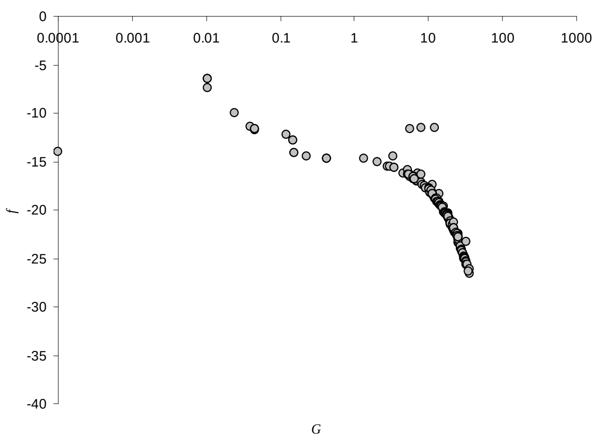


Fig. 12. Population after the 10[th] iteration; horizontal axis is the total violation *G* on a log scale; vertical axis is *f*(***x***)

After 20 iterations the best feasible solution is found to be *f*(***x***)=-14.9076345785146.

The population is shown in figure 12. The figures 12-17 demonstrate the robust convergence properties of the algorithm. Namely, the population members form a compact aggregation about the close vicinity of the optimum. This aggregation makes the mean of the violations g_j small, so that the decay constant λ_j of the exponential pdf becomes large, and consequently the slope of the penalty parameter *r* is large. Due to this, the convergence to the optimum is fast, accurate, and with precision. Due to the memoryless ness property of the exponential pdf, the populations form about the same patterns in any scale of the convergence process. This is clearly seen in the figures by the logarithmic scale of representations of the violations.

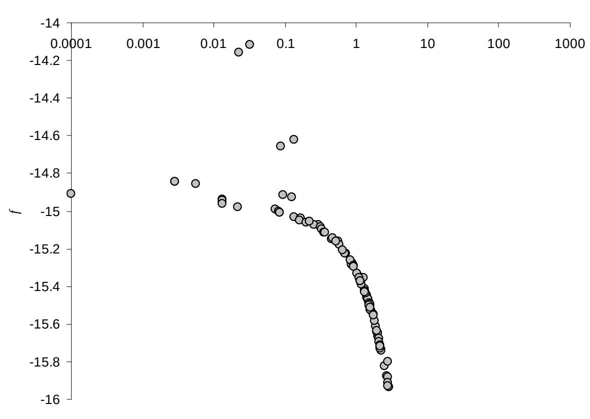


Fig. 13. Population after the 20th iteration; horizontal axis is the total violation *G* on a log scale; vertical axis is *f*(***x***)

After 30 iterations the best feasible solution is found to be *f*(***x***)=-14.9760230713287.
The population is shown in figure 14.

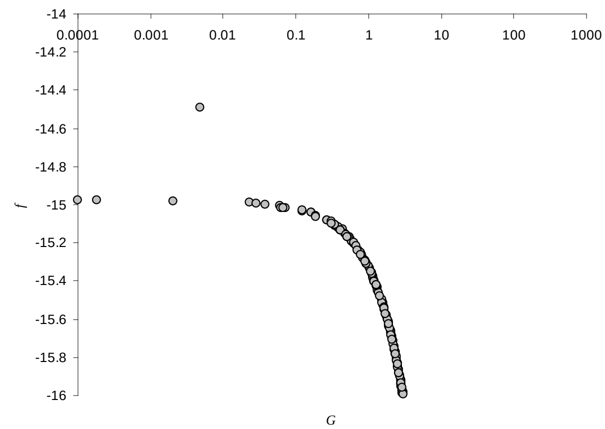


Fig. 14. Population after the 30th iteration; horizontal axis is the total violation *G* on a log scale; vertical axis is *f*(***x***)

After 50 iterations the best feasible solution is found to be *f*(***x***)=-14.9985221605613.

The population is shown in figure 14. The independent variable values of this solution are x_1=0.999997312719596; x_2=0.999997982197311; x_3=0.999999888524811; x_4=0.999999871166525; x_5=0.999994649877324; x_6=0.999987862005421; x_7=0.999984815877352; x_8=0.999999999750139; x_9=0.999926599531956; x_{10}=2.99995794671011; x_{11}=2.99961604207534; x_{12}=2.99907993443006; x_{13}=0.999999037205755.

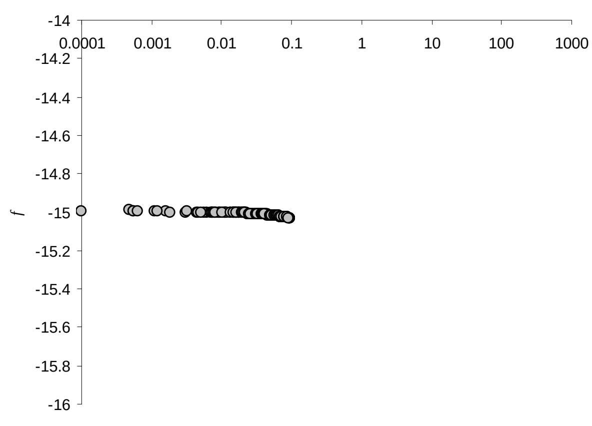


Fig. 15. Population after the 50th iteration; horizontal axis is the total violation *G* on a log scale; vertical axis is *f*(***x***)

After 80 iterations the best feasible solution is found to be *f*(***x***)=-14.999997075874.

The whole population is shown in figure 12. The independent variable values of this solution are x_1=0.999999970028148; x_2=1; x_3=0.999999993220535;

x_4=0.999999971015986; x_5=0.999999868494369; x_6=0.999999960482284; x_7=0.999999981750632; x_8=0.999999947363726; x_9=0.999999932369763; x_{10}=2.99999980083455; x_{11}=2.99999890685553; x_{12}=2.99999900736034; x_{13}=0.999999999039419.

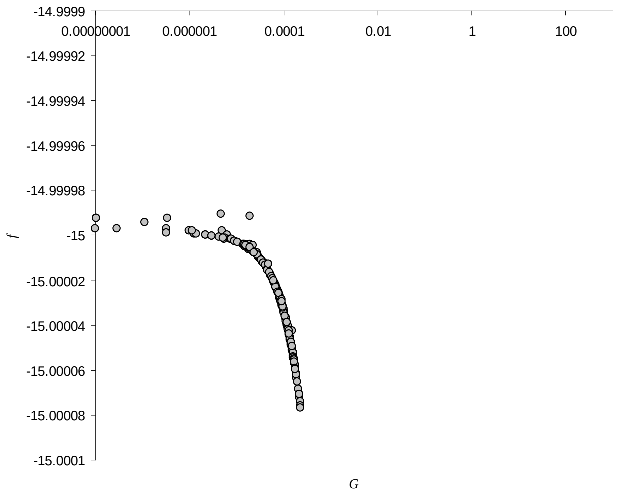


Fig. 16. Population after the 80th iteration; horizontal axis is the total violation G on a log scale; vertical axis is $f(\boldsymbol{x})$

After 100 iterations the best feasible solution is found to be $f(\boldsymbol{x})$=-14.9999999458368.

The whole population is shown in figure 15. The independent variable values of this solution are x_1=0.999999999790223; x_2=0.999999999809861; x_3=0.99999999933798; x_4=0.99999999995506; x_5=0.999999994377998; x_6=0.999999995023679; x_7=0.999999999831045; x_8=1; x_9=0.999999996761354; x_{10}=2.99999999225267; x_{11}=2.99999998081316; x_{12}=2.99999999477602; x_{13}=0.999999997535212.

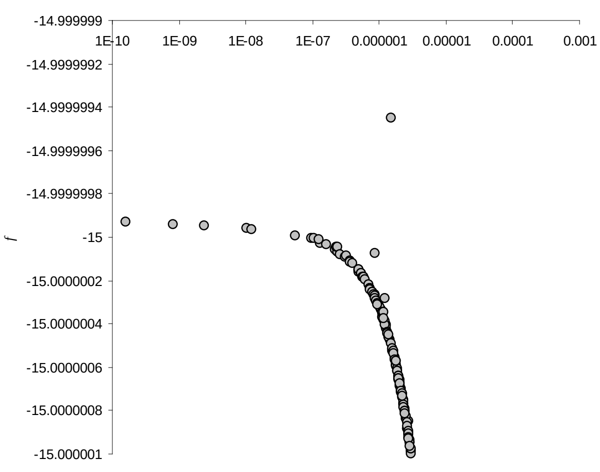


Fig. 17. Population after the 100th iteration; horizontal axis is the total violation G on a log scale; vertical axis is $f(\boldsymbol{x})$

VII. CONCLUSIONS

Probabilistic sorting for effective elitism in multi-objective evolutionary algorithms is presented. In the evolutionary optimization ranking of the genetic population members plays very important role on the performance of the algorithm. This work addresses this issue by a new non-linear ranking procedure, which eventually leads to an effective elitism and marked performance of the algorithm. Conventionally, in constrained or multi-objective optimization problems evolutionary computation turns out to be supported by auxiliary optimization means, in order to approach the optimum sufficiently close. In this respect, by means of the new methodology a marked improvement is achieved. The source of the improvement lies in the non-linearity of the ranking, achieved by the transformation of the objective space to a newly defined probabilistic distance domain. The transformation is adaptively carried out throughout the generations, so that the commensurate ranking with respect to the generation is maintained. Additionally, explicit definition of productive and non-productive chromosomes has been made, and accordingly maximum gain from the unproductive chromosomes is exported to the productive portion of the population at each generation. By means of the particular tournament selection procedure, the detrimental effect on the average violation by the stiff constraints, that is, by highly non-productive population members is prevented; namely, during two consecutive generations the progressive diminishing of the average is augmented against the contingent average increase that may be effective especially during the advanced stages of the convergence. Non-linear ranking plays two major roles at the same time. One is the accomplishment of an adaptive penalty parameter matching the optimality conditions during the search. The other is maintaining maximum gain constantly from unproductive to productive solutions. This allows accurate and systematic convergence with precision, which is also rapid. The probabilistic sorting is implemented in both, nonlinear tournament selection and elitism. The method showed outstanding performance as to speed of convergence, precision and approaches to the solution without auxiliary support like local search, memetic algorithm etc. This is exemplified by means of a standard problem chosen from the literature for the comparison of the results and demonstration of the effectiveness of the methodology. The reported results include not only the final outcomes but also the progress of the convergence throughout the optimization process, clearly showing the exact matching of the results with the theoretical underlying material. It is also noteworthy to mention that, due to the systematic convergence procedure established by the novel method, the search process is demonstrated to be transparent.

REFERENCES

[1] C. M. Fonseca, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary Computation,* vol. 3, pp. 1-16, 1995.

[2] C. A. C. Coello, "An updated survey of Ga-based multi-objective optimization techniques," *ACM Computing Surveys,* vol. 32, pp. 109-143, 2000.

[3] D. E. Goldberg, *Genetic Algorithms*. Reading, Massachusetts: Addison Wesley, 1989.

[4] C. A. C. Coello, D. A. Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multiobjective Problems*. Boston: Kluwer Academic Publishers, 2003.

[5] K. Deb, *Multiobjective Optimization using Evolutionary Algorithms*: John Wiley & Sons, 2001.

[6] K. Deb and R. Datta, "A bi-objective constrained optimization algorithm using a hybrid evolutionary and penalty function approach," *Engineering Optimization,* vol. 45, pp. 503-527, 2013.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation,* vol. 6, pp. 182-197, 2000.

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

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

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

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

[12] 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.

[13] P. L. Meyer, *Introductory Probability and Statistical Applications*: Addison-Wesley, 1970.

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

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

[16] O. Ciftcioglu, M. S. Bittermann, and I. S. Sariyildiz, "Precision Evolutionary Optimization - Part I: Nonlinear Ranking Approach," presented at the GECCO 2012, Philadelphia, 2012.

[17] C. Floundas and P. Pardalos, *A collection of test problems for constrained global optimization, Vol.455*. Berlin, Germany: Springer Verlag, 1987.

S. Serhat Seker was born in Istanbul on July 2, 1959. He started his education at Mathematics Engineering Department of Istanbul Technical University (ITU) and graduated from Electrical and Electronic Engineering Faculty where he started his career as a research assistant at Electrical Power Engineering Department in 1985. He got his master degree at ITU's Nuclear Energy Institute and his doctorate at Electrical Engineering Division of the same university's Science and Technology Institute with his research titled as "Stochastic Signal Processing with Neural Network in Power Plant Monitoring."

Dr. Seker studied during his PhD thesis at Energy Research Centre of the Netherlands (ECN) with his scholarship provided from Netherlands Organization for International Cooperation in Higher Education (NUFFIC) and worked on signal analysis techniques there. He was titled as Assistant Professor and Associate Professor at ITU in 1995 and 1996 respectively. Also, he worked on industrial signal processing at Nuclear Engineering Department and Maintenance and Reliability Centre of the University of Tennessee, Knoxville-USA, by getting the scholarship from the Scientific and Technological Research Council of Turkey (NATO-TUBITAK B1) in 1997. He had many administrative duties at ITU in the previous years, including his vice dean position at Electrical and Electronic Engineering Faculty during 2001 - 2004. He was a department head in Electrical engineering department between 2004 and 2007. Also, he was dean of Technology Faculty and vice rector as well as the founding dean of Engineering Faculty at Kırklareli University between 2011 and 2013. Dr. Seker was dean of the Electrical and Electronics Engineering Faculty at ITU between 2013-2016.

Michael S. Bittermann was born in Germany in 1976. He graduated with the highest honors as MSc in Architecture from Delft University of Technology (TU Delft), The Netherlands, in 2003. He received his PhD degree in Design Informatics with the highest honors from the same university in 2009. Bittermann was distinguished as one of the five eminent researchers among all 2008-2009 PhD graduates of TU Delft in the form of the Young Researcher Fellowship of Delft University of Technology (80k€). He joined Maltepe University, Istanbul, Turkey in 2013, where he contributed to the research project titled 'Design by Cognition and Comprehension,' funded by the Scientific and Technological Research Council of Turkey. His research interest is computational cognition.

Ramazan Caglar received the BSc., and MSc. diploma in Electrical Engineering from the Istanbul Technical University in 1984 and 1987 respectively, and PhD degrees in Graduate School of Science, Engineering and Technology of the same university in 1999. He accepted as Postdoctoral Researcher by the School of Electric and Computer Engineering at Georgia Institute of Technology, Atlanta, USA in 2001. He has conducted his research activities for two years in this university. His research interests are power system modeling, reliability of electric power system, the effects of deregulation on the electric power system, power transmission pricing, distributed generations, and fault diagnosis, risk management and signal processing.

In 1984 he joined the Faculty of Electrical Engineering, Istanbul Technical University as a research assistant, where he is presently an Assoc. Professor. He is active in teaching and research in the general modeling, analysis, and control of power systems. His has published article in some referred journals, and conference proceedings.

Dr. Caglar is a member of IEEE, IEEE-PES, a member of CIGRE, a member of Turkish National Committee on Illumination (ATMK), and a National Chamber of Turkish Electrical Engineering (EMO).

Rituparna Datta received the PhD degree from the Indian Institute of Technology Kanpur (IITK), Kanpur, India, in 2013.

He is a Post-doctoral researcher with Graduate School of Knowledge Service Engineering, Department of Industrial & Systems Engineering at Korea Advanced Institute of Science and Technology (KAIST). Prior to that he was a Post-doctoral fellow in French Institute for Research in Computer Science and Automation (Institut national de recherche en informatique et en automatique, INRIA), France, Project scientist in Smart Materials, Structures and Systems Laboratory, Indian Institute of Technology, Kanpur, India and Postdoctoral research fellow with the Robot Intelligence Technology (RIT) Laboratory at KAIST. His current research interests include investigation of evolutionary computation, constraint handling, multiobjective optimization, knowledge extraction from data, and robotics. His has two edited books with Springer, many international journals and international conferences in his credit.

Dr. Datta is a regular Reviewer of the IEEE Transactions on Evolutionary Computation, IEEE Transactions on Cybernetics, The Journal of Applied Soft Computing, and The Journal of Engineering Optimization.