Further Note on the Probabilistic Constraint Handling

Özer Ciftcioglu, Senior Member, IEEE
Department of Architecture
Delft University of Technology, Delft, The Netherlands
Maltepe University, Istanbul, Turkey
o.ciftcioglu@tudelft.nl ozerciftcioglu@maltepe.edu.tr

Michael S. Bittermann
Department of Architecture
Maltepe University
Istanbul, Turkey
michaelsbittermann@maltepe.edu.tr

Rituparna Datta, Member, IEEE
Department of Mechanical Engineering
Indian Institute of Technology, Kanpur
Kanpur, Pin 208016, India
rdatta@iitk.ac.in

***Abstract*—A robust probabilistic constraint handling approach in the framework of joint evolutionary-classical optimization has been presented earlier. In this work, the theoretical foundations of the method are presented in detail. The method is known as bi-objective method, where the conventional penalty function approach is implemented. The present work highlights the dynamic variation of the commensurate penalty parameter for each objective treated as constraint. It is shown that the constraint parameters collectively define the right slope of the tangent as to the optimal front during the search. The robust and sustained convergence throughout the search up to micro level in the range of 10^{-10} or beyond is explained. The work here is presented as a further note in connection with the previous publication, where the subtle theoretical considerations and their details had been omitted for the sake of detailed results of the experiments demonstrating the effective working of the approach. In contrast to the implementation-centered reporting of the previous work, this work can be considered as a description of the detailed probabilistic basis underlying the previous work. Therefore, this study is of great importance to let the researchers conveniently gain the insight into the work and its implications reported earlier.**

***Keywords*—evolutionary algorithm; multiobjective optimization; constrained optimization; probabilistic modeling**

I. Introduction

This work is a further note on the previous research [1], to introduce more insight into the working mechanism of a probabilistic approach for effective constraint handling in the context joint evolutionary-classical optimization. At the same time it forms the basis of another study, where the theoretical considerations concerning the probabilistic method are verified by an exclusively evolutionary implementation, i.e. without classical component [2]. The purpose of this paper is twofold. On one hand it provides detailed analysis of the method for probabilistic constraint handling in the joint evolutionary-classical case. Thus the work can be considered as a significant complimentary or supplementary work to let the researchers of that approach gain more insight into the probabilistic component, and to understand the working mechanism of the method, rather than only comprehending the method without thoroughly understanding its working principles, and its implications. In this work probabilistic considerations are prevailing in contrast to conventional constraint-handling procedures, together with some interesting features of the probabilistic method that are being pointed out. On the other hand the effectiveness of the probabilistic method alone, i.e. of its implementation without auxiliary means like local search, presented in [2], is theoretically explained in detail in this paper.

Since the advent of genetic algorithms for solving optimization problems some three decades ago, the advancements made along this line are surprisingly rapid. Eventually, today we are dealing with evolutionary computation encompassing many advanced optimization algorithms having the spirit of genetic algorithms in essence. The rapid developments may be broadly categorized as single optimizations, multiobjective optimizations in Pareto sense, and multiobjective constrained optimizations. Referring to the latter, the present work aims to shed some light on further probabilistic considerations as to continuous progress along this line. There are a number of excellent text books that contributed to the progress of evolutionary multiobjective optimization [3-5]. Evolutionary optimization algorithms are widely used to solve general optimization problems and updated surveys are reported in the literature from time to time, e.g. [6-8]. Such problems are extensively treated in literature [8-24]. Since multiobjective optimization can be formulated as a single objective with constraints, where the constraints are the rest of the objectives subject to minimization, it is interesting to tackle the constrained optimization with single objective function as a general case and this is the case in this work. A widely used method for constrained optimization is the penalty function method [25]. Penalty function method penalizes a solution, which deteriorates the fitness of a solution when it violates constraints. This penalization is accomplished by adding a value to the objective function value in proportion to the amount of constraint violation, where the proportionality factor is known as the penalty parameter. A strategy that did not require a penalty parameter in evolutionary constrained optimization was proposed by Deb in 2000 [26], which is superseded by another research with the penalty parameter [27]. In this approach during the tournament selection process an infeasible solution is always treated as inferior compared to a feasible one, or as inferior to a solution that violates the constraints to a lesser extent. Coello [28] proposed a self-adaptive penalty approach by using a co-evolutionary model to adapt the penalty factors. However, in general determination of a right penalty parameter still remained an issue [29].

This work addresses the multiobjective optimization as a single objective optimization together with a penalty function. The issues of penalty approach having been pointed out in above mentioned works, in this paper details and implications of a new approach is proposed, where a probabilistic model of the random solutions is used to derive a nonlinear distance measure that it is used for effective, i.e. robust ranking of genetic population members and efficient, i.e. fast convergence, and stable solutions. The measure is used for nonlinear ranking among the population members during the evolutionary process. The method is studied for several standard test problems in two implementation scenarios. One scenario concerns local search based optimization with evolutionary support. The test problem results for this implementation are reported earlier [1]. The second scenario is pure evolutionary computation, i.e. local search is omitted. The test problem results for this implementation are reported in [2]. The organization of the paper is as follows. In section two, problem of constrained optimization via multiobjective optimization is formulated, the issues of the approach are pointed out, and analyses of the penalty parameter are presented. In section three, based on these analyses the probabilistic modeling for nonlinear exponential ranking is described explaining the exact working mechanism of the method. In section four the implications of the analyses are presented. This is followed by discussion and conclusions.

II. WEIGHTING METHOD FOR MULTIOBJECTIVE OPTIMIZATION

A. Problem Formulation

The formulation in this research stems from the considerations known as *weighting method* [30-32]. 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,….,k so that a weighting problem can be stated as

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

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

$$\min \quad f(\mathbf{x}) \ \ subject\ to \ \ g(\mathbf{x}) = [g_1(x), g_2(x), ..., g_m(x)]^T \le 0 \tag{2}$$

We assume that the feasible region is of the form

$$S = \{x \in R^n \mid \ g(\mathbf{x}) = [g_1(x), g_2(x), ..., g_m(x)]^T \ge 0\} \tag{3}$$

One notes that in this formulation every constraint function $g_i(\mathbf{x})=-v_i(\mathbf{x})$ where v denotes the actual degree of violation of a constraint, and this degree is a non-negative number for a violated constraint. The functions $g_i(\mathbf{x})$ have a negative value for a violated constraint, so that $\langle g_j(\mathbf{x})\rangle$, where $\langle \alpha \rangle$ is the *bracket operator* that is equal to $-\alpha$ if $\alpha<0$, and zero otherwise, have a positive value for a violated constraint. Therefore, the sum of violations $\langle g_j(\mathbf{x})\rangle$ is another objective subject to minimization. That is, the problem formulation becomes a problem of two objective functions subject to minimization. In this case the formulation of the problem using weighting method becomes

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

where $G(\mathbf{x})=f_1(\mathbf{x})$ and $f(\mathbf{x})=f_2(\mathbf{x})$, and for k number of constraints $G(\mathbf{x})$ is given by

$$G(\mathbf{x}) = \sum_{i=1}^{k} \mu_i \langle g(\mathbf{x}) \rangle \tag{5}$$

where $\boldsymbol{\mu}$ are non-negative values that are not all zero. Thus, the problem definition becomes explicitly,

$$\begin{aligned} &\min \ \sum_{i=1}^{k} \mu_i \langle g_i(\mathbf{x}) \rangle + f(\mathbf{x}) = G(\mathbf{x}) + f(\mathbf{x}) \\ &S = \{x \in R^n \mid \langle g(\mathbf{x}) \rangle = [\langle g_1(\mathbf{x}) \rangle, \langle g_2(\mathbf{x}) \rangle, ..., \langle g_m(\mathbf{x}) \rangle]^T \ge 0\} \end{aligned} \tag{6}$$

where $w_1=\mu_i$, $w_{2i}=1$. Without deviating from generality, this formulation of the problem is equivalent to a single objective problem with the objective $f(\mathbf{x})$ and the constraints denoted by $\langle g_j(\mathbf{x})\rangle$. Such an approach is known as ε-*Constraint* method [32, 33]. Here one of the objective functions is selected to be optimized and all the other objective functions are converted into constraints by setting an upper bound to each of them. The problem to be solved is now of the form

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

where $l \in \{1,...,k\}$. Naturally, inequalities can be converted to equalities by taking $\varepsilon_j=0$ for all $j=1,2,....,k$, $j \ne l$.

B. Issues of the penalty function approach

Conventionally, (6) is written in the form

$$\min \; P(\boldsymbol{x}, R) = f(\boldsymbol{x}) + \sum_{j=1}^{J} R_j \left\langle g_j(\boldsymbol{x}) \right\rangle \qquad (7)$$

where function $\langle g_j(\boldsymbol{x})\rangle$ is considered to be a *penalty function* and the parameters R_j are the associated *penalty parameters*. Since each individual R_j is not known, conventionally a common penalty parameter R is defined so that (7) becomes

$$\min \; P(\boldsymbol{x}, R) = f(\boldsymbol{x}) + R \sum_{j=1}^{J} \left\langle g_j(\boldsymbol{x}) \right\rangle \qquad (8)$$

or taking $f_2(\boldsymbol{x}) = f(\boldsymbol{x})$ and the summation of the $\langle g_j(\boldsymbol{x})\rangle$ functions as $f_1(\boldsymbol{x})$, we can write

$$P_{opt} = \min\{f_2(\boldsymbol{x}) + R\, f_1(\boldsymbol{x})\} \qquad (9)$$

To solve the optimization problem given by (9) with the weighting method, one can consider the development of the optimal front is illustrated in figure 1. The final development is the theoretical front and the solution is denoted with the point T which is far from the optimal point denoted by P_{opt}. As result of this option some gradient-based search algorithm is necessary that tails up evolutionary computation to reach the optimal point if it is realizable at all due to the chance of getting trapped in some local optima. During the Pareto front formation the most of the attention of the chromosomes goes to the penalty function rather than the objective function. As result of this, the convergence is essentially due to the constraints and therefore there is a significant progress along that line, while the single objective is de facto subsumed under the constraints. This situation makes determination of R very critical and precarious at the same time.

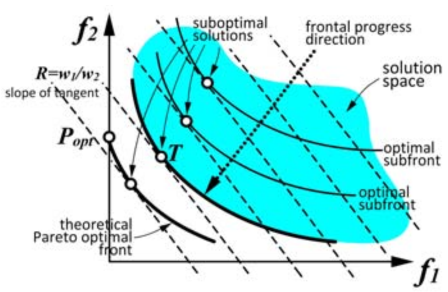


Fig. 1. Approach to the final optimal solution by penalty parameter R.

C. Analysis of penalty function parameter

Let us assume that optimal theoretical front compromises the solutions for the objectives $f_1(x)$ and $f_2(x)$, where objective $f_1(x)$ admits to be minimally zero. For the analysis viewpoint we assume that Pareto front is symmetrical with respect to $f_1(x)$ and $f_2(x)$, and the front is an envelope of a line crossing the $f_1(x)$ axis at the point t and crossing the $f_2(x)$ *axis* at the point P_{opt}-t; t is a parameter related to parametric representation of a line tangent to the Pareto front, and it is represented by

$$\frac{f_2(\boldsymbol{x})}{t} + \frac{f_1(\boldsymbol{x})}{P_{opt} - t} = 1 \qquad (10)$$

In (10), P_{opt} is the optimum solution, where $f_2(\boldsymbol{x}) = P_{opt} = t$ and $f_1(\boldsymbol{x}) = 0$, which represents the satisfaction of the constraint. From (10), we obtain

$$f_2(\boldsymbol{x}) = \frac{t}{t - P_{opt}(\boldsymbol{x})} f_1(\boldsymbol{x}) + t \qquad (11)$$

We can define the slope

$$r = \frac{t}{t - P_{opt}(\boldsymbol{x})} \qquad (12)$$

as a *kernel* penalty parameter representing the varying part of the general penalty parameter R in (8), and for each constraint we consider $r = r_j$. The envelope of the tangent in (10) is shown in figure 2.

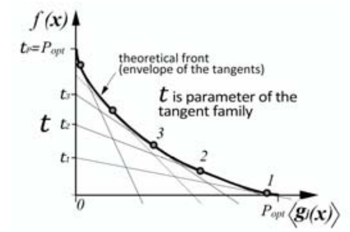


Fig. 2. The variation of the *new* penalty parameter $r = t/[t - P_{opt}(x)]$

In words, r is the gain in $f_2(\boldsymbol{x})$ per unit decrease in $f_1(\boldsymbol{x})$ at the point of tangent F and within infinitesimally small interval of $f_1(\boldsymbol{x})$. Incidentally, the envelope of the tangent is determined by the following condition obtained in Appendix A

$$t = f_2(x) + \sqrt{f_2(x) f_1(x)} \qquad (13)$$

And substitution of (13) in (11) yields the Pareto front expression as

$$[f_2(x) - f_1(x)]^2 - 2[f_1(x) + f_2(x)] + P_{opt} = 0 \qquad (14)$$

Variation of r during the minimization process for a given constraint j is shown in figure 3.

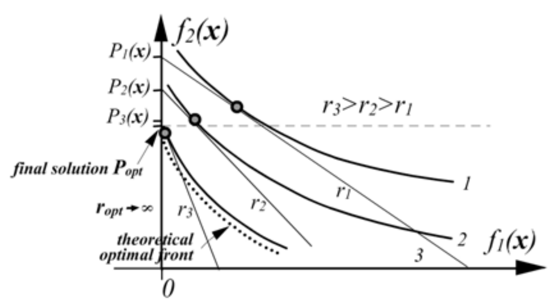


Fig. 3. Variation of the penalty parameter r.

As shown in the figure, as the process approaches to the minimum, the slope tends to approach infinity. Therefore, in this work penalty parameter R in (7) is not a constant, but it is a varying parameter, adapted during the search process, which is peculiar to this work.

The kernel penalty parameter r is zero for $t=0$ and it monotonically increases as t increases, as seen in (12), and t is given by (13). As P_{opt} is reached, at this point $f_2(x)=0$, and

$t=f_2(x)$ where $t=P_{opt}$. For $t=P_{opt}$ the kernel penalty parameter r goes to infinity, as seen in (11). Alternatively, this work shows that the kernel parameter r is a function of the objective functions f_1 and f_2, and at the end of the search process the intersection of the tangent given by (10) is the minimum being sought for, where $f_1=0$ and f_2 is the minimum. At that point Pareto front and tangent disappear, and they reduce to the point P_{opt}.

A convergence approach complying with (12) exhibits two gains:

- Approach to optimum is systematic and therefore robust without precarious tangent slope computations
- No local search for P_{opt} is necessary.

Implementation of the approach is due to a probabilistic modeling of the random solutions in the evolutionary computation and ensuing nonlinear ranking. These are presented in the following section

III. Probabilistic Modeling for Nonlinear Exponential Ranking

Referring to (6), in a general constrained optimization problem the problem formulation is written as

$$\min \; P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{j=1}^{J} \mu_j \left\langle g_j(\boldsymbol{x}) \right\rangle \tag{15}$$

where $f(\boldsymbol{x})$ is the single objective function to be minimized; $\langle g_j(\boldsymbol{x})\rangle$ is the violation of the j-th constraint, namely penalty function, μ_j is the associated parameter of the penalty function. Since $\langle g_j(\boldsymbol{x})\rangle$ is at each generation continually tried to be vanishing during the evolutionary minimization process, considering the population density of solutions, the probability density of $\langle g_j(\boldsymbol{x})\rangle$ is highest about zero violations, and its value 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. 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. That is 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

$$f_\lambda(y) = \lambda e^{-\lambda y} \tag{16}$$

where λ is the decay parameter. Denoting

$$y = \left\langle g_j(x) \right\rangle \tag{17}$$

the pdf in (16) becomes

$$f_{g_j}(\left\langle g_j \right\rangle) = \lambda_j e^{-\lambda_j \langle g_j \rangle} \tag{18}$$

The mean value of the exponential pdf function is equal to λ_j^{-1}. During the evolutionary search $\langle g_j(\boldsymbol{x})\rangle$ 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

$$f_{g_j}(\left\langle g_{j,s} \right\rangle) = \lambda_j e^{-\lambda_j \langle g_{j,s} \rangle} \tag{19}$$

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 $\langle g_j\rangle$ to the mean of the exponential pdf, namely

$$\lambda_j = 1 / \left\langle \bar{g}_j \right\rangle \tag{20}$$

One should note that the mean of the exponential probability density of $\langle g_j\rangle$ is equivalent to the mean of a uniform probability density applied to the violations $\langle g_j\rangle$. 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. Since a violation $\langle g_j\rangle$ spans all the violations starting from zero up to the point $\langle g_j\rangle$, the probability of the violation is expressed as cumulative distribution function whose implication is easy to comprehend by considering the extremes. The cumulative distribution function of (16) is given by

$$p(\left\langle g_j \right\rangle) = \frac{1}{\left\langle \bar{g}_j \right\rangle} \int_0^{\langle g_j \rangle} e^{-\frac{\langle g_j \rangle}{\langle \bar{g}_j \rangle}} dg_j = 1 - e^{-\frac{\langle g_j \rangle}{\langle \bar{g}_j \rangle}} \tag{21}$$

For $\langle g_j\rangle=0$ violation is zero and for $\langle g_j\rangle=\infty$, violation is 1, i.e., 100%. Explicitly $p(\langle g_j\rangle)$ is the probability of a violation in the range zero and $\langle g_j\rangle$. It is monotonically increasing function complying with the boundary conditions of $\langle g_j(x)\rangle$ which varies between zero and infinity. It is interesting to note that for zero constraint violation the exponential probability density is maximum and probability of violation is minimum.

The probability $p(\langle g_j\rangle)$ is an appropriate measure for the magnitude or effectiveness of a violation, and 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 [34, 35]. The important implication of the premise (21) will be seen shortly afterwards.

The optimization problem with constraints is formulated in this work as follows.

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{j=1}^{J} c_j r_j(\left\langle g_j \right\rangle) \left\langle g_j(\boldsymbol{x}) \right\rangle \tag{22}$$

where c_j is a penalty parameter belonging to the constraints and varying during the search process, and $r_j(\langle g_j\rangle)$ is a penalty parameter also varying during the search process and belonging to each constraint. Therefore r_j is called as *convergence parameter,* being related to the convergence properties of the search, which in general means that it is a function of $\langle g_j(\boldsymbol{x})\rangle$. For each constraint, separately, we can write

$$f_{1j}(\boldsymbol{x}) = c_j r_j\left(\left\langle g_j\right\rangle\right)\left\langle g_j(\boldsymbol{x})\right\rangle \tag{23}$$

And from (12) and (13)

$$r_j = \frac{f_2(\boldsymbol{x}) + \sqrt{f_2(\boldsymbol{x}) f_{1j}(\boldsymbol{x})}}{f_2(\boldsymbol{x}) + \sqrt{f_2(\boldsymbol{x}) f_{1j}(\boldsymbol{x})} - P_{opt}(\boldsymbol{x})} \tag{24}$$

In (23) $c_j rj(\langle g_j\rangle)\ \langle g_j\rangle$ is replaced by $p(\langle g_j\rangle)$, in the form

$$c_j r_j\left(\left\langle g_j\right\rangle\right)\left\langle g_j\right\rangle = c_j p_j\left(\left\langle g_j\right\rangle\right) \tag{25}$$

Hence (22) becomes

$$P(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{J} p_j\left(\left\langle g_j(\boldsymbol{x})\right\rangle\right) \tag{26}$$

The absolute value of r_j in (25) is due to the bracket operator mentioned with respect to (7). Justification of (29) can be seen by the limiting values, as follows. For $\langle g_j\rangle$ goes to infinity, then $p_j(\langle g_j\rangle)$ goes obviously to 1 due to (21). The product $p_j = c_j r_j(\langle g_j\rangle)\langle g_j\rangle$ is computed using (11), noting that $\langle g_j\rangle$ is equal to f_{2j} , and as $\langle g_j\rangle$ goes to infinity P_{opt} also goes to infinity. From (23)

$$\lim_{\langle g_j\rangle\to\infty} c_j r_j\left(\left\langle g_j\right\rangle\right)\left\langle g_j\right\rangle = \lim_{\langle g_j\rangle\to\infty} c_j \frac{t}{t-\left\langle g_j\right\rangle}\left\langle g_j\right\rangle = c_j t \tag{27}$$

Taking

$$\lim_{\langle g_j\rangle\to\infty} c_j \to \frac{1}{t} \tag{28}$$

then (27) becomes

$$\lim_{\langle g_j\rangle\to\infty} c_j r_j\left(\left\langle g_j\right\rangle\right)\left\langle g_j\right\rangle = 1 \tag{29}$$

It is to note that c_j is varying between $1/P_{opt}$ and 0 during the convergence. For g_j is equal to zero, $p_j(\langle g_j\rangle)$ in (21) goes to zero. In this case, the penalty term $c_j r_j(\langle g_j\rangle)\langle g_j\rangle$ becomes zero as it should be.

In view of (25), r_j is given by

$$r_j = f\left(\left\langle g_j\right\rangle\right) = p_j\left(\left\langle g_j\right\rangle\right) / \left\langle g_j\right\rangle \tag{30}$$

The new formulation (30) yields favourable, far reaching implications which are presented below. From (6), where we define

$$\sum_{j=1}^{J} \mu_j \left\langle g_j\right\rangle = G = \sum_{j=1}^{J} p\left(\left\langle g_j\right\rangle\right) \tag{31}$$

where μ is the weighting parameter. J is the number of constraints; The probability $p(\langle g_j\rangle)$ controls the penalty parameter in (8) R where $c_j r_j$ is absorbed in $p(\langle g_j\rangle)$ while c_j is a constant being dependent on the constraints. The importance of this nonlinear transformation, namely $p(\langle g_j\rangle)$ is mainly due to its use for ranking the population members during the genetic search. In (28), $p(\langle g_j\rangle)$ can admit several interpretations as follows.

- On one hand it is a penalty function obtained by a nonlinear interpolation applied to $\langle g_j\rangle$. In this process, the probabilistic considerations apparently are exercised as a nonlinear transformation to the penalty function $\langle g(x_j)\rangle$ to obtain another penalty function $p(\langle g_j\rangle)$ in order to bring $\langle g(x_j)\rangle$ from an infinite range to a finite range namely, between zero and unity.

- As another interpretation, the penalty function $p(\langle g_j\rangle)$ is the probability of a random variable G, namely cumulative probability of an exponentially distributed random variable.

- Yet another interpretation is to consider $p(\langle g_j\rangle)$ as another stochastic variable Y_j obtained from a function of stochastic variable $X_j=\langle g_j\rangle$.

The last interpretation is highlighted in this work so that several essential implications can be derived. For this aim first we consider the premise given by (21). The implication of this premise can be seen as follows.

Let us define

$$p\left(\left\langle g_j\right\rangle\right) = H\left(\left\langle g_j\right\rangle\right) \tag{32}$$

where $H(\langle g_j\rangle)$ is a function of random variable given by (21), $\langle g_j\rangle$ being the random variable in question.

$$\begin{aligned} p\left(\left\langle g_j\right\rangle\right) = H\left(\left\langle g_j\right\rangle\right) &= \int_0^{\langle g_j\rangle} \lambda_j e^{-\lambda_j\langle g_j\rangle} dg_j \\ &= 1 - e^{-\lambda_j\langle g_j\rangle} \end{aligned} \tag{33}$$

where

$$\lambda_j = \frac{1}{\left\langle \overline{g}_j\right\rangle} \tag{34}$$

The probability density of this random variable is exponential density function given by (16). The probability density $f_p(p)$ of a new random variable p is given by

$$f_p(p) = \frac{f_{g_j}\left(\left\langle g_j\right\rangle\right)}{\left| \dfrac{dH\left(\left\langle g_j\right\rangle\right)}{dg_j} \right|_{\langle g_j\rangle = H^{-1}(p)}} \tag{35}$$

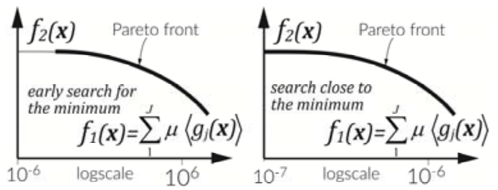
that gives the obvious result

$$f_p(p) = 1 \qquad 0 \le p \le 1 \tag{36}$$

which is a uniform pdf. That is, (21) implies the uniform probability density of p. The important implication of this result will be presented in the following section.

IV. Implications of the Probabilistic Modeling

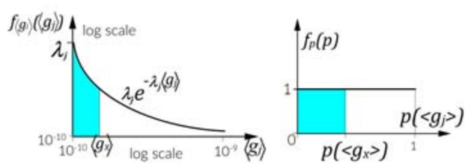
Adaptive zooming for ranking with precision is accomplished by accurate computation of $p(\langle g_j\rangle)$ in the range zero and unity as probabilistic distances, even though the actual constraint $\langle g_j(x)\rangle$ values may be close to the minimal point as much as the computer precision can allow, say at the range of 10^{-10}. To illustrate this, a sketch of the Pareto front at the early stage of the genetic search is shown in figure 4a. A sketch of the Pareto front at the last stage of the genetic search is given in figure 4b. The shape of the curves is because of the log scale.



(a) (b)

Fig. 4. Sketch of formation of the Pareto front at the early stage (a); at the last stage of the GA search (b).

The probabilistic distance to the minimum is illustrated as a typical example in figure 5a by the indicated area where the computation of the shaded area 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 5b by taking simply $p(\langle g_j \rangle)$ as the probability distance to the minimum. The indicated shaded areas in



(a) (b)

Fig. 5. Mathematical lense; pdf of the violations in the objective functions space (a); in the probabilistic space (b)

figures 5a and 5b are the same. This means if the constraint $g_j(\langle x \rangle)$ can be close to the optimal point in a micro scale, say in the range of 10^{-10}, as shown in figure 5a the penalty function $p(\langle g_x \rangle)$ takes place always in a macro scale in the range of between 0 and unity, as shown in figure 5b. This situation is equivalent to applying a commensurate magnifying glass to the space formed by actual objective function and the constraints functions to carry out the convergence process without being effected by any scale of convergence happening in this $\langle g_j \rangle$ space. More precisely, in the micro-scale the genotype of the chromosomes is limited to a narrower favourable region determined by $\langle g_j(x) \rangle$, the corresponding accurate computation of $p_j(\langle g_j \rangle)= c_j r_j(\langle g_j \rangle)\langle g_j \rangle$ in (26) is computed equivalently by (21) always. This is the crucial point to see not only how the method works but also why it performs an improved convergence. Namely, the computation of p_j is straightforward by (21) rather than by the precarious product given by $p_j(\langle g_j \rangle)=c_j r_j(\langle g_j \rangle)\langle g_j \rangle$. With the probabilistic distance measure we obtain robust progress for convergence at each generation due to improved population forming for a new generation in the optimization process. This is independent of the method of evolutionary algorithm being used, although in the research we have used NSGA-II in combination with a local search.

In the course of the generations' production the change of the tangent in (10) is favourably adjusted. One can see this from (29) by considering

$$r_j = \frac{p(\langle g_j \rangle)}{c\,\langle g_j \rangle} = \frac{1-\mathrm{e}^{-\lambda_j \langle g_j \rangle}}{c\,\langle g_j \rangle} \qquad (37)$$

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

$$\lim_{\langle g_j \rangle \to 0} \; r_j = \frac{p(\langle g_j \rangle)}{c_j \langle g_j \rangle} = \lim_{\langle g_j \rangle \to 0} \frac{\lambda_j e^{-\lambda_j \langle g_j \rangle}}{c_j} \lim_{\langle g_j \rangle \to 0} \frac{\lambda_j}{c_j} \to \infty \qquad (38)$$

which indicates the variation of the penalty parameter r_j during the convergence. In the limiting case to the minimum, i.e. P_{opt} in figure 3 r_j goes to infinity, as one should expect. Explicitly, the penalty parameter r_j goes to infinity as $\langle g_j \rangle$ goes to zero being dependent on the decay parameter of the exponential function given by (20). c_j is given by (28) and for this limiting case is determined to be

$$\lim_{\langle g_j \rangle \to 0} \; c_j \to \frac{1}{t} = \frac{1}{P_{opt}} \qquad (39)$$

Please note that the above described probabilistic computations are the main machinery of the effectiveness of the probabilistic constrained handling due to the accurate computation of $p(\langle g_j \rangle)$ in (26). Otherwise the same computation is problematic because of the precarious product involved. This can be noted easily by considering a limiting case. Namely, while $r_j \to \infty$ then $\langle g_j \rangle \to 0$ so that the product $p_j \langle g_j \rangle)= c_j r_j(\langle g_j \rangle)\langle g_j \rangle$ in (26) becomes undetermined.

It is also to note that the above considerations to compute p_j corroborate the premise given by (21) by which p_j is computed, as it should be.

V. CONCLUSIONS

The details and the implications of a new probabilistic approach for multiobjective evolutionary optimization and constrained single objective optimization are presented. Conventionally the problem is handled in the form of single objective and the sum of constraints. This means, the essential optimization process is focused on the constraints during the optimal front formation. This is due to the involvement of the sum of a number of constraints. As consequence the single objective is minimally attended, so that progress with regards to its minimization is relatively poor. As result, conventionally in this problem formulation evolutionary computation has to be supported by auxiliary local search algorithms. The new methodology is briefly presented in an earlier work which is centered for applications and a marked improvement is achieved[1]. The herewith reported details and the implications of the probabilistic constraint handling approach can be exclusively summarized as follows. Firstly, the new approach can work as a mathematical lens where the characteristic exponential probability distribution of the constraint violations remains the same. The implication of this is the adaptive decay parameter computation in concert with the constraint violation yielding a continuous and stable convergence during the search process. In this way the same convergence effectiveness during the search is preserved, being independent of the level of convergence to the optimum, i.e., number of generations. This means the method forms a dynamic "lens," the magnifying power of which is commensurate with the scale of convergence. That is, the convergence is accomplished effectively and systematically, at any range allowed by machine or genotype coding precision. Relative to the conventional approach, the method shows outstandingly better performance as to precision as well as accuracy, approaching to the solution. Secondly, the analytical form of the Pareto front is approximately determined by (14) that can be of interest providing more insight into the convergence properties of the algorithm used. Thirdly, the dependence of each individual constraint penalty parameter on the objectives is established by (24). The limit of each such constraint penalty parameter is established by (38). This can also be of interest providing more insight into the convergence properties of the algorithm used. The research is an important account of a new probabilistic method providing the essentials which explain not only how the method works, but also why it performs better in the context of join-classical optimization for instance. Therefore the work is an important follow-up study which makes the related earlier research with local search [1] appreciable and accessible for everyone easily. At the same time, the effectiveness of the basic form of the algorithm presented in [2], where local search is omitted and precision optimization is accomplished by evolutionary computation alone, is explained, as well.

APPENDIX A

For the development of an envelope for a family of curves, for each value of t the relation $F(x,y,t)=0$ defines a curve in the x y plane. The total collection of such curves forms a family of curves. Some families of curves possess an envelope that is a curve which touches each member of a family. The envelope may be determined as the solutions to the simultaneous equations

$$F(x,y,t)=0,\ \ F_t^{'}(x,y,t)=0 \qquad (40)$$

In this work $F(x,y,t)$ is given by

$$\begin{aligned} F(x,y,t) &= \frac{y}{t}+\frac{x}{P_{opt}-t}-1=0 \\ F_t^{'}(x,y,t) &= -\frac{y}{t^2}+\frac{x}{\left(P_{opt}-t\right)^2}=0 \end{aligned} \qquad (41)$$

From these two equations above, we obtain

$$t=y+\sqrt{xy} \qquad (42)$$

The substitution of (42) into (41) yields

$$(x-y)^2-2(x+y)+1=0 \qquad (43)$$

REFERENCES

[1] R. Datta, M. S. Bittermann, K. Deb, and O. Ciftcioglu, "Probabilistic constraint handling in the framework of joint evolutionary-classical optimization with engineering applications," presented at the IEEE Congress on Evolutionary Computation, Brisbane, Australia, 2012.

[2] O. Ciftcioglu and M. S. Bittermann, "Precision constrained optimization by exponential ranking," presented at the IEEE World Congress on Computational Intelligence - WCCI 2016, Vancouver, Canada, 2016 (under review for publication in this conference).

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

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

[8] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms," Reliability Engineering and Systems Safety vol. 91, pp. 992-1007, 2006.

[9] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert, "Performance assessment of multiobjective optimizers:An analysis and review," IEEE Trans. on Evolutionary Computation, vol. 7, pp. 117-132, 2003.

[10] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," IEEE Transactions on Evolutionary Computation, vol. 7 pp. 117 - 132 2003

[11] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach," IEEE Trans. Evolutionary Computation, vol. 3, pp. 257-271, 1999.

[12] D. A. V. Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithms: Analyzing the state-of-the-art," Evolutionary Computation, vol. 8, pp. 125-147, 2000.

[13] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization " Evolutionary Computation, vol. 10, pp. 262-282, 2002.

[14] D. C. J. Knowles, K. Deb (Eds.), , Springer, 2008. , "Multiobjective problem solving from nature: From concepts to applications."

[15] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Diversity improvement by non-geometric binary crossover in evolutionary multiobjective optimization," IEEE Transactions on Evolutionary Computation, vol. 14, pp. 985 - 998, 2010

[16] J. Horn, N. Nafploitis, and D. E. Goldberg, "A niched pareto genetic algorithm for multiobjective optimization," in First IEEE Conf. on Evolutionary Computation, 1994, pp. 82-87.

[17] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization and multiple constraint handling with evolutionary algorithms-part ii: Application example," IEEE Trans. Syst., Man, Cybern, vol. 28, pp. 38-47, 1998.

[18] M. Farina, K. Deb, and P. Amato, "Dynamic multiobjective optimization problems: Test cases, approximations, and applications " IEEE Transactions on Evolutionary Computation, vol. 8, pp. 425 - 442, 2004

[19] M. S. Bittermann and O. Ciftcioglu, "Architectural design computing supported by multi-objective optimization," presented at the IEEE Congress Evolutionary Computation - CEC 2015, Sendai, Japan, 2015.

[20] Ö. Ciftcioglu and M. S. Bittermann, "Adaptive formation of pareto front in evolutionary multi-objective optimization," in Evolutionary computation, W. P. d. Santos, Ed., ed Vienna: In-Tech, 2009, pp. 417-444.

[21] N. Nassif, S. Kajil, and R. Sabourin, "Evolutionary algorithms for multi-objective optimization in hvac system control strategy," in NAFIPS '04, 2004, pp. 51-56.
[22] K. Deb, J. Sundar, U. Bhaskara, and S. Chaudhuri, "Reference point based multi-objective optimization using evolutionary algorithm," Int. J. Comp. Intelligence Research, vol. 2, pp. 273-286, 2006.
[23] 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.
[24] J. Branke, T. Kaussler, and H. Schmeck, "Guiding multi-objective evolutionary algorithms towards interesting regions," AIFB University of Karlsruhe, Germany2000.
[25] R. Datta and K. Deb, Evolutionary constrained optimization: Springer, 2015.
[26] K. Deb, "An efficient constraint handling method for genetic algorithms.," Computer Methods in Applied Mechanics and Engineering, vol. 186, p. 28, 2000.
[27] 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.
[28] C. A. C. Coello, "Use of a self-adaptive penalty approach for engineering optimization problems," Ciomputers in Industry, vol. 41, pp. 113-127, 2000.
[29] H. J. C. Barbosa, A. C. C. Lemonge, and S. B. Heder, "A critical review of adaptive penalty techniques in evolutionary computation," in Evolutionary constrained optimization, R. Datta and K. Deb, Eds., ed: Springer, 2015, pp. 1-27.
[30] S. Gass and T. Saaty, "The computational algorithm for the parametric objective function," Naval Research Logistics Quarterly, vol. 2, p. 7, 1955.
[31] L. Zadeh, "Non-scalar-valued performance criteria," IEEE Trans. Automatic Control, vol. 8, p. 2, 1963.
[32] K. Miettinen, Nonlinear multiobjective optimization. Boston: Kluwer Academic, 1999.
[33] 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.
[34] G. Bachman and L. Narici, Functional analysis. New York: Dover, 2000.
[35] J. T. Oden and L. F. Demkowicz, Applied functional analysis: CRC Press, 1996.