An Optimality Theory Based Proximity Measure for Set Based Multi-Objective Optimization

Kalyanmoy Deb, Fellow, IEEE and Mohamed Abouhawwash
Department of Electrical and Computer Engineering
Computational Optimization and Innovation (COIN) Laboratory
Michigan State University, East Lansing, MI 48824, USA
Email: kdeb@egr.msu.edu, mhawwash@msu.edu

COIN Report Number 2014015

***Abstract*—Set-based multi-objective optimization methods, such as evolutionary multi-objective optimization (EMO) methods, attempt to find a set of Pareto-optimal solutions, instead of a single optimal solution. To evaluate these algorithms from their convergence to the efficient set, the current performance metrics require the knowledge of the true Pareto-optimal solutions. In this paper, we develop a theoretically-motivated KKT proximity measure that can provide an estimate of the proximity of a set of trade-off solutions from the true Pareto-optimal solutions without any prior knowledge of them. Besides theoretical development of the proposed metric, the proposed KKT proximity measure is computed for iteration-wise trade-off solutions obtained from specific EMO algorithms on two, three, five, and 10-objective optimization problems. Results amply indicate the usefulness of the proposed KKT proximity measure as a metric for evaluating different sets of trade-off solutions and also as a possible termination criterion for an EMO algorithm. Other possible uses of the proposed metric are also highlighted.**

I. Introduction

Multi-objective optimization problems give rise to a set of trade-off optimal solutions, known as Pareto-optimal solutions [1], [2]. In solving these problems, one popular approach has been to first find a representative set of Pareto-optimal solutions and then use higher-level information involving one or more decision-makers to choose a preferred solution from the set. Evolutionary multi-objective optimization (EMO) methodologies follow this principle of solving multi-objective optimization problems [2] and have received extensive attention for the past two decades. Many books, journal and conference papers, and public-domain and commercial softwares are available for make EMO an emerging field of research and application.

From an optimization point of view, the main task of a set based optimization methodology is to find, not one, but a set of trade-off solutions at the end of a simulation run. Since multiple solutions are targets for a set based methodology, it has always been a difficulty to evaluate the performance of such algorithms and therefore develop a termination criterion for ending a simulation run. In the EMO literature, for example, several performance metrics, such as hypervolume measure [3], [4] and inverse generational distance (IGD) measure [5], are suggested, but they are not appropriate particularly for the purpose of terminating a simulation run. For the hypervolume measure, there is no pre-defined target that can be set for an arbitrary problem, thereby making it difficult to determine an expected hypervolume value for terminating a run. On the other hand, the IGD measure requires the knowledge of true Pareto-optimal solutions and their corresponding objective values and hence is not applicable to an arbitrary problem.

For terminating a simulation run, it is ideal if some knowledge about the proximity of the current solution(s) from the true optimal solution(s) can be obtained. For single-objective optimization algorithms, recent studies on approximate Karush-Kuhn-Tucker (A-KKT) points have been suggested [6], [7], [8], [9], [10]. The latter studies have also suggested a *KKT proximity measure* that monotonically reduced to zero as the iterates approach a KKT point for a single-objective constrained optimization problem. On many standard constrained test problems, the study has shown that the population-best solutions of a real-parameter genetic algorithm (RGA) – an evolutionary optimization procedure [2] – demonstrate a reducing trend to zero with RGA iterations. In this paper, we extend the definition of an A-KKT point for multi-objective optimization using the achievement scalarizing function concept proposed in the multiple criterion decision making (MCDM) literature and suggest a KKT proximity measure that can suitably used for evaluating an EMO's performance in terms of convergence to the KKT points. The proposed KKT proximity measure can also be used as a termination criterion for an EMO. The proposed KKT proximity measure provides an uniform value for solutions on a non-dominated front parallel to the true efficient front, thereby providing an equal metric value near Pareto-optimal solutions.

In the remainder of the paper, we first describe the principle of classical point-by-point and set based methods for solving multi-objective optimization problems in Section II. Thereafter, we present the KKT proximity measure concept developed for single-objective optimization problems in Section III. Section IV develops the KKT proximity measure concept for multi-objective optimization by treating every trade-off solution as an equivalent achievement scalarizing problem. The concept is then applied on an illustrative problem to test its working. In Section V, the KKT proximity measure is computed for the entire objective space on a well-known two-

objective ZDT test problems [11] to illustrate that as non-dominated solutions approach the Pareto-optimal set front-wise, the KKT proximity measure reduces to zero. Thereafter, KKT proximity measure is computed for trade-off sets found using specific EMO algorithms, such as NSGA-II [12] and NSGA-III [5], [13] on two, three, five and 10-objective test problems. Three engineering design problems are also used for the above purpose. The use of the proposed KKT proximity measure as a termination criterion is then evaluated on all the above problems in Section VI by performing an extensive computer simulation study. The section also discusses a number of viable extensions of this study. Finally, conclusions are drawn in Section VII.

II. Multi-Objective Optimization

Let us consider a n-variable, M-objective optimization problem with J inequality constraints:

$$\begin{array}{rl} \text{Minimize}_{(\mathbf{x})} & \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x})\}, \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \ldots, J. \end{array} \tag{1}$$

Variable bounds of the type $x_i^{(L)} \leq x_i \leq x_i^{(U)}$ can be splitted into two inequality constraints: $g_{J+2i-1}(\mathbf{x}) = x_i^{(L)} - x_i \leq 0$ and $g_{J+2i}(\mathbf{x}) = x_i - x_i^{(U)} \leq 0$. Thus, in the presence of all n pairs of specified variable bounds, there are a total of $J+2n$ inequality constraints in the above problem. We do not consider equality constraints in this paper, but our analysis can be easily modified to consider them as well.

For a multi-objective optimization, an optimal solution (known as a *Pareto-optimal* solution) is defined as a solution that is not worse than any other solution in the search space in all objectives and is strictly better in at least one objective. For a problem having multiple conflicting objectives, usually a number of Pareto-optimal solutions, instead of one, exists. When their objective vectors are computed, they often line up as a front in the M-dimensional objective space. The respective front is called the *efficient front* [1], [2].

A. Classical Point-by-Point Based Methods

Multi-objective optimization problems are traditionally solved by converting them into a parametric scalarized single-objective optimization problem [1], [14]. In the concept of *generative multi-objective optimization* methods, the resulting single-objective optimization problem is expected to be solved multiple times for different scalarizing parameter values. Most of the scalarized single-objective optimization methods are expected to find a *weak* or a *strict* Pareto-optimal solution (see Figure 1), provided the chosen numerical optimization method is able to find the true optimum of the resulting problem. However, very few methods have the converse property – find every possible Pareto-optimal solution using certain parameter values of the scalarization process. For example, the well-known weighted-sum approach cannot find any Pareto-optimal solution which lies on the non-convex part of the efficient frontier, whereas the ϵ-constraint [14] or the normal constraint method [15] can, in principle, find any Pareto-optimal solution.

However, the main disadvantage of the generative approaches is that they are computationally inefficient, as one application using a particular parameter value of scalarization does not help in solving other applications with different parameter values. This is because each parametric application is an independent optimization process and each Pareto-optimal solution needs to be obtained independently in separate runs and without any help from other runs. A recent study [16] has clearly shown that an evolutionary multi-objective optimization (EMO) approach that attempts to find multiple Pareto-optimal solutions simultaneously in a single simulation run can constitute a *parallel* search of finding multiple solutions simultaneously and is often computationally faster than the generative methods. In the following subsection, we discuss this principle of set-based multi-objective optimization methods for solving multi-objective optimization problems.

B. Set Based Multi-Objective Optimization Methods

Instead of finding a single Pareto-optimal solution at a time, set based multi-objective optimization methods, such an evolutionary multi-objective optimization (EMO) method, attempts to find a set of Pareto-optimal solutions in a single simulation run [2]. An EMO's goal is two-fold:

1) Find a set of well-converged set of trade-off solutions, and
2) Find a set of well-diversified set of solutions across the entire efficient front.

The population approach of an EMO and the flexibilities in its operators allow both the above goals to be achieved for multi-objective problems. EMO algorithms, such as NSGA-II [12] and SPEA2 [17] are well-practiced methods for handling two and three-objective problems. Recent extensions (such as NSGA-III [5] and MOEA/D [18]) are able to handle four and more objectives (even more than 10 objectives). The topic is so important today that handling such large number of objectives is given a separate name – evolutionary *many*-objective optimization.

Figure 1 illustrates the principle of set based multi-objective optimization. In each iteration, a set of points gets operated

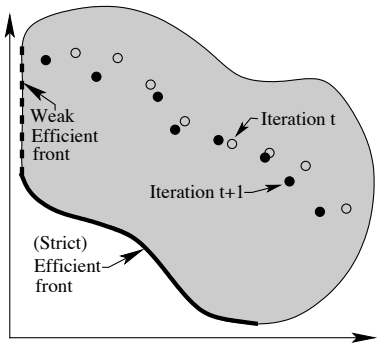

Fig. 1. Principle of set based multi-objective optimization procedure is illustrated.

byan EMO procedure and a new set of points – hopefully towards the strict efficient front and having more diversity – are generated. This process is expected to continue with

iterations before converging on the true efficient front covering its entire range.

When a set based multi-objective optimization algorithm transits from one set of non-dominated points to hopefully a better set of non-dominated points, there is a need to evaluate the convergence and diversity of the new set of solutions vis-a-vis the old solutions. This is because, if we can measure both aspects well through one or more performance metrics, they can also be used to determine the termination of the optimization algorithm. In this paper, we do not consider the diversity aspect much and concentrate on defining a theoretical KKT optimality based convergence measure which will enable us to achieve two aspects:

1) Provide a theoretical underpinning of the working of an algorithm, and
2) Establish a theoretically motivated termination criterion for stopping an algorithm.

III. KKT Proximity Measure for Single-objective Optimization

Dutta et al. [9] defined an approximate KKT solution to compute a KKT proximity measure for any iterate $\mathbf{x}^k$ for a single-objective optimization problem of the following type:

$$\begin{array}{rl} \text{Minimize}_{(\mathbf{x})} & f(\mathbf{x}), \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \ldots, J. \end{array} \tag{2}$$

After a long theoretical calculation, they suggested a procedure of computing the KKT proximity measure for an iterate ($\mathbf{x}^k$) as follows:

$$\begin{array}{rl} \text{Minimize}_{(\epsilon_k, \mathbf{u})} & \epsilon_k \\ \text{Subject to} & \|\nabla f(\mathbf{x}^k) + \sum_{j=1}^{J} u_j \nabla g_j(\mathbf{x}^k)\|^2 \leq \epsilon_k, \\ & \sum_{j=1}^{J} u_j g_j(\mathbf{x}^k) \geq -\epsilon_k, \\ & u_j \geq 0, \quad \forall j, \end{array} \tag{3}$$

However, that study restricted the computation of the KKT proximity measure for feasible solutions only. For an infeasible iterate, they simply ignored the computation of the KKT proximity measure. However, a little thought on the associated KKT optimality conditions reveal the following two additional conditions associated with equality constraints:

$$g_j(\mathbf{x}^k) \geq 0, \quad j = 1, 2, \ldots, J, \tag{4}$$

$$u_j g_j(\mathbf{x}^k) = 0, \quad j = 1, 2, \ldots, J. \tag{5}$$

The first constraint set makes sure that the iterate $\mathbf{x}^k$ is feasible and the second constraint set ensures that the *complementary slackness* condition is satisfied. However, the complementary slackness condition is relaxed in the above KKT proximity measure formulation (Equation 3). In the above problem, the variables are $(\epsilon_k, \mathbf{u})$. The KKT proximity measure was then defined as follows:

$$\text{KKT Proximity Measure}(\mathbf{x}^k) = \epsilon_k^*, \tag{6}$$

where ϵ_k^* is the optimal objective value of the problem stated in Equation 3.

IV. KKT Proximity Measure for Multi-objective Optimization

For the M-objective optimization problem stated in Equation 1, the Karush-Kuhn-Tucker (KKT) optimality conditions are given as follows [19], [1]:

$$\sum_{k=1}^{M} \lambda_k \nabla f_k(\mathbf{x}^k) + \sum_{j=1}^{m} u_j \nabla g_j(\mathbf{x}^k) = \mathbf{0}, \tag{7}$$

$$g_j(\mathbf{x}^k) \leq 0, \quad j = 1, 2, \ldots, J, \tag{8}$$

$$u_j g_j(\mathbf{x}^k) = 0, \quad j = 1, 2, \ldots, J, \tag{9}$$

$$u_j \geq 0, \quad j = 1, 2, \ldots, J, \tag{10}$$

$$\lambda_k \geq 0, \quad k = 1, 2, \ldots, M, \text{ and } \boldsymbol{\lambda} \neq \mathbf{0}. \tag{11}$$

Multipliers λ_k are non-negative, but at least one of them must be non-zero. The parameter u_j is called the Lagrange multiplier for the j-th inequality constraint and they are also non-negative. Any solution $\mathbf{x}^k$ that satisfies all the above conditions is called a KKT point [20]. Equation 7 is called the *equilibrium equation*. Equation 9 for every constraint is called the complimentary slackness equation, as mentioned above. Note that the conditions given in 8 ensure feasibility for $\mathbf{x}^k$ while the equation 10 arises due to the minimization nature of the problem given in equation 2.

A. A Naive Measure from KKT Optimality Conditions

The above KKT conditions can be used to naively define a *KKT error* measure. For a given iterate $\mathbf{x}^k$, the parameters λ-vector and $\mathbf{u}$-vector are unknown. Different values of these parameters will cause the above conditions to be either satisfied completely or not. We can propose a method to identify suitable λ and $\mathbf{u}$-vectors so that the all inequality constraints (conditions 8, 9, 10 and 11) are satisfied and the equilibrium condition (7) is violated the least. Thus, we formulate an optimization procedure to find such optimal ($\boldsymbol{\lambda}$, $\mathbf{u}$)-vectors:

$$\begin{array}{rl} \text{Minimize}_{(\boldsymbol{\lambda}, \mathbf{u})} & \|\sum_{k=1}^{M} \lambda_k \nabla f_k(\mathbf{x}^k) + \sum_{j=1}^{m} u_j \nabla g_j(\mathbf{x}^k)\|, \\ \text{Subject to} & g_j(\mathbf{x}^k) \leq 0, \quad j = 1, 2, \ldots, J, \\ & u_j g_j(\mathbf{x}^k) = 0, \quad j = 1, 2, \ldots, J, \\ & u_j \geq 0, \quad j = 1, 2, \ldots, J, \\ & \lambda_k \geq 0, \quad k = 1, 2, \ldots, M, \text{ and } \boldsymbol{\lambda} \neq \mathbf{0}. \end{array} \tag{12}$$

The operator $\|\cdot\|$ can be the L_2-norm. It is clear that if $\mathbf{x}^k$ is a KKT point, the norm would be zero for the associated and feasible $(\boldsymbol{\lambda}, \mathbf{u})$-vectors. For a non-KKT but feasible $\mathbf{x}^k$ point, the norm need not be zero, but we are interested in investigating if the violation of the KKT conditions can be used a convergence metric in the neighborhood of a KKT point. To achieve this, we attempt to find a suitable $(\boldsymbol{\lambda}, \mathbf{u})$-vector that will minimize the above norm. We then explore if this minimum norm value can behave as an indicator to how close a point is from a suitable Pareto-optimal point. Thus, a KKT Error for a feasible iterate $\mathbf{x}^k$ is defined by solving the above problem and computing the following error value:

$$\text{KKT Error}(\mathbf{x}^k) = \|\sum_{k=1}^{M} \lambda_k^* \nabla f_k(\mathbf{x}^k) + \sum_{j=1}^{m} u_j^* \nabla g_j(\mathbf{x}^k)\|, \tag{13}$$

where λ_k^* and u_j^* are optimal values of the problem given in Equation 12.

In order to investigate whether the above KKT error can be used as a metric for evaluating *closeness* of an iterate $\mathbf{x}^k$ to a Pareto-optimal solution, we consider a simple bi-objective optimization problem, given as follows:

$$P1 : \left\{ \begin{array}{ll} \text{Minimize}_{(x,y)} & \{x, \frac{1+y}{1-(x-0.5)^2}\}, \\ \text{Subject to} & 0 \leq (x,y) \leq 1. \end{array} \right. \quad (14)$$

The Pareto-optimal solutions are given as follows: $y^* = 0$ and $x^* \in [0, 0.5]$. The efficient front can be represented as $f_2^* = 1/(1-(f_1^*-0.5)^2))$. We attempt to compute the above KKT error for any given point or iterate $\mathbf{x}^k = (x^k, y^k)^T$ exactly by solving the following optimization problem, considering two inequality constraints $g_1(\mathbf{x}) = -x^k \leq 0$ and $g_2(\mathbf{x}) = -y^k \leq 0$:

$$\begin{array}{ll} \text{Min.}_{(\lambda_1,u_1,u_2)} & \epsilon = \|\lambda \nabla f_1 + (1-\lambda_1)\nabla f_2 + u_1 \nabla g_1 + u_2 \nabla g_2\|, \\ \text{Subject to} & -x^k \leq 0, \quad -y^k \leq 0, \\ & -u_1 x^k = 0, \quad -u_2 y^k = 0, \\ & u_1 \geq 0, \quad u_2 \geq 0, \\ & 0 \leq \lambda_1 \leq 1. \end{array} \quad (15)$$

The use of $\lambda_2 = 1 - \lambda_1$ and restricting λ_1 to lie within $[0, 1]$ ensures that (i) both λ values are non-negative and (ii) both are not zero at the same time. The optimal solution of the above problem can be computed exactly, as follows:

1) Case 1: When $x^k > 0$ and $y^k > 0$: This means that both u_1 and u_2 are zero. Thus, the minimum ϵ^* can be calculated as:
$$\epsilon^* = \frac{K_2}{\sqrt{K_2^2 + (1-K_1)^2}}, \quad (16)$$
where
$$\begin{array}{rcl} K_1 & = & \dfrac{2(x^k - 0.5)(1+y^k)}{(1-(x^k-0.5)^2)^2}, \\ K_2 & = & \dfrac{1}{1-(x^k-0.5)^2}. \end{array}$$
Corresponding $\lambda_1^* = \frac{K_1^2 + K_2 - K1}{K_2^2 + (1-K_1)^2}$.
2) Case 2: When $x > 0$ and $y = 0$: This means that $u_1 = 0$. The minimum error can be founds as $\epsilon^* = 0$ with $\lambda_1^* = \frac{K_1}{K_1 - 1}$ and $u_2^* = \frac{K_2}{1-K_1}$.
3) Case 3: When $x = 0$ and $y > 0$: This means $u_2 = 0$. The minimum error can be founds as $\epsilon^* = 0$ with $\lambda_1^* = 1$ and $u_1^* = 1$.
4) Case 4: When $x = 0$ and $y = 0$: The minimum error can be computed as $\epsilon^* = 0$ with $u_1^* = \lambda(1-K_1) + K_1$ and $u_2^* = (1-\lambda_1)K_2$ for any $\lambda_1 \in [0, 1]$.

We consider 10,000 equi-spaced points in the variable space and compute the above KKT error value for each of them. The KKT error value is then plotted in Figure 2. It is clear that the KKT error is zero for all Pareto-optimal solutions ($y = 0$), but surprisingly the KKT error value increases as the points get closer to the Pareto-optimal solution, as shown in Figure 3. This figure is plotted for a fixed value of x: $x = 0.2$, but for different values of y. The Pareto-optimal solution along this line corresponds to the point $(0.2, 0)^T$. The KKT error at this point $(0.2, 0)^T$ is zero, as it is evident from Figure 3, but it jumps to a large value in its vicinity. Ironically, the figure depicts that for points away from this Pareto-optimal point the KKT error gets smaller. A similar naive approach [21] was also found to produce near zero KKT error for feasible solutions far away from the Pareto-optimal solutions, thereby making the naive KKT error measure as an unreliable metric for estimating convergence pattern to the Pareto-optimal solutions.

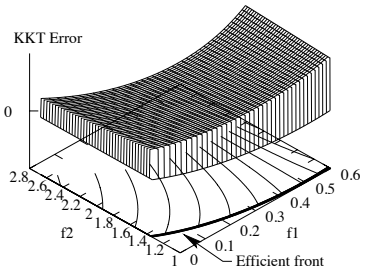

Fig. 2. KKT Error value increases towards the efficient front in Problem P1.

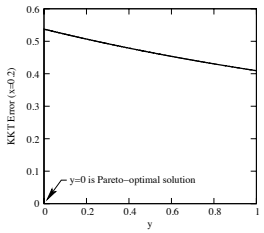

Fig. 3. KKT Error value increases towards the Pareto-optimal point along a line $x = 0.2$ for Problem P1.

The above phenomenon suggests that KKT optimality condition is a *singular* set of conditions valid only at the KKT points. A minimal violation of equilibrium KKT condition (we describe here as a 'naive' measure) cannot provide an idea about how close a solution is to a KKT point in general. To investigate the suitability of the above KKT Error metric as a potential termination condition or as a measure for judging closeness of obtained non-dominated solutions to the true Pareto-optimal solutions, we consider the problem stated in Equation 14 and attempt to solve it using NSGA-II procedure [12]. The problem has two variables, two objectives, and four constraints. NSGA-II is used with standard parameter settings: population size 20, SBX operator [22] with probability $p_c = 0.9$ and distribution index (η_c) of 20 and polynomial mutation operator [2] with probability $p_m = 0.5$ and index $\eta_m = 50$. The non-dominated solutions from each generation is extracted and the optimization problem stated in Equation 15 is solved using Matlab's `fmincon()` routine. Non-dominated solutions at various generations are shown in Figure 4. It can be seen that with generations the non-dominated solutions approach the efficient frontier. At generation 50, the population is well spread over the entire efficient frontier.

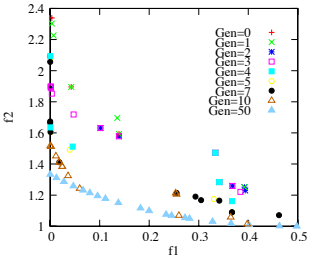

Fig. 4. NSGA-II obtained non-dominated solutions approach the efficient frontier with generations for Problem P1.

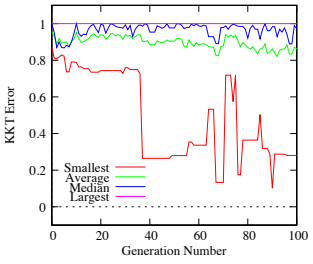

Fig. 5. Variation of smallest, average, median and largest KKT Error metric values for NSGA-II obtained non-dominated points for Problem P1.

The resulting KKT Error (ϵ^*) for each non-dominated solution is recorded and the minimum, mean, median, and maximum KKT Error value is plotted in Figure 5 at every generation. It is interesting to observe that although the smallest KKT Error value reduces in the initial few generations, the variation of the KKT Error value is too large even up to generation 100 for it to be considered as any stable performance metric. On the contrary, Figure 4 shows that the non-dominated solutions approach the true efficient front with generations. NSGA-II uses an elite preserving operator in emphasizing non-dominated and less-crowded (based on the crowding distance operator [12]) solutions, but the above figure suggests that although better non-dominated and distributed points are continuously found by NSGA-II (as Figure 4 also supports), their satisfaction of KKT optimality conditions need not be improved with generations. This behavior of the KKT Error measure does not allow us to use it to either as a performance metric or as a termination condition. However, the KKT proximity measure suggested in earlier studies [9], [10] for single-objective optimization problems relaxed both the equilibrium condition and complementary slackness condition, thereby enabling us to derive a monotonic KKT proximity measure in that study. In the following section, we extend the KKT proximity metric concept to multi-objective optimization and investigate if it can alleviate the difficulties with the KKT error metric studied above.

B. Proposed KKT Proximity Measure

One of the common ways to solve the generic multi-objective optimization problem stated in Equation 1 is to solve a parameterized *achievement scalarization function (ASF)* optimization problem repeatedly for different parameter values. The ASF approach was originally suggested by Wierzbicki [23]. For a specified reference point $\mathbf{z}$ and a weight vector $\mathbf{w}$ (parameters of the ASF problem), the ASF problem is given as follows:

$$\begin{array}{ll} \text{Minimize}_{(\mathbf{x})} & \text{ASF}(\mathbf{x},\mathbf{z},\mathbf{w}) = \max_{i=1}^{M}\left(\frac{f_i(\mathbf{x})-z_i}{w_i}\right), \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0, \quad j=1,2,\ldots,J. \end{array} \tag{17}$$

The reference point $\mathbf{z} \in R^M$ is any point in the M-dimensional objective space and the weight vector $\mathbf{w} \in R^M$ is an M-dimensional unit vector for which every $w_i \geq 0$ and $\|\mathbf{w}\| = 1$. To avoid division by zero, we shall consider strictly positive weight values. It has been proven that for above conditions of $\mathbf{z}$ and $\mathbf{w}$, the solution to the above problem is always a Pareto-optimal solution [1]. Figure 6 illustrates the ASF procedure of arriving at a weak or a strict Pareto-optimal solution. For illustrating the working principle of the ASF procedure, we consider specific reference vector $\mathbf{z}$ (marked in the figure) and weight vector $\mathbf{w}$ (marked as $\mathbf{w}$in the figure). For any point $\mathbf{x}$, the objective vector $\mathbf{f}$ is computed (shown as F). Larger of two quantities ($(f_1 - z_1)/w_1$ and $(f_2 - z_2)/w_2$) is then chosen as the ASF value of the point $\mathbf{x}$. For the point G, the above two quantities are identical, that is, $(f_1 - z_1)/w_1 = (f_2 - z_2)/w_2 = p$ (say). For points on line GH, the first term dominates and the ASF value is identical to p. For points on line GK, the second term dominates and the ASF value is also identical to p. Thus, it is clear that for any point on lines GH and GK, the corresponding ASF value will be the same as p. This is why an iso-ASF line for an objective vector at F traces any point on lines KG and GH. For another point A, the iso-ASF line is shown in dashed lines passing through A, but it is important to note that its ASF value will be larger than p, since its iso-ASF lines meet at the $\mathbf{w}$-line away from the direction of minimum of both objectives. Since the ASF function is minimized, point F is considered better than A in terms of the ASF value. Similarly, point A will be considered better than point B. A little thought will reveal that when the ASF problem is optimized the point O will correspond to the minimum ASF value in the entire objective space (marked with a shaded region), thereby making point O (an efficient point) as the final outcome of the optimization problem stated in Equation 17 for the chosen $\mathbf{z}$ and $\mathbf{w}$ vectors. By keeping the reference point $\mathbf{z}$ fixed and by changing the weight vector $\mathbf{w}$ (treating it like a parameter of the resulting scalarization process), different points on the efficient front can be generated by the above ASF minimization process.

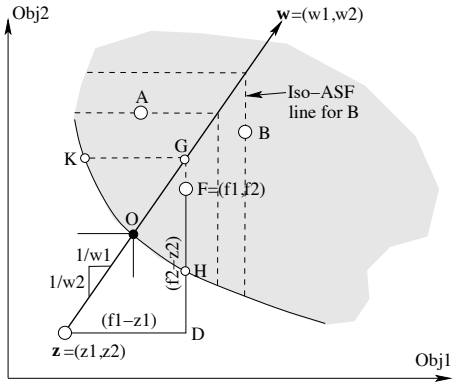

Fig. 6. ASF procedure of finding a Pareto-optimal solution is illustrated.

For our purpose of estimating a KKT proximity measure for multi-objective optimization, we fix the reference point $\mathbf{z}$ to an utopian point: $z_i = z_i^{\text{ideal}} - \epsilon_i$ for all i, where z_i^{ideal} is the ideal point and is set as the optimal function value of the following minimization problem:

$$\begin{array}{ll} \text{Minimize}_{(\mathbf{x})} & f_i(\mathbf{x}), \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0, \quad j=1,2,\ldots,J. \end{array} \tag{18}$$

Thus, the above optimization task needs to be performed M times to obtain the M-dimensional $\mathbf{z}$-vector. For all our computations of this paper, we set $\epsilon_i = 0.01$. Note that $\epsilon_i = 0$ can also be chosen.

Setting the reference point as an utopian vector, we are left with a systematic procedure for setting the weight vector $\mathbf{w}$. For this purpose, we compute the direction vector from $\mathbf{z}$ to the objective vector $\mathbf{f} = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x}))^T$ computed at the current point $\mathbf{x}$ for which the KKT proximity measure needs to be computed. The weight value for the i-th objective function is computed as follows:

$$w_i = \frac{f_i(\mathbf{x}) - z_i}{\sqrt{\sum_{k=1}^{M}(f_k(\mathbf{x}) - z_k)^2}}. \tag{19}$$

In the subsequent analysis, we treat the above $\mathbf{w}$-vector fixed for a given iterate $\mathbf{x}^k$ and set $\mathbf{w}^k = \mathbf{w}$. Figure 6 illustrates the reference point and the weight vector. Notice, how the ASF optimization problem results in a corresponding Pareto-optimal solution. Thus, the idea of the above formulation of the ASF problem is that if $\mathbf{x}^k$ under consideration is already a Pareto-optimal solution, the solution to the resulting ASF optimization problem will be the same solution and the KKT optimality conditions are expected to be satisfied. However, if the point $\mathbf{x}$ is away from the efficient frontier, our proposed KKT proximity measure is expected to provide a monotonic metric value related to the proximity of the point from the corresponding ASF optimal solution.

To formulate the KKT proximity measure, first, we use a reformulation of optimization problem stated in Equation 17 into a smooth problem:

$$\begin{array}{rl} \text{Min.}_{(\mathbf{x},x_{n+1})} & F(\mathbf{x},x_{n+1}) = x_{n+1} \\ \text{Subject to} & \left(\frac{f_i(\mathbf{x})-z_i}{w_i^k}\right) - x_{n+1} \leq 0, \quad i = 1,2,\ldots,M, \\ & g_j(\mathbf{x}) \leq 0, \quad j = 1,2,\ldots,J. \end{array} \tag{20}$$

Since we plan to use Matlab's `fmincon()` optimization algorithm for solving the ASF minimization problem, this conversion is helpful. A new variable x_{n+1} is added and M additional inequality constraints arise to make the problem smooth [1]. Thus, to find the corresponding Pareto-optimal solution, the above optimization problem has $(n+1)$ variables: $\mathbf{y} = (\mathbf{x}; x_{n+1})$. For an ideal or an utopian point as $\mathbf{z}$, the optimal solution to the above problem is expected to make $x^*_{n+1} \geq 0$.

Since the above optimization problem is a single-objective problem, we can use the KKT proximity metric discussed in the previous section to estimate a proximity measure for any point $\mathbf{x}$. Note that above problem has $M + J$ inequality constraints:

$$G_j(\mathbf{y}) = \left(\frac{f_j(\mathbf{x}) - z_j}{w_j^k}\right) - x_{n+1} \leq 0, \; j = 1,\ldots,M, \tag{21}$$

$$G_{M+j}(\mathbf{y}) = g_j(\mathbf{x}) \leq 0, \quad j = 1,2,\ldots,J. \tag{22}$$

The KKT proximity measure can now be computed for a given *feasible* $\mathbf{x}^k$ as follows. Note that $\mathbf{x}^k$ is known, but the variable x_{n+1} is unknown. Thus, we first construct an $(n+1)$-dimensional vector $\mathbf{y} = (\mathbf{x}; x_{n+1})$ and formulate the following optimization problem using Equation 20 to compute the KKT proximity measure:

$$\begin{array}{rl} \text{Minimize}_{(\epsilon_k,x_{n+1},\mathbf{u})} & \epsilon_k + \sum_{j=1}^{J} \left(u_{M+j} g_j(\mathbf{x}^k)\right)^2, \\ \text{Subject to} & \|\nabla F(\mathbf{y}) + \sum_{j=1}^{M+J} u_j \nabla G_j(\mathbf{y})\|^2 \leq \epsilon_k, \\ & \sum_{j=1}^{M+J} u_j G_j(\mathbf{y}) \geq -\epsilon_k, \\ & u_j \geq 0, \quad j = 1,2,\ldots,(M+J), \\ & -x_{n+1} \leq 0. \end{array} \tag{23}$$

The added term in the objective function allows a penalty associated with the violation of complementary slackness condition. If the iterate $\mathbf{x}^k$ is Pareto-optimal or a KKT point, the complementary slackness condition will be zero (by either $g_j(\mathbf{x}^k) = 0$ or $u_{M+j} = 0$) and hence the above optimization should end up with a zero objective value. For any other solution, the left side expressions of first two constraints in Equation 23 are not expected to be zero, but the above minimization problem should drive the process to find minimum possible value of ϵ_k and constraint violation. In some sense, the optimal ϵ_k value should provide us with an idea of the extent of violation of KKT equilibrium condition and complementary slackness condition at every iterate $\mathbf{x}^k$. The addition of the penalty term in the objective function allows to have a larger (and worse) KKT proximity measure value for feasible but non-Pareto-optimal solutions. The final constraint is added to make sure a positive ASF value (x_{n+1}) is assigned to every feasible point for an ideal or utopian reference point $\mathbf{z}$.

The optimal objective value (ϵ_k^*) to the above problem will correspond to the proposed KKT proximity measure. Let us now discuss the variables of the above optimization problem. They are ϵ_k, x_{n+1}, and the Lagrange multiplier vector $u_j \in R^{M+J}$. Thus, there are $(M+J+2)$ variables and $(M+2J+2)$ inequality constraints to the above problem. Let us describe the constraints next. The first constraint requires gradient of F and G functions:

$$\left\| \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} + \sum_{i=1}^{M} u_i \begin{pmatrix} \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_1} \\ \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_2} \\ \vdots \\ \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_n} \\ -1 \end{pmatrix} + \sum_{j=1}^{J} u_{M+j} \begin{pmatrix} \frac{\partial g_j(\mathbf{x}^k)}{\partial x_1} \\ \frac{\partial g_j(\mathbf{x}^k)}{\partial x_2} \\ \vdots \\ \frac{\partial g_j(\mathbf{x}^k)}{\partial x_n} \\ 0 \end{pmatrix} \right\| \leq \sqrt{\epsilon_k}. \tag{24}$$

Here, the quantity $\frac{\partial f_i(\mathbf{x}^k)}{\partial x_j}$ is the partial derivative of objective function $f_i(\mathbf{x})$ with respect to variable x_j evaluated at the given point $\mathbf{x}^k$. A similar meaning is associated with the partial derivative of the constraint g_j above. Using the L_2-norm and squaring both sides, we obtain the following constraint:

$$\left\| \sum_{j=1}^{M} \frac{u_j}{w_j^k} \nabla f_j(\mathbf{x}^k) + \sum_{j=1}^{J} u_{M+j} \nabla g_j(\mathbf{x}^k) \right\|^2 + \left(1 - \sum_{j=1}^{M} u_j\right)^2 - \epsilon_k \leq 0. \tag{25}$$

The second constraint in Equation 23 can be rewritten as follows:

$$-\epsilon_k - \sum_{j=1}^{M} u_j \left(\frac{f_j(\mathbf{x}^k) - z_j}{w_j^k} - x_{n+1}\right) - \sum_{j=1}^{J} u_{M+j} g_j(\mathbf{x}^k) \leq 0. \tag{26}$$

For *infeasible* iterates, we simply compute the constraint violation as the KKT proximity measure. Therefore, the KKT proximity measure for any iterate $\mathbf{x}^k$ is calculated as follows:

$$\text{KKT Proximity Measure}(\mathbf{x}^k) = \begin{cases} \epsilon_k^*, & \text{if } \mathbf{x}^k \text{ is feasible,} \\ 1 + \sum_{j=1}^{J} \left\langle g_j(\mathbf{x}^k)\right\rangle^2, & \text{otherwise.} \end{cases} \tag{27}$$

An analysis of the optimization problem stated in Equation 23 and the subsequent simplifications reveal an important result. The optimal KKT proximity measure for feasible solutions is

given as follows:

$$\epsilon_k^* = 1 - \sum_{j=1}^{M} u_j^* - \sum_{j=1}^{J} \left(u_{M+j}^* g_j(\mathbf{x}^k)\right)^2. \quad (28)$$

Since each $u_j^* \geq 0$ and third term is always positive, ϵ_k^* is always bounded in $[0, 1]$ for any feasible solution $\mathbf{x}^k$. To make infeasible solutions to have worse KKT proximity measure value than that for any feasible solution, we add one in the expression of ϵ_k^* value for an infeasible solution in Equation 27. For a KKT point $\mathbf{x}^k = \mathbf{x}^*$, the complementary slackness $u_{M+j}^* g_j(\mathbf{x}^*) = 0$ for all constraints and Equation 24 reduces to $\epsilon_k \geq \left(1 - \sum_{j=1}^{M} u_j\right)^2$. In such a case, along with other conditions including the fact that each u_j is independent, it can be shown that $\epsilon_k^* = 0$, meaning that at the KKT point the KKT proximity measure is always zero.

We now illustrate the working of the above KKT proximity measure on Problem P1, which was used in illustrating the working of the naive KKT error metric earlier. The ideal point for this problem is $\mathbf{z}^{\text{ideal}} = (0, 1)^T$. The variable bounds are converted into four inequality constraints and the optimization problem stated in Equation 23 is solved for 10,000 grid points in the variable space. Each optimization problem is solved using Matlab's fmincon() optimization routine. Figure 7 shows the KKT proximity surface and its contour plot on the objective space. The efficient frontier can be clearly seen from the contour plot. Moreover, it is interesting to note how the contour lines become almost parallel to the efficient frontier, thereby indicating that when a set of non-dominated solutions close to the efficient frontier is found, their KKT proximity measure values will be more or less equal to each other. Also, unlike in the naive approach (shown in Figure 2), a monotonic reduction of the KKT proximity measure towards the efficient frontier is evident from this figure and also from Figure 8 plotted for a fixed value of $x = 0.2$ and spanning y in its range [0,1]. The proposed KKT proximity measure monotonically reduces to zero at the corresponding Pareto-optimal point. A comparison of this figure with 3 clearly indicates that the proposed KKT proximity measure can be used to assess closeness of points to Pareto-optimal points and can also be used to possibly determine a termination of an EMO simulation run reliably.

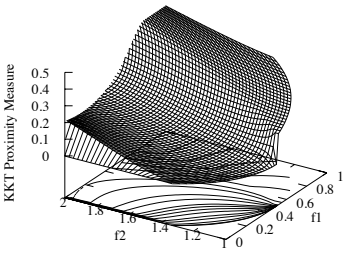

Fig. 7. KKT proximity measure is plotted for Problem P1.

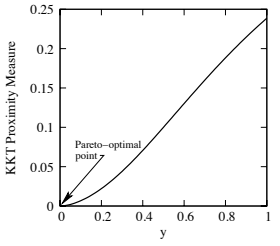

Fig. 8. KKT proximity measure reduces to zero at the Pareto-optimal solution along a line $x = 0.2$ for Problem P1.

To investigate further whether the proposed KKT Proximity metric can be used as a proximity indicator of non-dominated solutions to the true Pareto-optimal solutions, we calculate the smallest, average, median and largest KKT proximity value for all NSGA-II non-dominated solutions generation-wise obtained earlier for Problem P1 and plot them in Figure 9. The inset figure shows a fairly steady decrease in the KKT proximity value with generation to zero, which is a true reflection of the generation-wise NSGA-II non-dominated points shown in Figure 4. This figure clearly indicates the possibility of using the proposed KKT proximity measure as a termination criterion or as a metric for evaluating convergence properties of an EMO algorithm.

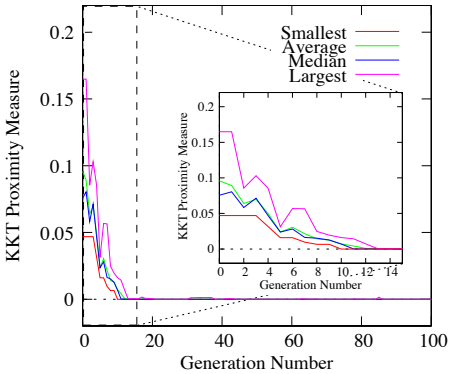

Fig. 9. Smallest, average, median and largest KKT proximity measure values are plotted for generation-wise NSGA-II obtained non-dominated points for Problem P1.

Next, we consider a constrained optimization problem (P2) having two inequality constraints, which may cause an set-based multi-objective optimization algorithm to face both feasible and infeasible points along the way of converging to the Pareto-optimal set:

$$P2: \begin{cases} \text{Minimize}_{(x,y)} & \{x, y\}, \\ \text{Subject to} & g_1(\mathbf{x}) = (x-1)^2 + (y-1)^2 \leq 0.81, \\ & g_2(\mathbf{x}) = x + y \leq 2, \\ & 0 \leq (x, y) \leq 2. \end{cases} \quad (29)$$

The above problem introduces six constraints in total considering the variable bounds as constraints. The Pareto-optimal solutions lie on the first constraint boundary: $y^* = 1 - \sqrt{0.81 - (x^* - 1)^2}$ for $x^* \in [0.1, 1]$. To investigate the KKT proximity measure at different infeasible and feasible points in the search space, we compute the KKT proximity metric value for 10,000 points in the search space and plot the metric surface in Figure 10. Both the surface and the contours of the KKT proximity measure indicates that it is zero at the Pareto-optimal solutions and increases as the points move away from the efficient front. Once again, we observe that the metric value increases monotonically from the feasible region towards infeasible region and also inside the feasible region, the proximity metric value monotonically reduces to zero to the efficient front.

NSGA-II is applied to Problem P2 with identical parameter settings as in Problem P1. To make the problem a bit more difficult, upper bounds on both variables were chosen to be 10, instead of 2. Constrained non-dominated solutions are shown

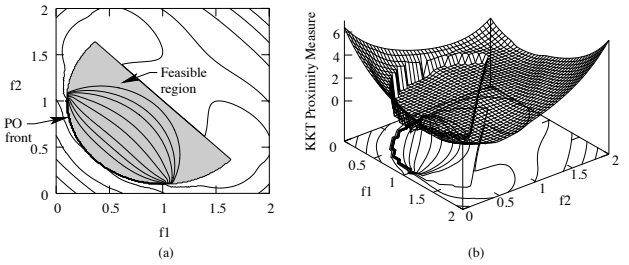

Fig. 10. KKT proximity measure is plotted for Problem P2. The figure (a) shows the feasible space and the contour plot of the KKT proximity measure. The figure (b) shows the KKT proximity measure surface.

in Figure 11. It is observed that up to the sixth generation,

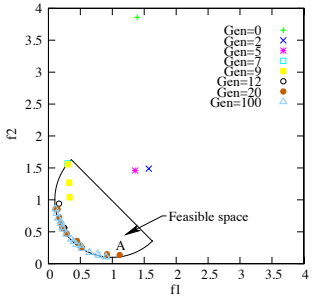

Fig. 11. NSGA-II obtained constrained non-dominated solutions approach the efficient frontier with generations for Problem P2.

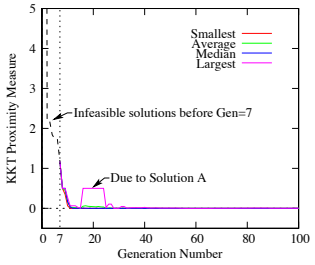

Fig. 12. Variation of smallest, average, median and largest KKT proximity metric values for NSGA-II obtained constrained non-dominated points for Problem P2.

no feasible solution is present in the NSGA-II population. Thereafter, NSGA-II finds solutions in the feasible space and is able to find solutions close to the efficient frontier with generations, as shown in the figure. KKT proximity measure values are then computed for the non-dominated solutions. When no feasible solution exist in a population, the constrained domination principle declares the least constraint violated solution as the constrained non-dominated solution. Figure 12 shows that the KKT proximity value reduces from a large value until generation 6, indicating that infeasible solutions get closer to the feasible region with generations. Thereafter, with more feasible solutions being found, the KKT proximity measure values reduce close to zero. At generation 16, solution A (shown in Figure 11), which is a non-dominated solution to the rest, but is not a Pareto-optimal solution, is created. As can be seen from Figure 10(a), this point has a relatively large KKT proximity value, but when a near-Pareto-optimal solution is found to dominate Solution A at generation 23, the largest KKT proximity measure value reduces. After about 40 generations, all non-dominated points are close to the true efficient front and the the variation in KKT proximity values among these solutions is also small and the KKT proximity values are very close to zero.

The KKT proximity measure on both the above test-problems indicates it can be used as a measure to performance of a set-based multi-objective optimization algorithm. In the following section, we compute the KKT proximity measure for generation-wise sets of non-dominated solutions obtained by different multi and many-objective evolutionary algorithms from the literature.

C. Augmented KKT Proximity Measure to Avoid Weak Pareto-optimal Solutions

The ASF procedure calculated from an ideal or an utopian point may result in a weak Pareto-optimal solution [2], as shown in Figure 13. Consider a weak Pareto-optimal point A

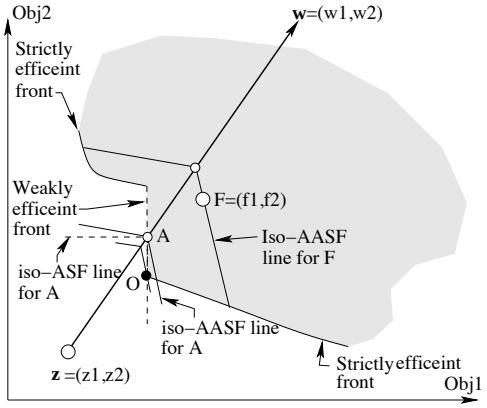

Fig. 13. The augmented achievement scalarizing function (AASF) approach is illustrated.

shown in the figure. The iso-ASF contour line passing through this point will have the same ASF value as for the strict efficient point O. A minimization of the ASF problem for the chosen $\mathbf{z}$ and $\mathbf{w}$-vectors will find any weak Pareto-optimal solutions on the line OA. However, most EMO algorithms are expected to find strict Pareto-optimal solutions (such as point O only), which are not weak. As discussed elsewhere [1], the weak Pareto-optimal solutions can be avoided if an augmented ASF is used instead. We describe the AASF optimization problem as follows:

$$\begin{aligned} \text{Minimize}_{(\mathbf{x})} \quad & \text{ASF}(\mathbf{x},\mathbf{z},\mathbf{w}) = \max_{i=1}^{M}\left(\frac{f_i(\mathbf{x})-z_i}{w_i}\right) \\ & \quad +\rho\sum_{j=1}^{M}\left(\frac{f_i(\mathbf{x})-z_i}{w_i}\right), \\ \text{Subject to} \quad & g_j(\mathbf{x}) \leq 0, \quad j=1,2,\ldots,J. \end{aligned} \tag{30}$$

Here, the parameter ρ takes a small value ($\sim 10^{-4}$). The additional term on the objective function has an effect of making the iso-AASF lines inclined to objective axes, as shown in the figure. The inclination angle is more for larger value of ρ. The iso-AASF line for an arbitrary point F is shown in the figure. A little thought will reveal that with some positive ρ value, the optimal solution to the above AASF problem is the point O, as point A or other weak Pareto-optimal points will make the AASF value larger than that at A, due to its meeting the $\mathbf{w}$-line at a higher point. The difference between ASF and AASF values at point A is clear from the iso-ASF and iso-AASF lines shown in the figure.

The above-mentioned calculation procedure of KKT proximity measure can be modified using the above AASF formulation, which is provided in the appendix of this paper. We shall demonstrate the effect of using the AASF approach compared to the ASF approach on some test problems having weak

Pareto-optimal points. The next section discusses simulation results using the above KKT proximity measures on several multi and many-objective problems from the literature.

V. Results

In this section, we consider two and three-objective test problems to demonstrate the working of the proposed KKT proximity measure with an EMO algorithm. Two-objective problems are solved using NSGA-II [12] and three or more objective problems are solved using NSGA-III [5]. In all problems, we use the SBX recombination operator [22] with $p_c = 0.9$ and $\eta_c = 30$ and polynomial mutation operator [2] with $p_m = 1/n$ (where n is the number of variables) and $\eta_m = 20$. Other parameters are mentioned in discussions on individual problems.

A. Two-Objective ZDT Problems

First, we consider commonly used two-objective ZDT problems. ZDT1 and ZDT2 has 30 variables and the efficient solutions occur for $x_i = 0$ for $i = 2, 3, \ldots, 30$. Before we evaluate the performance of an EMO algorithm on these problems, we first consider ZDT1 problem and 30 variables are set using two parameters: $x = x_1$ and $y = x_2 = x_3 = \ldots = x_{30}$. Thereafter, we calculate the proposed KKT proximity measure for two parameters x and y in the range $[0, 1]$. 10,000 equi-spaced points are chosen i x-y plane to make the plot. Since at $x_1 = 0$, the derivative of f_2 does not exist, we do not compute the KKT Proximity measure for such solutions. Figure 14 clearly indicates the following two aspects:

1) The KKT proximity measure reduces to zero at the efficient frontier.
2) The KKT proximity measure increases almost parallely to the efficient frontier in the local vicinity of the efficient frontier.

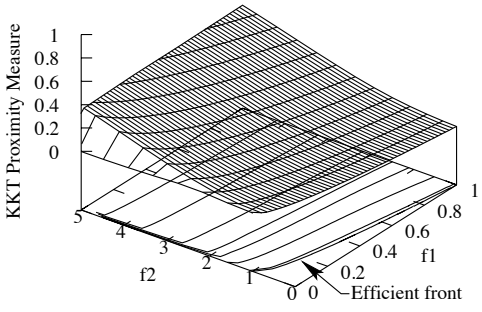

Fig. 14. KKT proximity measure is plotted for Problem ZDT1.

We now apply the KKT proximity measure procedure with ASF formulation to each non-dominated solution obtained by NSGA-II (with a population size of 100) at every generation until 200 generations. Thereafter, the smallest, median, and largest KKT proximity measure value for all non-dominated solutions at each generation is plotted in Figure 15. If the proposed KKT proximity measure is used to terminate the

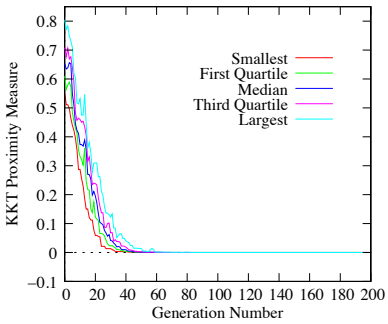

Fig. 15. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for Problem ZDT1.

NSGA-II run, the run would have terminated at around 60-th generation, as the difference between minimum and median KKT proximity measure for all non-dominated points is insignificant thereafter. It is also interesting to note how the KKT proximity measure values reduce to zero with generation. Interestingly, EMO algorithms work without any derivative information or without using any KKT optimality conditions in its operations. Even then, the an EMO algorithm (NSGA-II here) is able to approach the theoretical efficient frontier with increasing generations. The first population member to reach the efficient front took 32 generations, while 50% population members converge within 38 generations and all 100 population members converge to the efficient front in just 59 generations. Not only the KKT proximity measure depicts the dynamics of convergence behavior of NSGA-II on ZDT1, it also indicates that generation span (27 generations) within which a rapid convergence of all population members take place. The close variation of all five indicators in the figure suggests that NSGA-II converges front-wise in the ZDT1 problem.

Next, we consider 30-variable ZDT2 problem. Considering $x = x_1$ and $y = x_2 = x_3 = \ldots = x_{30}$, we compute KKT proximity measure for 10,000 points in the x-y space and resulting measure values are plotted on the ZDT2 objective space in Figure 16. ZDT2 has a non-convex efficient front, as

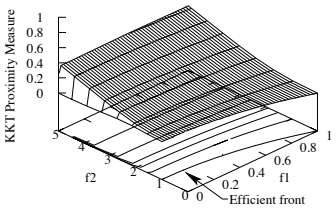

Fig. 16. KKT proximity measure is plotted for Problem ZDT2.

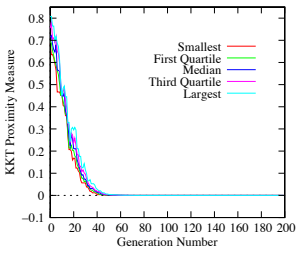

Fig. 17. Generation-wise KKT proximity measure is computed for non-dominated solutions (except $x_1 = 0$ solutions) of NSGA-II populations for Problem ZDT2.

shown in the figure. Notice, how the KKT proximity measure monotonically reduces to zero at the efficient frontier. The contour plots indicate almost parallel contour lines to the

efficient frontier, as the points move away from the efficient frontier. The smoothness of the KKT proximity measure surface is another interesting aspect which can be exploited in various ways.

Now, we demonstrate how the non-dominated sets of points obtained from NSGA-II are evaluated by the KKT proximity measure in Figure 17. A population of size 100 is used and a maximum of 200 generations are run. Here, we include all non-dominated points for KKT Proximity measure computation except points having $x_1 = 0$. This is because $x_1 = 0$, makes $f_1 = 0$ and these points correspond to weak Pareto-optimal points and thus they will be judged as KKT points having our KKT proximity measure value as zero. Figure 16 demonstrates this aspect clearly. It is clear that without considering weak Pareto-optimal points having $x_1 = 0$ points (some of which are also strict Pareto-optimal points), the KKT proximity measure reduces to zero with generation. Finally, at around generation 39, there is one solution that lies on the true efficient front. Soon thereafter, all non-dominated solutions fall on the true efficient front.

If, however, all non-dominated solutions including $x_1 = 0$ are considered for the KKT proximity measure computation, the convergence to the zero KKT proximity measure value is faster, as shown in Figure 18. The first efficient solution is found at generation 1 itself and this point is a weak Pareto-optimal point.

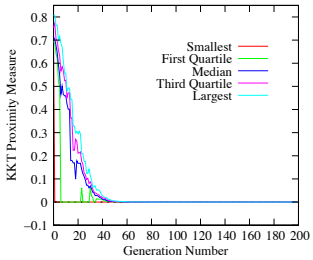

Fig. 18. Generation-wise KKT proximity measure is computed using ASF for all non-dominated solutions of NSGA-II populations for Problem ZDT2.

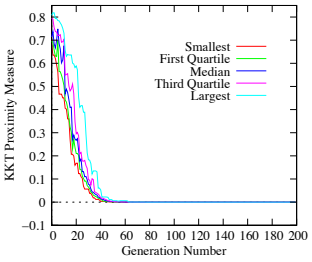

Fig. 19. Generation-wise KKT proximity measure is computed using augmented ASF for *all* non-dominated solutions of NSGA-II populations for Problem ZDT2.

Next, we consider the augmented ASF (or AASF) for the KKT proximity measure computation for all non-dominated solutions of NSGA-II runs on ZDT2 problem. Since weak Pareto-optimal solutions will now not be optimal solutions for the augmented ASF computation, the weak Pareto-optimal solutions will produce non-zero proximity measure value. Figure 19 shows the variation of the KKT proximity measure with AASF. It takes 39 generations to find the first strictly efficient solution. Once again, the measure reduces with generation counter, but the convergence is a bit slower than that in Figure 18. The reduction of KKT proximity measure to zero with generation is an indication that it can be used as a termination condition or can be used as a measure of proximity of non-dominated sets to the efficient frontier. The figure reveals that NSGA-II requires 43 generations to have 50% population members to converge to the efficient front and 50 generations for all 100 solutions to converge. Since AASF formulation is capable of differentiating weak from strict Pareto-optimal points, we use the AASF formulation to compute the KKT proximity measure for the rest of the problems here.

Now, we consider 30-variable ZDT3 problem. KKT proximity measures for NSGA-II non-dominated solutions (obtained with a population size of 40 and run until 200 generations) using the AASF metric are shown in Figure 20. The first Pareto-optimal solution was found at generation 7 and half of all Pareto-optimal solutions were found at generation 26, but it took 159 generations to find all Pareto-optimal solutions. The disjoint efficient front in this problem makes NSGA-II to find 100% population members to converge to the true efficient front. The KKT proximity measure variation supports the original test-problem construction philosophy for designing ZDT3 problem, as it was then thought that the disconnected nature of the efficient sets could provide difficulties to a multi-objective algorithm [24].

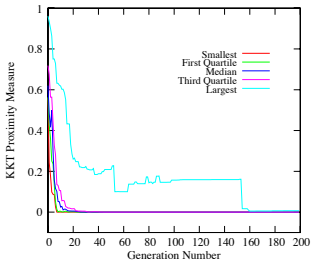

Fig. 20. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for Problem ZDT3.

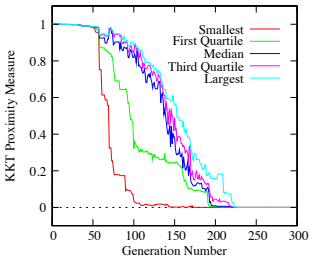

Fig. 21. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for Problem ZDT4.

The 10-variable ZDT4 problem is multimodal and has about 100 local efficient fronts and is more difficult to solve than ZDT1 and ZDT2. Using identical NSGA-II parameter values and with a population size of 100, Figure 21 shows how generation-wise variation of the KKT proximity measure takes place for the non-dominated population members of a typical NSGA-II run. Once again, a monotonic decrease in the KKT proximity measure indicates its suitability as a termination condition for an EMO algorithm. Here again, we do not consider points having $x_1 = 0$ to avoid points for which the derivative does not exist.

Finally, KKT proximity measure values for NSGA-II population members (population size 40) for ZDT6 problem are shown in Figure 22. The first, half and all population members require 21, 27, and 28 generations to converge to the efficient front. This plot suggests almost simultaneous convergence of 40 points on the true efficient set. Although a maximum generations of 200 were fixed for the NSGA-II run, the figure indicates that a maximum of 28 generations were enough to have all solutions to converge on the efficient front for the above simulation run.

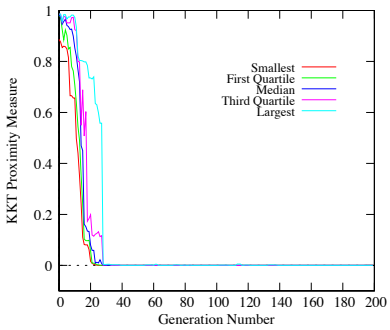

Fig. 22. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for Problem ZDT6.

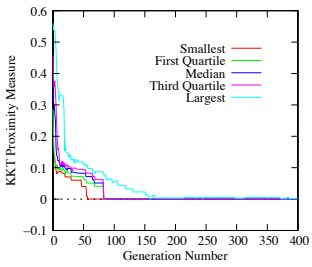

Fig. 24. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for three-objective DTLZ2.

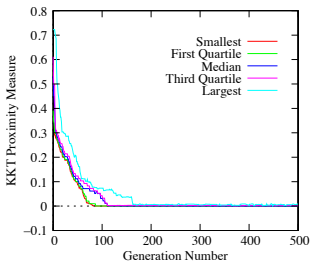

Fig. 25. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for three-objective DTLZ5.

B. Three-objective DTLZ Problems

DTLZ1 problem has a number of locally efficient fronts on which some points can get stuck, hence it is relatively difficult problem to solve to global optimality. Recently proposed NSGA-III procedure [5] is applied to DTLZ1 problems with 92 population members for 1,000 maximum generations. Figure 23 shows the variation of KKT proximity measure values (with AASF formulation) versus the generation counter. Although a Pareto-optimal solution is found fairly early (68

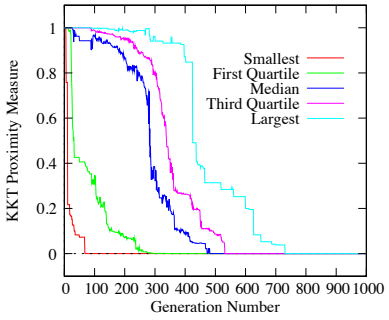

Fig. 23. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for three-objective DTLZ1.

generations), half of the non-dominated set took more than 480 generations to converge to the true efficient front, however all 92 population members took an overall 731 generations for the same purpose. The presence of multiple efficient fronts causes such a delayed overall convergence in this problem.

Next, we apply NSGA-III with a population size of 92 to three-objective DTLZ2 problem, which has a concave efficient front. NSGA-III was run for a maximum of 400 generations. Figure 24 shows the KKT proximity measure values. Interestingly, the convergence of the entire non-dominated set to the respective Pareto-optimal solutions happens almost at the same time in this problem, as dictated by the small differences in smallest, median and largest KKT proximity measure values. The first, 50% and all 92 population members took 55, 83, and 157 generations, respectively, to converge to the efficient front for the above particular run.

DTLZ5 problem has a degenerate efficient front. Although the problem has three objectives, the efficient front is two-dimensional. Figure 25 shows the KKT proximity measure values versus generation number for NSGA-III with a population size of 92. In this problem, a number of 'redundant' solutions remain as non-dominated with Pareto-optimal solutions and these solutions cause a challenge to an EMO algorithm to converge to the true efficient front. The first population member took 80 generations to converge to the efficient front. Half the population took 110 generations to converge to the efficient front, but it tool 163 generations for all 92 population members to converge to the efficient front.

C. Many-Objective Optimization Problems

As the number of objectives increase, EMO algorithms find DTLZ1 and DTLZ2 problems difficult to optimize as evident from Figures 26 and 27. For five and 10-objective versions

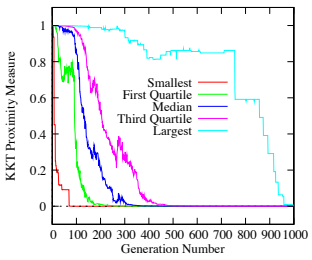

Fig. 26. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for five-objective DTLZ1.

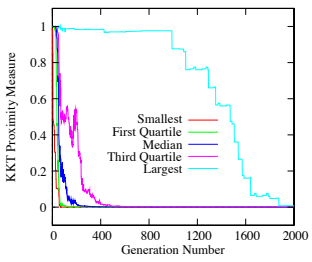

Fig. 27. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for 10-objective DTLZ1.

of DTLZ1 and DTLZ2, we have used 212 ad 276 population members. While the five-objective DTLZ1 problem took 317 and 959 generations to have half and all 212 population members to reach the true efficient front, respectively, the 10-objective problems takes about 215 and 1,875 generations, respectively.

Since DTLZ2 problem is relatively easier to solve to Pareto-optimality, KKT proximity measure values for five and 10-objective DTLZ2 problems are found to be steadily approaching to zero, as evident from Figures 28 and 29, respectively. The five-objective DTLZ2 problem requires 102 and 329 generations for half and all 212 population members to converge

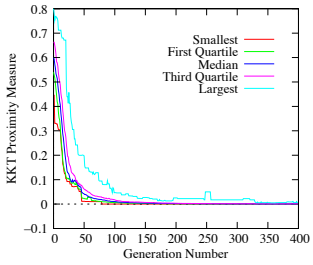

Fig. 28. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for five-objective DTLZ2.

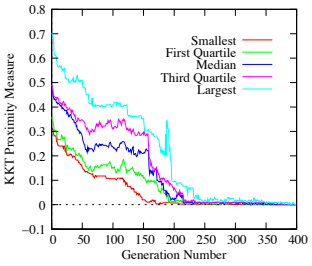

Fig. 29. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-III populations for 10-objective DTLZ2.

to the efficient front, while the 10-objective DTLZ2 problem requires 215 and 348 generations, respectively, for half and all 276 population members to converge to the efficient front.

The above results amply demonstrate the scalability of the proposed KKT proximity metric to many-objective optimization problems and behavior of NSGA-III in solving these problems.

D. Constrained Test Problems

We now consider some standard constrained test problems. First, we consider the problem TNK, which has two constraints, two variables and two objectives, as shown below:

$$\begin{aligned} \text{Min.} \quad & f_1(\mathbf{x}) = x_1, \\ \text{Min.} \quad & f_2(\mathbf{x}) = x_2, \\ \text{s.t.} \quad & g_1(\mathbf{x}) \equiv x_1^2 + x_2^2 - 1 - 0.1\cos\left(16\arctan\tfrac{x_1}{x_2}\right) \geq 0, \\ & g_2(\mathbf{x}) \equiv (x_1 - 0.5)^2 + (x_2 - 0.5)^2 \leq 0.5, \\ & 0 \leq x_1 \leq \pi, \quad 0 \leq x_2 \leq \pi. \end{aligned} \tag{31}$$

NSGA-II is run with 40 population members. For the specific run, the initial population has only six feasible solutions, but since they were away from the Pareto-optimal region, their KKT proximity measure value was calculated to be large. As shown in Figure 30, thereafter, with iterations more and more solutions became feasible and all 40 population members becomes feasible. Importantly, the non-dominated solutions started to approach the efficient front. The corresponding KKT proximity measure values (smallest, median and largest values among the non-dominated solutions) reduce with generation and eventually at generation 48 all 40 population members reach very close to the efficient front. If a termination criterion was used based on KKT proximity measure, generation 48 would have been the earliest opportunity to have all non-dominated solutions close to the efficient front for NSGA-II to terminate. Without such information, it is difficult to decide on an appropriate termination point. Interestingly, NSGA-II or most other EMO methods do not use any gradient information or any KKT optimality conditions in their operators. Even then, the operator interactions of the algorithms allow the whole population to continuously come closer to the efficient front with iterations. This remains as an interesting matter which could encourage theoretical optimization researchers to pay attention to the working procedure of EMO algorithms for multi-objective optimization.

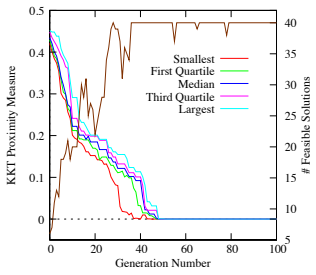

Fig. 30. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for problem TNK.

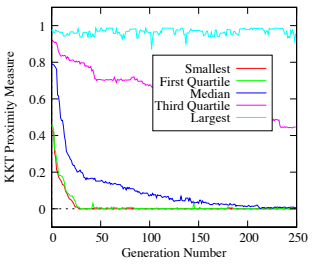

Fig. 31. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for problem BNH.

Next, we consider problem BNH which has two variables and two constraints:

$$\begin{aligned} \text{Minimize} \quad & f_1(\mathbf{x}) = 4x_1^2 + 4x_2^2, \\ \text{Minimize} \quad & f_2(\mathbf{x}) = (x_1 - 5)^2 + (x_2 - 5)^2, \\ \text{subject to} \quad & g_1(\mathbf{x}) \equiv (x_1 - 5)^2 + x_2^2 \leq 25, \\ & g_2(\mathbf{x}) \equiv (x_1 - 8)^2 + (x_2 + 3)^2 \geq 7.7, \\ & 0 \leq x_1 \leq 5, \quad 0 \leq x_2 \leq 3. \end{aligned} \tag{32}$$

Due to the known difficulty in solving this problem, NSGA-II is run with 200 population members. Figure 31 shows the variation of KKT proximity measure with generation for problem BNH. At generation 26, one of the population members falls on the efficient front, while half of the population requires 212 generations. The remaining 50% of population members do not converge to the efficient front in 250 generations.

Next, we consider the consider the SRN problem, stated below:

$$\begin{aligned} \text{Minimize} \quad & f_1(\mathbf{x}) = 2 + (x_1 - 2)^2 + (x_2 - 1)^2, \\ \text{Minimize} \quad & f_2(\mathbf{x}) = 9x_1 - (x_2 - 1)^2, \\ \text{subject to} \quad & g_1(\mathbf{x}) \equiv x_1^2 + x_2^2 \leq 225, \\ & g_2(\mathbf{x}) \equiv x_1 - 3x_2 + 10 \leq 0, \\ & -20 \leq x_1 \leq 20, \quad -20 \leq x_2 \leq 20. \end{aligned} \tag{33}$$

The Pareto-optimal points were found to have the following properties: $x_1 = -2.5$ and $x_2 \in [2.50, 14.79]$. KKT proximity measure values computed for 160,801 uniformly-spaced gridded points in the search space ($x = x_1$ and $y = x_2$) are shown in a contour plot in Figure 32. It is clear that as the solutions get closer to the above Pareto-optimal points, the KKT proximity measure value gets smaller. To illustrate this aspect better, we have plotted the KKT proximity measure for certain fixed $y = x_2$ values in the range $[2.50, 14.79]$ in Figure 33 for different $x = x_1$ values. It is clear from the plot that closer a point is to $x = -2.5$, smaller is the corresponding KKT proximity measure for each y value. For $y = 2.5$, any x larger than -2.5 is infeasible and the KKT proximity measure value is also large. The penalty term in our KKT proximity measure expression (Equation 27) makes a more infeasible solution (having a larger g value) to have a larger KKT proximity measure, as also evident from the figure.

Figure 34 plots the KKT proximity measure for generation-wise NSGA-II non-dominated solutions. A population of size

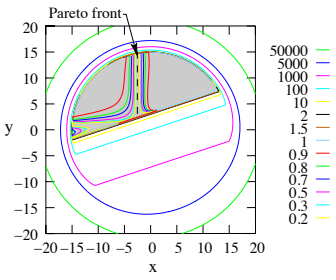

Fig. 32. Contour plots of KKT proximity measure on the constrained problem SRN shows how it monotonically reduces for points close to the true Pareto-optimal solutions ($x = -2.5$ and $y \in [2.5, 14.79]$).

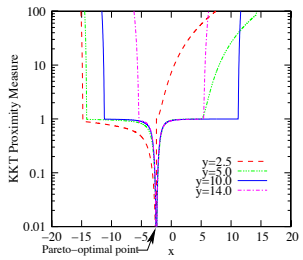

Fig. 33. KKT proximity measure values are shown for certain fixed y values, clearly indicating the monotonic reduction close to Pareto-optimal value ($x = -2.5$).

40 is used. Although one population member with a KKT proximity value less than 0.01 was first discovered at generation 169, it took 363 generations for 25% population members to converge to the efficient front. At 500-th generation, the 50-th and 75-th percentile KKT proximity measure values are 0.018 and 0.026, respectively. The reason for this wide spread

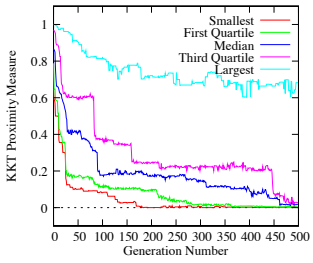

Fig. 34. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for problem SRN.

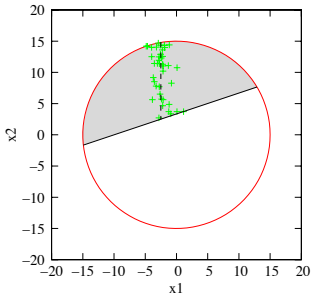

Fig. 35. Non-dominated NSGA-II solutions at generation 250 for problem SRN.

in KKT proximity measure can be understood from Figure 35, which presents the non-dominated solutions in the x_1-x_2 plane at generation 500. It is clear that due to the lack of convergence of all solutions on the exact Pareto-optimal set ($x_1 = -2.5$ and $x_2 \in [2.5, 14.79]$ [2], shown by the dashed line) by NSGA-II procedure, the spread in KKT proximity measure values is also large.

Next, we consider problem OSY, stated below:

$$\begin{aligned}
\text{Min.} \quad & f_1(\mathbf{x}) = -\left[25(x_1-2)^2 + (x_2-2)^2 + (x_3-1)^2 + (x_4-4)^2 + (x_5-1)^2\right], \\
\text{Min.} \quad & f_2(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2, \\
\text{s.t.} \quad & g_1(\mathbf{x}) \equiv x_1 + x_2 - 2 \leq 0, \\
& g_2(\mathbf{x}) \equiv x_1 + x_2 - 6 \leq 0, \\
& g_3(\mathbf{x}) \equiv -x_1 + x_2 - 2 \leq 0, \\
& g_4(\mathbf{x}) \equiv x_1 - 3x_2 - 2 \leq 0, \\
& g_5(\mathbf{x}) \equiv (x_3-3)^2 + x_4 - 4 \leq 0, \\
& g_6(\mathbf{x}) \equiv -(x_5-3)^2 - x_6 + 4 \leq 0, \\
& 0 \leq x_1, x_2, x_6 \leq 10, \quad 1 \leq x_3, x_5 \leq 5, \quad 0 \leq x_4 \leq 6.
\end{aligned} \tag{34}$$

Pareto-optimal solutions for this problem were identified in another study [2]. NSGA-II is run with 200 population size.

Figure 36 shows the KKT proximity measure values (smallest, 25-th percentile, median, 75-th percentile and the largest for all non-dominated solutions) as they vary with generations. A steady decrease in these values with the generation counter can be observed from the figure. The first population member comes close to the efficient front at generation 33. Thereafter, 50-th and 75-th percentile population members took 148 and 173 generations, respectively, to converge to the efficient front. However, until generation 250, all 200 population members could not come close to the efficient front. To identify the

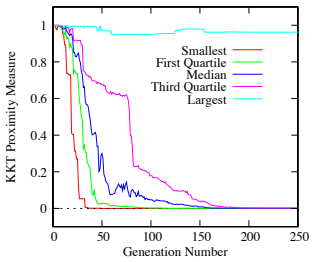

Fig. 36. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for problem OSY.

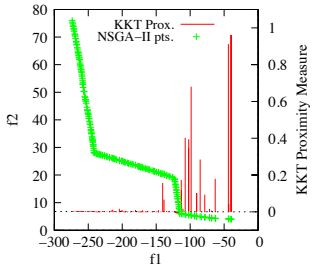

Fig. 37. 200 NSGA-II points on the objective space and their respective KKT proximity measure values are shown for problem OSY at generation 250.

reason for this lack of convergence, Figure 37 shows the NSGA-II points in the objective space and their corresponding KKT proximity measure values at generation 250. It is clear from the figure that a few NSGA-II points have a large KKT proximity measure value. Clearly, minimum f_1 solutions have converged quickly, whereas minimum f_2 solutions have still not converged to their true efficient points even until 250 generations. This information about differential convergence pattern can be used to employ a local search to improve non-converged solutions, a matter we discuss in Section VI.

E. Engineering Design Problems

Next, we consider two multi-objective optimization problems from practice. In both problems, we have used the AASF formulation for the KKT proximity computation.

1) Welded Beam Design Problem: The welded beam design problem has two objectives and a few non-linear constraints [2]. This problem is solved using NSGA-II with 60 population members, initially created at random. Figure 38 shows the variation of the smallest, median and largest KKT proximity measure values with generation counter. The figure shows that although the first Pareto-optimal solution was found at generation 26, half of the population came close to the efficient front after 33 generations. Thereafter, in few more generations most non-dominated points came close to the efficient front. The 75-th percentile of the population converges at generation 49, but the last population member converges at 496-th generation. In practical problems, such behavior is expected. The use of a local search method to improve the final few non-converged population members will constitute a faster computational approach.

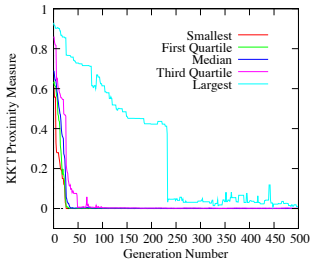

Fig. 38. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for the welded-beam design problem.

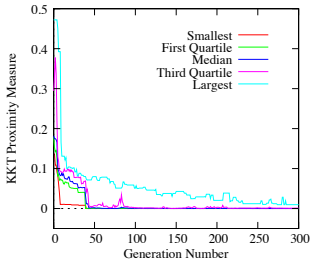

Fig. 39. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for the car side impact problem.

2) Car Side Impact Problem: Next, we consider a three-objective CAR cab design problem having seven variables and 10 constraints and 14 variable bounds. NSGA-III [13] is applied with 92 initial population members. Figure 39 shows the variation of KKT proximity measure with generations. The first Pareto-optimal solution took 23 generations to appear, thereafter, it took another 21 generations for 50% of the population members to converge to the efficient front. The final solution took a total of 274 generations to converge to the efficient front. The behaviors of NSGA-III on this problem and NSGA-II on the welded beam design problem seem to similar.

3) WATER Problem: Finally, we consider a five-objective WATER problem having three variables and seven constraints. NSGA-III [13] is applied with 68 initial population members and using identical other parameter values as in previous problems. Figure 40 shows the variation of KKT proximity measure with generations. An interesting and unknown property of this problem becomes clear. Since there are three variables and five objective functions, the efficient front is at most three-dimensional. Most random solutions lie on the Pareto-optimal set in this problem. It is clear from the figure that the 75-th percentile population member is on the efficient front at the initial population. After 47 generations, the last population member reaches close (within $\epsilon^* \leq 0.01$) to the efficient front. Since this problem is now found to be simple

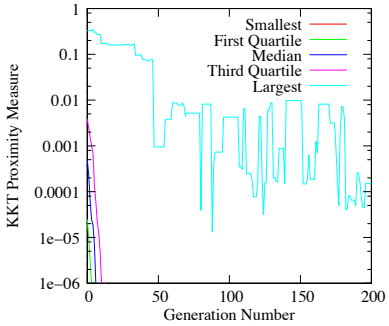

Fig. 40. Generation-wise KKT proximity measure is computed for non-dominated solutions of NSGA-II populations for the WATER problem.

to get to the efficient front easily, we do not recommend this problem to be used as a test problem in many-objective studies.

VI. KKT Proximity Measure for Improved EMO Applications

The KKT proximity measure value for each non-dominated solution in an EMO population can also be used differently to either improve the convergence behavior or to set up a termination condition of an EMO. We discuss these possible future studies in the following.

A. Termination Criterion

Above results amply demonstrate that the proposed KKT proximity measure can be used to terminate a simulation run, as the measure is able to indicate whether multiple trade-off points are truly all close to the efficient front or not. In this section, we demonstrate the possibility of using the KKT proximity measure as a termination criterion by presenting further simulation runs.

For each of the problems discussed above, we perform the following study. To reduce computational burden, we compute the KKT proximity measure (KKTPM) values for all current non-dominated solutions in the population after every five generations. When the median KKTPM value of the current non-dominated set reaches a threshold value or less (say, η_T), we terminate the run. The procedure is repeated from 25 different initial populations, and minimum, median and maximum generations to termination are recorded. They are tabulated in Table I for two different threshold (η_T) values. All two variable

TABLE I
MINIMUM, MEDIAN AND MAXIMUM NUMBER OF GENERATIONS (OUT OF 25 RUNS) NEEDED FOR 50% OF NON-DOMINATED SOLUTIONS TO ACHIEVE A KKT PROXIMITY MEASURE VALUE OF $\eta_T = 0.01$ AND 0.001 FOR TERMINATING AN EMO ALGORITHM. 'II' MEANS NSGA-II AND 'III' MEANS NSGA-III.

Problem	$\eta_T = 0.01$			$\eta_T = 0.001$		
	Best	Median	Worst	Best	Median	Worst
ZDT1	10	15	30	20	45	75
ZDT2	10	20	35	20	30	50
ZDT3	10	15	20	15	20	35
ZDT4	45	60	125	50	80	145
ZDT6	15	20	25	15	25	30
TNK	0	5	5	5	10	15
OSY	65	90	145	95(14)	180(14)	245(14)
BNH (II)	350(1)	350(1)	350(1)	–	–	–
BNH (III)	45	70	165	330(11)	455(11)	490(11)
SRN (II)	310(5)	350(5)	370(5)	–	–	–
SRN (III)	110	235	420	295(3)	325(3)	385(3)
DTLZ1-3	300	370	430	400	440	480
DTLZ1-5	310	360	400	430	470	520
DTLZ1-10	220	250	300	520	550	600
DTLZ2-3	20	25	30	60	80	90
DTLZ2-5	60	70	80	190	230	265
DTLZ2-10	80	100	130	200	290	380
DTLZ5-3	15	15	25	30	40	45
WELD	20	35	155	35(20)	95(20)	265(20)
WATER	0	0	0	0	0	0
CAR	15	25	40	35	60	120

problems are solved using NSGA-II [12] and all three and many-objective problems are solved using NSGA-III [5]. The

numbers in brackets indicate the number of runs (out of 25) the median KKT proximity measure value has reduced below the threshold value of η_T in the event of not all 25 runs satisfying the specified termination condition. The table indicates that the KKT proximity measure can be used as a termination condition, as in most cases the same order of magnitude in the terminating generation number was found to be adequate over multiple runs. For ZDT problems, the termination happens within a few tens of generations. For DTLZ1 problem, NSGA-III satisfies the termination condition within a few hundreds of generations consistently over multiple runs, but for DTLZ2 problem, NSGA-III requires a few tens of generations. BNH and SRN problems were found to be difficult to solve to KKT-optimality using NSGA-II, but NSGA-III seems to perform better on these two problems due to its external guidance property. The WATER problem is found to be extremely easy and 50% non-dominated solutions were found to lie on the efficient front right from the initial generation. In the above experiments, KKT proximity measure was computed after every 5 generations, but a future parametric study is needed to establish a reasonable gap in generations for KKT proximity measure computation so as to make a good balance of extra computations involved in computing it versus the extent of change in KKT proximity measure between two such termination checks.

B. Develop a More Convergent EMO

The KKT proximity measure values for different non-dominated population members provide a relative understanding of convergence pattern at various regions of the efficient front. Refer to Figure 37 in which it was observed that after 250 generations, NSGA-II is able to find solutions very close to the true Pareto-optimal solutions for the minimum f_1 solutions, but many solutions near minimum f_2 region were not close to their respective Pareto-optimal solutions. The solutions having a larger KKT proximity measure value, indicating their relatively slow convergence properties, can be recognized early on during a simulation run. This information can be utilized in an EMO in many different ways to improve EMO's overall convergence behavior. For example, the region having poor KKT proximity values can be subjected to additional focused search, such as using a local search method, so that a faster convergence can be achieved for those regions. Since the variation of KKT proximity measure for different regions of the trade-off frontier can now be available, the rate of convergence of different regions can also be easily calculated and an adequate emphasis of search can be provided at regions having slower rate of convergence. This was not possible before on non-test problems for which no knowledge of Pareto-optimal solutions are known beforehand. KKTPM allows us to get a comprehensive idea of relative convergence behavior of different non-dominated solutions without knowing the true location of the Pareto-optimal solutions.

C. Other Types of Multi-Objective Optimization Problems

The KKT proximity measure can also be used for other types of multi-objective optimization studies, such as for evaluating convergence behavior of preference-based EMO methods – Reference point based EMO [25] or 'light beam' based EMO [26], etc.

Although gradients of objective and constraint functions are needed to compute the KKT proximity measure value at any point, the use of numerical gradients (from function value information at two or more neighboring points) can be explored. This may allow the applicability of the proposed method to a wide variety of problems. For this purpose, the subdifferentials [27] and related KKT optimality theories [19] can be used with our proposed KKTPM construction methodology and applied to certain non-differentiable problems.

D. Develop an Efficient Island-based Parallel EMO

In an island-based parallel application of EMO, a different EMO with separate cone-domination concept was suggested to be to applied on a different processor [28]. A KKT proximity measure computation for each processor would provide the knowledge of fast and slow-converging islands (or processors). To have an overall efficient search method, best solutions from a convergent processor can be migrated to other less-convergent processors in the hope of speeding up the overall convergence of the algorithm to improve the performance of island-based parallel EMO algorithms.

VII. Conclusions

This paper has extended the concept of approximate KKT point definition proposed in an earlier study but for single-objective optimization problems to multi-objective optimization. For each point, an equivalent achievement scalarizing function is developed and a KKT proximity metric is defined to estimate the proximity of the solution from a corresponding Pareto-optimal solution. The idea is novel and has been demonstrated to provide a reducing proximity measure value with the solutions getting closer to the true efficient front on many test problems and a few engineering design problems.

The KKT proximity measure of a set of non-dominated solutions of an EMO population can be utilized to constitute a faster and more convergent method. Some viable methods have been discussed in this paper. One important outcome is that the proposed metric can be used to terminate an EMO run as an indicator of overall convergence behavior (and their proximity) of non-dominated points to the the true Pareto-optimal solutions.

The computation of the metric requires gradients of the objective and constraint functions at each point, hence the metric is computable only for differentiable problems at this stage. Further studies are needed to extend the idea to non-differentiable and discrete problems. To reduce the computational complexity of overall method, it is also suggested to compute the KKT proximity metric value after every five or 10 generations. Despite these limitations, the connection of EMO-obtained solutions being close to theoretical KKT points in a problem remains as a hallmark achievement of this paper. Although gradients are not used in an EMO algorithm for updating one population to its next, the fact that the

overall search process directs the population towards theoretical Pareto-optimal points remains interesting and intriguing. Certainly, such studies close the gap between theoretical and computational optimization studies and bring respects to both methods in utilizing tools of one approach to evaluate the behavior of the other.

ACKNOWLEDGMENTS

Authors acknowledge the efforts of Mr. Haitham Saeda, a PhD student of Michigan State University, in providing us with results of NSGA-III procedure on certain many-objective test problems.

REFERENCES

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

[2] K. Deb, *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: Wiley, 2001.

[3] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 29–38, 2006.

[4] J. Bader, K. Deb, and E. Zitzler, "Faster hypervolume-based search using Monte Carlo sampling," in *Proceedings of Multiple Criteria Decision Making (MCDM 2008)*. Heidelberg: Springer, 2010, pp. 313–326, lNEMS-634.

[5] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point based non-dominated sorting approach, Part I: Solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

[6] G. Haeser and M. L. Schuverdt, "Approximate KKT conditions for variational inequality problems," *Optimization Online*, 2009. [Online]. Available: http://www.optimization-online.org/DB_HTML/2009/10/2415.html

[7] R. Andreani, G. Haeser, and J. M. Martinez, "On sequential optimality conditions for smooth constrained optimization," *Optimization*, vol. 60, no. 5, pp. 627–641, 2011.

[8] R. Andreani, J. M. Martinez, and B. F. Svaiter, "A new sequential optimality condition for constrained optimization and algorithmic consequences," *SIAM. J. OPT*, vol. 20, pp. 3533–3554, 2010.

[9] J. Dutta, K. Deb, R. Tulshyan, and R. Arora, "Approximate KKT points and a proximity measure for termination," *Journal of Global Optimization*, vol. 56, no. 4, pp. 1463–1499, 2013.

[10] R. Tulshyan, R. Arora, K. Deb, and J. Dutta, "Investigating ea solutions for approximate KKT conditions for smooth problems," in *Proceedings of Genetic and Evolutionary Algorithms Conference (GECCO-2010)*. ACM Press, 2010, pp. 689–696.

[11] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation Journal*, vol. 8, no. 2, pp. 125–148, 2000.

[12] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[13] H. Jain and K. Deb, "An evolutionary many-objective optimization algorithm using reference-point based non-dominated sorting approach, Part II: Handling constraints and extending to an adaptive approach," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 602–622, 2014.

[14] V. Chankong and Y. Y. Haimes, *Multiobjective Decision Making Theory and Methodology*. New York: North-Holland, 1983.

[15] A. Messac and C. A. Mattson, "Normal constraint method with guarantee of even representation of complete Pareto frontier," *AIAA Journal*, in press.

[16] P. Shukla and K. Deb, "On finding multiple Pareto-optimal solutions using classical and evolutionary generating methods," *European Journal of Operational Research (EJOR)*, vol. 181, no. 3, pp. 1630–1652, 2007.

[17] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization," in *Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems*, K. C. Giannakoglou, D. T. Tsahalis, J. Périaux, K. D. Papailiou, and T. Fogarty, Eds. Athens, Greece: International Center for Numerical Methods in Engineering (CIMNE), 2001, pp. 95–100.

[18] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *Evolutionary Computation, IEEE Transactions on*, vol. 11, no. 6, pp. 712–731, 2007.

[19] C. R. Bector, S. Chandra, and J. Dutta, *Principles of Optimization Theory*. New Delhi: Narosa, 2005.

[20] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1996.

[21] K. Deb, R. Tiwari, M. Dixit, and J. Dutta, "Finding trade-off solutions close to KKT points using evolutionary multi-objective optimization," in *Proceedings of the Congress on Evolutionary Computation (CEC-2007)*. Piscatway, NJ: IEEE Press, 2007, pp. 2109–2116.

[22] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Systems*, vol. 9, no. 2, pp. 115–148, 1995.

[23] A. P. Wierzbicki, "The use of reference objectives in multiobjective optimization," in *Multiple Criteria Decision Making Theory and Applications*, G. Fandel and T. Gal, Eds. Berlin: Springer-Verlag, 1980, pp. 468–486.

[24] K. Deb, "Multi-objective genetic algorithms: Problem difficulties and construction of test problems," *Evolutionary Computation Journal*, vol. 7, no. 3, pp. 205–230, 1999.

[25] K. Deb, J. Sundar, N. Uday, and S. Chaudhuri, "Reference point based multi-objective optimization using evolutionary algorithms," *International Journal of Computational Intelligence Research (IJCIR)*, vol. 2, no. 6, pp. 273–286, 2006.

[26] K. Deb and A. Kumar, "Light beam search based multi-objective optimization using evolutionary algorithms," in *Proceedings of the Congress on Evolutionary Computation (CEC-07)*, 2007, pp. 2125–2132.

[27] F. H. Clarke, *Optimization and nonsmooth analysis*. Wiley-Interscience, 1983.

[28] K. Deb, P. Zope, and A. Jain, "Distributed computing of Pareto-optimal solutions using multi-objective evolutionary algorithms," in *Proceedings of the Second Evolutionary Multi-Criterion Optimization (EMO-03) Conference (LNCS 2632)*, 2003, pp. 535–549.

APPENDIX

Here we present the computational procedure of the KKT proximity measure with the *augmented* achievement scalarizing function approach to avoid zero KKT proximity metric value for weak Pareto-optimal solutions.

The optimization problem described in Equation 20 is now changed for the augmented ASF approach, as follows:

$$\begin{array}{ll} \text{Min.}_{(\mathbf{x},x_{n+1})} & F(\mathbf{x},x_{n+1}) = x_{n+1} \\ \text{s.t.} & \left(\frac{f_i(\mathbf{x})-z_i}{w_i^k}\right) + \rho\sum_{j=1}^{M}\left(\frac{f_i(\mathbf{x})-z_i}{w_i^k}\right) - x_{n+1} \leq 0, \ \forall i, \\ & g_j(\mathbf{x}) \leq 0, \quad j=1,2,\ldots,J. \end{array} \tag{35}$$

This makes the $G_j(\mathbf{y})$ with $\mathbf{y}=(\mathbf{x};x_{n+1})$ updated as follows:

$$\begin{array}{l} G_j(\mathbf{y}) = \left(\frac{f_j(\mathbf{x})-z_j}{w_j^k}\right) + \rho\sum_{j=1}^{M}\left(\frac{f_i(\mathbf{x})-z_i}{w_i^k}\right) - x_{n+1} \leq 0, \\ \quad j=1,2,\ldots,M. \end{array} \tag{36}$$

Thus, for a given iterate (or solution) $\mathbf{x}^k$ (feasible or infeasible), the optimization problem leading to the computation of the KKT proximity measure with the AASF approach remains as in Equation 23 with the above-mentioned modification in the $G_j(\mathbf{y})$ function. The first constraint now becomes as

follows:

$$\left\| \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} + \sum_{i=1}^{M} u_i \begin{pmatrix} \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_1} + \rho \sum_{j=1}^{M} \frac{\partial f_j(\mathbf{x}^k)}{w_j^k} \\ \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_2} + \rho \sum_{j=1}^{M} \frac{\partial f_j(\mathbf{x}^k)}{w_j^k} \\ \vdots \\ \frac{1}{w_i^k}\frac{\partial f_i(\mathbf{x}^k)}{\partial x_n} + \rho \sum_{j=1}^{M} \frac{\partial f_j(\mathbf{x}^k)}{w_j^k} \\ -1 \end{pmatrix} + \sum_{j=1}^{J} u_{M+j} \begin{pmatrix} \frac{\partial g_i(\mathbf{x}^k)}{\partial x_1} \\ \frac{\partial g_i(\mathbf{x}^k)}{\partial x_2} \\ \vdots \\ \frac{\partial g_i(\mathbf{x}^k)}{\partial x_n} \\ 0 \end{pmatrix} \right\| \leq \sqrt{\epsilon_k}. \tag{37}$$

The above inequality becomes as follows:

$$\left\| \sum_{i=1}^{M} \left(\frac{u_i}{w_i^k} \nabla f_i(\mathbf{x}^k) + \rho u_i \sum_{j=1}^{M} \frac{\nabla f_j(\mathbf{x}^k)}{w_j^k} \right) + \sum_{i=1}^{J} u_{M+i} \nabla g_i(\mathbf{x}^k) \right\|^2 + \left(1 - \sum_{i=1}^{M} u_i \right)^2 - \epsilon_k \leq 0. \tag{38}$$

The second constraint in problem described in Equation 23 becomes as follows:

$$-\epsilon_k - \sum_{i=1}^{M} u_i \left(\frac{f_i(\mathbf{X}^k) - z_i}{w_i^k} + \rho \sum_{j=1}^{M} \frac{f_j(\mathbf{X}^k) - z_j}{w_j^k} - x_{n+1} \right) - \sum_{j=1}^{J} u_{M+j} g_j(\mathbf{x}^k) \leq 0. \tag{39}$$

This completes the formulation of the optimization problem for the AASF approach.

Like before, the optimization problem given in Equation 23 can be used with the above modifications and an optimal solution $(\epsilon^*, x_{n+1}^*, \mathbf{u}^*)$ can be found. Then, the KTT proximity measure for an iterate $\mathbf{x}^k$ can be computed as follows:

$$\text{Augmented KKT Proximity Measure}(\mathbf{x}^k) = \begin{cases} \epsilon_k^*, & \text{if } \mathbf{x}^k \text{ is feasible,} \\ 1 + \sum_{j=1}^{J} \left\langle g_j(\mathbf{x}^k) \right\rangle^2, & \text{otherwise.} \end{cases} \tag{40}$$