Finding Trade-off Solutions Close to KKT Points Using Evolutionary Multi-Objective Optimization

Kalyanmoy Deb, Rahul Tewari, Mayur Dixit, and Joydeep Dutta

***Abstract*— Despite having a wide-spread applicability of evolutionary optimization procedures over the past few decades, EA researchers still face criticism about the theoretical optimality of obtained solutions. In this paper, we address this issue for problems for which gradients of objectives and constraints can be computed either exactly, or numerically or through subdifferentials. We suggest a systematic procedure of analyzing a representative set of Pareto-optimal solutions for their closeness to satisfying Karush-Kuhn-Tucker (KKT) points, which every Pareto-optimal solution must also satisfy. The procedure involves either a least-square solution or an optimum solution to a set of linear system of equations involving Lagrange multipliers. The procedure is applied to a number of differentiable and non-differentiable test problems and to a highly nonlinear engineering design problem. The results clearly show that EAs are capable of finding solutions close to theoretically optimal solutions in various problems. As a by-product, the error metric suggested in this paper can also be used as a termination condition for an EA application. Hopefully, this study will bring EAs and its research closer to classical optimization studies.**

I. Introduction

Evolutionary algorithms (EAs) for solving single or multi-objective optimization problems are often criticized for their lack of theoretical relevance. Often questions are raised whether the obtained solutions are close to being an optimal solution. This is a difficult question to answer for solving any arbitrary problem, not only using an evolutionary optimization technique, but also using any other optimization method. Often in the EA literature, such questions are addressed by first solving a set of test problems for which the optimal solutions are known a priori. Such an exercise provides enough confidence to a reader about the efficacy of the proposed procedure. With the confidence built by solving a test suite of problems, the proposed procedure is then applied to a real problem for which the optimal solutions are not known. Although such a dual-stage optimization is better than a direct application of a proposed procedure to the real-world problem without doing any validation study, such a dual-stage study is still not adequately convincing to a theoretical mind about the worthiness of a heuristic optimization procedure.

In this paper, we suggest a verification procedure based on Karush-Kuhn-Tucker (KKT) conditions [8] to build confidence about the near-optimality of solutions obtained using an evolutionary optimization procedure. The proposed procedure can be used for both single and multi-objective optimization problems, however, here we discuss the procedure for solving multi-objective optimization problems only. Since KKT conditions are used, the procedure is applicable for real-parameter optimization problems with or without availability of gradients. If the functions are differentiable, direct gradients or numerically-computed gradients can be used. However, if one or more functions are continuous but non-differentiable at the critical points in the search space, the concept of subdifferentials [2], [3] can be used. The motivation of this study is to show that evolutionary optimization procedures aided with a local search strategy can lead to KKT solutions in multi-objective optimization problems. Moreover, the suggested error-based metric can also be used as a termination criterion for an evolutionary optimization procedure.

In the remainder of the paper, we briefly discuss the KKT conditions and their relationship with the optima of the problem. Thereafter, we propose the KKT-based validation procedure for solutions obtained using an EMO procedure. The proof-of-principle results are shown by applying the procedure in three non-linear optimization problems. Thereafter, we extend the concept with subdifferentials and show simulation results on a non-differentiable problem. Finally, conclusions of the study are made.

II. Karush-Kuhn-Tucker (KKT) Conditions and Points

KKT conditions are among the most important theoretical optimization results. Let us consider the following problem:

$$\begin{array}{rlll} \text{Minimize} & f_i(\mathbf{x}) & i = 1,\ldots,M, & \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0 & j = 1,\ldots,J, & (1)\\ & h_k(\mathbf{x}) = 0 & k = 1,\ldots,K. & \end{array}$$

The following theorem describes the KKT conditions [7]:

Theorem 1 *Assume that* $\mathbf{x}$ *is a Pareto-optimal (and feasible) solution to the above problem. Let* f_i*,* g_j *and* h_k *functions are real-valued differentiable functions. Assume that gradient vectors of constraints* $\nabla g_i(\mathbf{x})$ *for active inequality constraints and* $\nabla h_k(\mathbf{x})$ *are linearly independent (constraint qualification condition). Then, there exist non-negative scalars* λ_i *for* $i = 1,\ldots,M$ *(with at least one* λ_i *strictly positive), non-negative scalars* u_i *for* $j = 1,\ldots,J$ *and scalars* v_k *for* $k = 1\ldots,K$ *such that*

K. Deb holds Shri Deva Raj Chair Professor, Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India (email: deb@iitk.ac.in)

R. Tewari and M. Dixit are undergraduate students of Mechanical Engineering, IIT Kanpur (email: {rtewari,mdixit}@iitk.ac.in)

J. Dutta is an Assistant Professor of Department of Mathematics and Statistics, IIT Kanpur, India (email: jdutta@iitk.ac.in)

1-4244-1340-0/07$25.00

1) $\sum_{i=1}^{M} \lambda_i \nabla f_i(\mathbf{x}) + \sum_{j=1}^{J} u_j \nabla g_j(\mathbf{x}) + \sum_{k=1}^{K} v_k \nabla h_k(\mathbf{x}) = 0$,
2) $u_j g_j(\mathbf{x}) = 0$ *for* $j = 1, \ldots, J$.

If the feasible search space is convex and all objective functions are also convex, a KKT point becomes a Pareto-optimal solution of the problem [1], [8]. However, for a violation of any of the above convexity assumptions, a KKT point need not be a Pareto-optimal solution, but it is certain that a Pareto-optimal solution must always be a KKT point. Due to this property, many classical optimization algorithms attempt to find KKT points for a nonlinear optimization problem, such as augmented Lagrangian method, generalized reduced gradient method etc. [8]. Here, we attempt to investigate if solutions obtained by a hybrid EMO-local-search procedure are KKT points.

III. Proposed Procedure

The procedure suggested here is simple. After a set of trade-off (non-dominated) solutions (P) are found using an EMO-cum-local-search procedure, the following methodology is suggested. Either all solutions in P or a subset of well-distributed solutions can be obtained from P by using a clustering procedure. We call this set P'. Now, from each member $\mathbf{x}^{(i)}$ of P' having function values $\mathbf{f}^{(i)} = (f_1^{(i)}, f_2^{(i)}, \ldots, f_M^{(i)})$, we perform a local search procedure to try to improve the solution. Various single-objective procedure can be used for this purpose. In this paper, we suggest the ϵ-constraint procedure in which all but one objectives are converted to constraints by restricting their values to lie smaller than their current function values:

$$\begin{array}{ll} \text{Minimize} & f_M(\mathbf{x}), \\ \text{Subject to} & f_j(\mathbf{x}) \le f_j^{(i)}, \quad j = 1, 2, \ldots, (M-1), \\ & \mathbf{x} \in \mathcal{S}. \end{array} \quad (2)$$

To solve this single-objective optimization problem, an evolutionary optimization procedure or any other numerical optimization techniques can be used. In this paper, we use the SQP procedure of MATLAB for finding the locally optimal solution (say $\mathbf{z}^{(i)}$) of the above problem for each solution i.

At this locally optimal solution, we compute gradients of the objective functions and all *active* constraints (constraints on which the solution $\mathbf{z}^{(i)}$ lies). For this purpose, all variable bounds are also converted into inequality constraints and checked for their activeness. We check to see if the solution satisfies the constraint qualification condition.

Next, we construct the following KKT condition:

$$\sum_{i=1}^{M} \lambda_i \nabla f_i(\mathbf{z}^{(i)}) + \sum_{j=1}^{J} u_j \nabla g_j(\mathbf{z}^{(i)}) + \sum_{k=1}^{K} v_k \nabla h_k(\mathbf{z}^{(i)}) = 0. \quad (3)$$

In the above vector equation, parameters λ_i, u_j and v_k are unknown. The KKT optimality conditions can be restated for a given solution $\mathbf{z}^{(i)}$ as follows. If there exist vectors $\lambda \ge \mathbf{0}$ (but $\lambda \ne \mathbf{0}$), $\mathbf{u} \ge 0$ and any $\mathbf{v}$ such the above vector equation is satisfied, the point $\mathbf{z}^{(i)}$ is a KKT point. Since u_j is expected to be zero for inactive constraints, the gradient term for inactive constraints need not be included in the above equation. From now on, we use J to indicate the number of active inequality constraints at a point of interest.

In equation 3, all gradient vectors are known and only unknown parameters are λ, $\mathbf{u}$ and $\mathbf{v}$. Since all λ_i values cannot be zero and the right side value is a zero, we normalize the λ_i parameters by using $\sum_{i=1}^{M} \lambda_i = 1$, such that

$$\lambda_M = 1 - \sum_{i=1}^{M-1} \lambda_i. \quad (4)$$

Eliminating λ_M from equation 3, we obtain the following vector equation (for simplicity in writing, we drop $\mathbf{z}^{(i)}$ from the expressions):

$$\begin{array}{l}(\nabla f_1 - \nabla f_M)\lambda_1 + (\nabla f_2 - \nabla f_M)\lambda_2 + \cdots + \\ (\nabla f_{M-1} - \nabla f_M)\lambda_{M-1} + \sum_{j=1}^{J} \nabla g_j u_j + \sum_{k=1}^{K} \nabla h_k v_k = -\nabla f_M. \end{array} \quad (5)$$

The above equation is a well-known matrix equation: $Ay = b$, in which

$$\begin{array}{rcl} A & = & ((\nabla f_1 - \nabla f_M), \ldots, (\nabla f_{M-1} - \nabla f_M), \\ & & \nabla g_1, \ldots, \nabla g_J, \nabla h_1, \ldots, \nabla h_k), \\ b & = & -\nabla f_M, \\ y & = & (\lambda_1, \ldots, \lambda_{M-1}, u_1, \ldots, u_J, v_1, \ldots, v_K)^T. \end{array}$$

Thus, the matrix A is a $n \times p$ matrix, where n is the number of decision variables and p is the number of unknown parameters ($p = (M-1) + J + K$).

Based on the size of A matrix, we shall have three different possible scenarios of solving the $Ay = b$ system. First, we consider the case for which $n > p$, which signifies that there are more equations than unknown and in general there may not exist a solution to the above system. In such a scenario, we find a least-square solution to the above system and obtain:

$$\bar{y} = (A^T A)^{-1}(A^T b). \quad (6)$$

Thereafter, we define the error measure as $\tilde{e} = \|b - A\bar{y}\|/\|b\|$. If this error is close to zero, it can be assumed that the system of equation $Ay = b$ has a solution and the corresponding $\mathbf{z}^{(i)}$ is a KKT solution.

On the other hand, if $n < p$ meaning that the number of equations is less than the number of unknowns, we pose it as an optimization problem as follows:

$$\begin{array}{ll} \text{Minimize} & e(\mathbf{y}) = \|b - A\bar{\mathbf{y}}\|/\|b\|, \\ \text{Subject to} & \mathbf{y}_g \ge \mathbf{0}, \end{array} \quad (7)$$

where $\mathbf{y}_g$ are $\mathbf{y}$ parameters for inequality constraints. If the optimal solution for the above problem has an error value close to zero, it can said that the corresponding solution $\mathbf{z}^{(i)}$ is a KKT solution.

In the third scenario (when $n = p$), the matrix A is a square matrix. If the matrix is not ill-conditioned, we can simply compute $\bar{y} = A^{-1}b$ and calculate the error measure $e(\bar{y}) = \|b - A\bar{y}\|/\|b\|$. If this error value is close to zero, the corresponding solution $\mathbf{z}^{(i)}$ is a KKT solution.

Thus, depending on the size and condition of matrix A, we use either a least-square solution, an optimization procedure or a simple matrix inverse to check if the solution $\mathbf{z}^{(i)}$ is close to being a KKT solution.

IV. Simulation Results

In all simulations here, we use NSGA-II with SBX recombination operator with $p_c = 0.9$ and $\eta_c = 10$ and polynomial mutation with $p_m = 1/n$ and $\eta_m = 20$ [4], [5]. The SQP procedure of MATLAB software is used as the local search operator. In each case, we run the procedure 20 times from different initial populations and present a representative set of results here.

A. Problem 1

First, we use a unconstrained problem with convex objective functions:

$$\begin{array}{ll} \text{Minimize} & f_1(x,y) = x^2 + y^2, \\ \text{Minimize} & f_2(x,y) = (x-2)^2 + y^2. \end{array} \tag{8}$$

Thus, a KKT point is guaranteed to be an optimal solution. We use a population of size 100 and maximum generation of 100. In this problem, we do not use the local search procedure. Figure 1 shows the reduction in error (best and median among non-dominated solutions) metric with generation counter. Figure 2 shows the non-dominated solutions after 10 generations. Although points close to the true Pareto-optimal front (shown in a solid line) are found, many of them possess an error value larger than 0.01. Also the number of generations are not enough for NSGA-II to find a good distribution of points. Figure 3 shows the non-dominated solutions after 50 generations. By this time, almost all obtained trade-off solutions are Pareto-optimal, as also confirmed by Figure 1. Since an error of 10^{-4} or smaller is achieved by 50 generations, these solutions are very close to being KKT solutions and by virtue of being convex objective functions, these points are also very close to the true Pareto-optimal solutions.

B. Problem 2

Next, we use a two-variable constrained problem [4]:

$$\begin{array}{ll} \text{Minimize} & f_1(\mathbf{x}) = x_1, \\ \text{Minimize} & f_2(\mathbf{x}) = \frac{1+x_2}{x_1}, \\ \text{subject to} & g_1(\mathbf{x}) \equiv x_2 + 9x_1 \geq 6, \\ & g_2(\mathbf{x}) \equiv -x_2 + 9x_1 \geq 1, \\ & 0.1 \leq x_1 \leq 1, \quad 0 \leq x_2 \leq 5. \end{array} \tag{9}$$

A part of the Pareto-optimal front lies on the first constraint. Figure 4 shows how the best and median error values reduce with generation counter. 100 population members are used. Figure 5 shows the non-dominated solutions after 5 generations. It is clear that most solutions possess a large error at such an early stage of the optimization run. However, at generation 50 (Figure 6), many solutions have reached near the constraint-free part of the Pareto-optimal front showing an error value smaller than 0.01. Most solutions near the constrained Pareto-optimal region have a large error value, indicating that this region are difficult to optimize.

Next, we show the effect of a local search after NSGA-II solutions are found. We terminate NSGA-II run at 50 generations and then perform a local search from each obtained non-dominated solution ($\mathbf{f}^*$) by forming a ϵ-constraint problem by restricting objective f_1 to be less than $\epsilon = f_1^*$. The solutions obtained by the local search method are also shown in Figure 7. The figure shows that most NSGA-II solutions which were on the constrained part of the Pareto-optimal front are now brought on the constraint by the local search method. It is also interesting to note that the solutions in the constraint-free part of the Pareto-optimal front did not improve by the local search, since NSGA-II could already find the solutions with zero error value.

C. Welded Beam Design

Third, we use a welded beam design problem having four variables ($\mathbf{x} = (h, \ell, t, b)^T$) and four nonlinear constraints:

$$\begin{array}{ll} \text{Minimize} & f_1(\mathbf{x}) = 1.10471 h^2 \ell + 0.048\,11 tb(14.0 + \ell), \\ \text{Minimize} & f_2(\mathbf{x}) = \delta(\mathbf{x}) = \frac{2.1952}{t^3 b}, \\ \text{Subject to} & g_1(\mathbf{x}) \equiv 13600 - \tau(\mathbf{x}) \geq 0, \\ & g_2(\mathbf{x}) \equiv 30000 - \sigma(\mathbf{x}) \geq 0, \\ & g_3(\mathbf{x}) \equiv b - h \geq 0, \\ & g_4(\mathbf{x}) \equiv P_c(\mathbf{x}) - 6000 \geq 0, \\ & 0.125 \leq h, b \leq 5, \quad 0.1 \leq \ell, t \leq 10. \end{array} \tag{10}$$

The terms $\tau(\mathbf{x})$, $\sigma(\mathbf{x})$, and $P_c(\mathbf{x})$ are given below:

$$\begin{aligned} \tau(\mathbf{x}) &= \\ &\sqrt{(\tau'(\mathbf{x}))^2 + (\tau''(\mathbf{x}))^2 + \ell\tau'(\mathbf{x})\tau''(\mathbf{x})/\sqrt{0.25[\ell^2 + (h+t)^2]}}, \\ \sigma(\mathbf{x}) &= \frac{504000}{t^2 b}, \\ P_c(\mathbf{x}) &= 64746.022(1 - 0.028\,234\,6t)tb^3, \end{aligned}$$

where $\tau'(\mathbf{x}) = \frac{6000}{\sqrt{2}h\ell}$ and

$$\tau''(\mathbf{x}) = \frac{6000(14 + 0.5\ell)\sqrt{0.25[\ell^2 + (h+t)^2]}}{2\left\{0.707h\ell[\ell^2/12 + 0.25(h+t)^2]\right\}}.$$

The objective functions and constraints are nonlinear but differentiable. We compute the gradients exactly for the optimality verification. In the KKT analysis, only active constraints and variable bounds are considered at a given point.

Figure 8 shows the decrease in median and best error of the NSGA-II (population size 200) solutions in the non-dominated front at each population of every generation. Figure 9 shows the non-dominated solutions after 25 generations of NSGA-II. In this problem, the Pareto-optimal region near the minimum-cost solution is difficult to find, simply due to more number of active constraints. However, the region near the minimum-deflection solution only makes one constraint active [6] and is comparatively easy to find. The final non-dominated front obtained by NSGA-II-cum-local-search method is also shown in a solid line. The figure shows that solutions near the minimum-cost solutions are not quite on the final front after 25 generations. Figure 10 shows the non-dominated solutions after 200 generations. Although solutions are close to the final front, many solutions have an error value which is more than 2%. However, when a

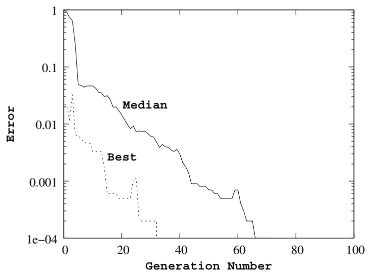

Fig. 1. Error versus generation number for problem 1.

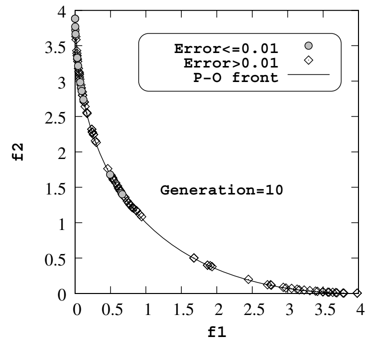

Fig. 2. Non-dominated solutions at generation 10 for problem 1.

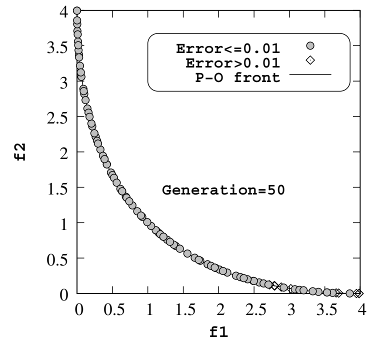

Fig. 3. Non-dominated solutions at generation 50 for problem 1.

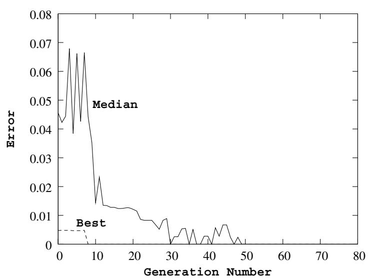

Fig. 4. Error versus generation number for problem 2.

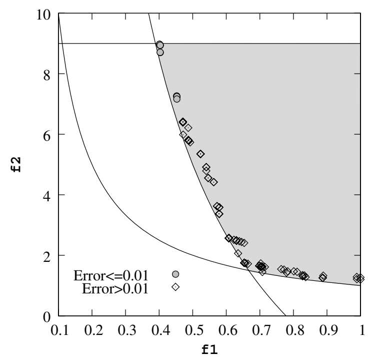

Fig. 5. Non-dominated solutions at generation 5 for problem 2.

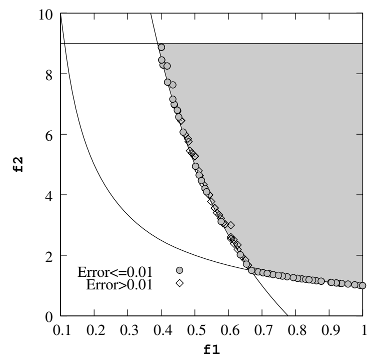

Fig. 6. Non-dominated solutions at generation 50 for problem 2.

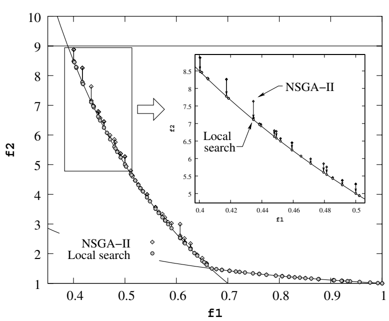

Fig. 7. Solutions after local search for problem 2.

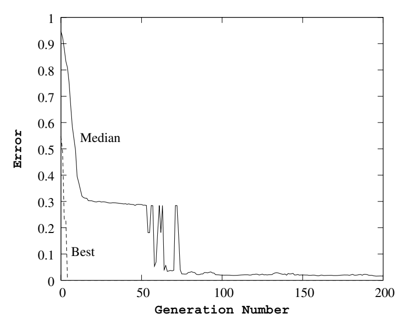

Fig. 8. Error values versus generation counter for the welded beam design problem.

local search method (SQP) is applied from each solution of the population at 200 generation, they improve slightly and approach a local optimal solution. Figure 11 shows the non-dominated solutions after the local search procedure. The error after the local search is smaller than $1(10^{-7})$, except at the minimum cost solution where the error is 0.000378. Figure 12 shows the change in error values before (a maximum of 0.434151) and after ($1(10^{-7})$) the local search. This welded beam design problem has been attempted to solve by many researchers in past, but this is first time, we present a trade-off frontier with a theoretical confidence of their closeness to optimum. We present a set of obtained trade-off solutions in Table I. It is interesting to note that how λ_1 parameter reduces from minimum-cost solution to the

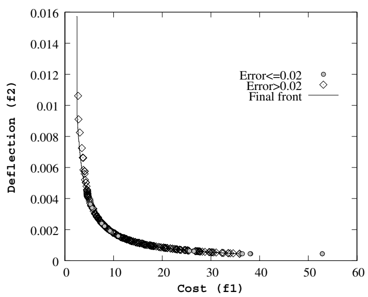

Fig. 9. Non-dominated solutions at generation 25 for welded beam design problem.

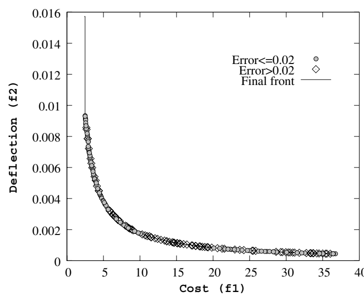

Fig. 10. Non-dominated solutions at generation 200 for welded beam design problem.

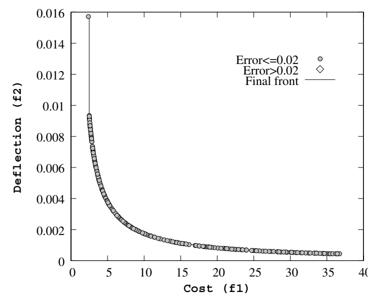

Fig. 11. Non-dominated solutions after a local search of population at 200 generation.

TABLE I

10 WELL-DISTRIBUTED PARETO-OPTIMAL SOLUTIONS. THE FIRST AND 10-TH SOLUTIONS ARE MINIMUM-COST AND MINIMUM-DEFLECTION SOLUTIONS, RESPECTIVELY.

h	ℓ	t	b	f_1	f_2	Error	λ_1	u_1	u_3	u_4	u_t	u_b
0.244	6.214	8.298	0.244	2.381	0.016	3.780e-04	1.054e-01	0.0627	0.0437	0.000	0.145	
0.235	5.178	10.000	0.235	2.486	0.009	0.000e+00	5.891e-03	0.0071	0.0035	1.0e-06	0.022	
0.310	3.606	10.000	0.310	3.003	0.007	0.000e+00	3.193e-03	0.0036	0.0013		0.016	
0.438	2.350	10.000	0.438	3.942	0.005	0.000e+00	1.541e-03	0.0018	0.0003		0.011	
0.653	1.470	10.000	0.714	6.009	0.003	0.000e+00	5.780e-04	0.0008			0.007	
0.857	1.075	10.000	1.263	10.033	0.002	0.000e+00	1.900e-04	0.0003			0.004	
1.139	0.777	10.000	2.235	17.006	0.001	0.000e+00	6.200e-05	0.0001			0.002	
1.341	0.644	10.000	3.070	22.911	0.001	0.000e+00	3.300e-05	7.7e-05			0.001	
1.558	0.542	10.000	4.088	30.055	0.001	0.000e+00	1.900e-05	4.9e-05			0.001	
1.223	0.751	10.000	5.000	36.723	0.000	0.000e+00	0.000e+00				0.001	0.0004

minimum-deflection solution. Since this parameter signifies the importance of the first objective (cost objective, here), this is expected. The table also presents the Lagrange multipliers for constraints and variable bounds which are active at the respective point. It is clear that for the first nine solutions constraint g_1 is active and the minimum-deflection solution makes only two variable bounds (upper bounds on t and b) active.

Figures 13 till 16 show the variation of decision variables h, ℓ, t and b with the first objective for both before and after the local search. It is evident from the figures that the local search is able to bring out salient properties of optimal trade-off solutions for the welded beam design problem:

1) The parameter h (weld thickness) must be increased for better-deflection solutions.
2) The parameter ℓ (overhung length) must be reduced for better-deflection solutions. From the minimum-cost solution, the length ℓ drastically reduces with an improved deflection solution.
3) Besides the minimum-cost solution, all trade-off solutions require the largest possible value of parameter t (beam width) or $t = 10$ (upper bound).
4) Interestingly, the parameter b (beam height) must be increased linearly from almost the lower bound of b to its upper bound for a better deflection solution.
5) Figure 17 depicts that for all Pareto-optimal solutions constraint g_1 is active. For feasible solutions, constraint g_1 and g_2 must take values less than one and g_3 and g_4 must take values more than one.
6) The minimum-cost solution makes three constraints active (g_1, g_3 and g_4). Thereafter, till about cost of 5 units, two constraints ($g1$ and g_3) are active.
7) Constraint g_2 is not active on any Pareto-optimal solutions and is the least important constraint.
8) We have already observed that all Pareto-optimal solutions must have the highest allowed value of t or $t = 10$ in. Only for the minimum-deflection solution, the parameter b takes its maximally allowed value of $b = 5$ in.

NSGA-II solutions without the local search do not adequately bring out such properties.

D. A Possible Termination Criterion

It is interesting to note that the error metric defined above reduces with generation and as the error metric value becomes close to zero, the corresponding solution is a KKT solution. Thus, the suggested error metric can be used to define a termination criterion for an optimization run including an evolutionary algorithm. Another recent study [9] suggested a KKT-based stopping criterion for unconstrained problems. If for all solutions (or a set of clustered solutions from the optimized front) the error metric values are smaller than a threshold (say, 0.01), the simulation can be terminated. However, it is clear from the description that the computation of the error metric may be computationally expensive,

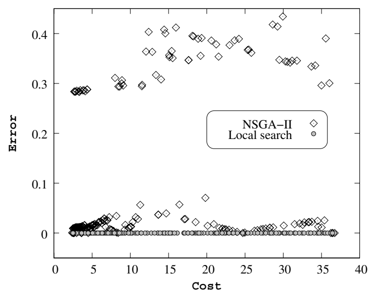

Fig. 12. Error values drastically reduce after local search.

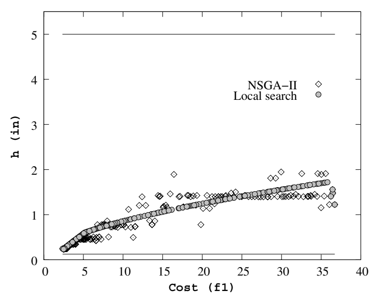

Fig. 13. Parameter h with cost.

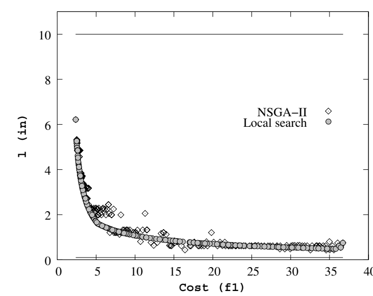

Fig. 14. Parameter ℓ with cost.

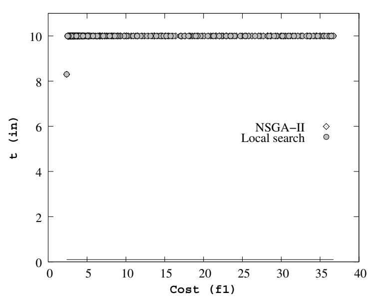

Fig. 15. Parameter t with cost.

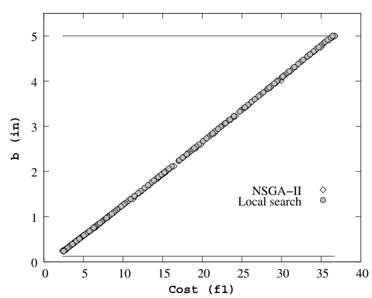

Fig. 16. Parameter b with cost.

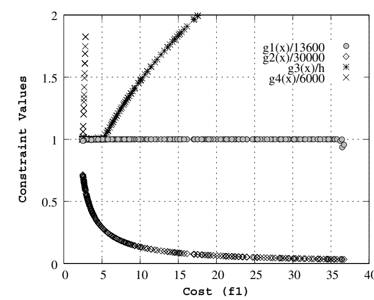

Fig. 17. Normalized constraints with cost.

thereby the use of it in every generation for checking the termination condition is questionable. But, the procedure can be applied after every few generations and the termination, if any, can be determined.

V. Non-differentiable Problems

In the event of non-differentiable objective functions and constraints, the concept of Clarke subdifferentials [2], [3] can be utilized. To define the Clarke subdifferential, we first define Clarke directional derivative at $\mathbf{x} \in R^n$ and in the direction $\mathbf{v} \in R^n$ for a locally Lipschitz function f as follows:

$$f^o(\mathbf{x}, \mathbf{v}) = \lim_{\mathbf{y}\to\mathbf{x}} \sup_{t\to 0+} \frac{f(\mathbf{y}+t\mathbf{v}) - f(\mathbf{y})}{t}, \qquad (11)$$

where $\|\mathbf{v}\| = 1$, $\mathbf{y} \in R^n$ and $t > 0$. For locally Lipschitz function, the right-hand side is bounded and the limit is finite. The Clarke subdifferential is then defined as follows:

$$\partial^c f(\mathbf{x}) = \{\zeta \in R^n : f^o(\mathbf{x}, \mathbf{v}) \geq \langle \zeta, \mathbf{v} \rangle, \forall \mathbf{v} \in R^n\}. \qquad (12)$$

In other words, the Clarke subdifferential is a set of all vectors whose component along any direction $\mathbf{v}$ is smaller than or equal to the Clarke directional derivative defined above. For locally Lipschitz functions, an important result is that the Clarke subdifferential is a convex and compact set made up with limiting derivatives $\lim_{i\to\infty} \nabla f(\mathbf{x}^{(i)})$ at neighboring points $\mathbf{x}^{(i)} \to \mathbf{x}$. For example, the function $f(x) = |x|$ is not differentiable at $x = 0$. However, the limiting derivatives for neighboring solutions $x^{(i)} > 0$ is 1 and for $x^{(i)} < 0$ is -1. Thus, any real-value in $[-1, 1]$ is a Clarke subdifferential of $f(x)$ at $x = 0$. It is interesting to note that Clarke subdifferential contains the $\nabla f(\mathbf{x})$ at a point $\mathbf{x}$ if the function is continuously differentiable.

For the problem given below having non-differentiable, locally Lipschitz functions f_i, $i = 1, \ldots, M$ and g_j, $j = 1, \ldots, J$:

$$\begin{array}{ll} \text{Minimize} & f_i(\mathbf{x}), \\ \text{Subject to} & g_j(\mathbf{x}) \leq 0, \end{array} \qquad (13)$$

the following theorem provides the definition of a necessary condition for an optimal solution $\mathbf{x}$ to the above problem [2], [3].

Theorem 2 *Assume that at least one constraint is active at the given point* $\mathbf{x}$*and there exists a direction* $\mathbf{d}$*, such that* $g_i^o(\mathbf{x}, \mathbf{d}) < 0$ *for all active constraints. Then, there exists scalars* $\lambda_i \geq 0$ *for* $i = 1, \ldots, M$ *(but not* λ_i *equal to zero) and* $u_j \geq 0$ *for* $j = 1, \ldots, J$ *such that*

1) $\mathbf{0} \in \sum_{i=1}^{M} \lambda_i \partial^c f_i(\mathbf{x}) + \sum_{j=1}^{J} u_j \partial^c g_j(\mathbf{x})$
2) $u_j g_j(\mathbf{x}) = 0$, *for* $i = 1, \ldots, J$.

The second conditions ensures that if the constraint is inactive at $\mathbf{x}$, $u_j = 0$. Otherwise, $u_j \geq 0$. The right-hand side of the first condition represents a convex hull and the condition states that if the zero-vector is included in the convex hull, the solution $\mathbf{x}$ is a candidate for the minimum of the original problem.

In our approach, the second condition is easily handled by only considering active constraints and by ensuring u_j and λ_i values are never negative. To handle the first condition, we can introduce additional decision vectors $\mathbf{s}$ and $\mathbf{t}$ for non-differentiable objective functions and constraints, respectively. Knowing the convex hull in which any subdifferential may lie, we then form the following conditions:

$$\mathbf{0} \in \sum_{i=1}^{M} \lambda_i s_i + \sum_{j=1}^{J} u_j t_j, \tag{14}$$

where for i and j, $s_i \in \partial^c f_i(\mathbf{x})$, and $t_j \in \partial^c g_j(\mathbf{x})$. The right-hand side of the above condition can be written as $e(\mathbf{s},\mathbf{t}) = A(\mathbf{s},\mathbf{t})y - b(\mathbf{s})$. To find suitable vectors $\mathbf{s}$ and $\mathbf{t}$, we form an optimization problem to find if there exist any $(\mathbf{s},\mathbf{t})$ which will make the above error term $|e(\mathbf{s},\mathbf{t})|$ close to zero. The vector $y = (\lambda, \mathbf{u})^T$. Thus, we solve the following optimization problem:

$$\begin{array}{ll} \text{Minimize} & |e(\mathbf{s},\mathbf{t})|, \\ \text{Subject to} & s_i \in \partial^c f_i(\mathbf{x}), \\ & t_j \in \partial^c g_j(\mathbf{x}). \end{array} \tag{15}$$

If the optimized absolute error value is close to zero, the solution $\mathbf{x}$ can be said to be close to being a candidate solution for minimum of the original optimization problem.

A. Simulation Results

We illustrate the above procedure by choosing the CTP3 test problem [4] for which all Pareto-optimal solutions lie on the constraint boundary:

$$\begin{array}{rl} \text{Min.} & f_1(\mathbf{x}) = x_1, \\ \text{Min.} & f_2(\mathbf{x}) = g(\mathbf{x}) \left(1 - \frac{f_1(\mathbf{x})}{g(\mathbf{x})}\right), \\ \text{s.t.} & C(\mathbf{x}) \equiv \cos(\theta)[f_2(\mathbf{x}) - e] - \sin(\theta) f_1(\mathbf{x}) \geq \\ & a \left|\sin\{b\pi [\sin(\theta)(f_2(\mathbf{x}) - e) + \cos(\theta) f_1(\mathbf{x})]^c\}\right|^d, \\ \text{where} & g(\mathbf{x}) = 1 + \sum_{i=2}^{n} (x_i - 0.5)^2, \\ & 0 \leq x_i \leq 1, \quad i = 1, 2, \ldots, n. \end{array} \tag{16}$$

We choose the following parameter values for CTP3:

$$\theta = -0.2\pi,\ a = 0.1,\ b = 10,\ c = 1,\ d = 0.5,\ e = 1.$$

The problem is such that there exist a finite number of Pareto-optimal solutions and at all such solutions the constraint function (absolute value of a sine function) is non-differentiable. Using the chain-rule of Clarke subdifferential [2] for a composite function $(g \circ F)(\mathbf{x})$ (for which a non-differentiable function $g : R \to R$ at $\mathbf{x}$ and a differentiable function $F : R^n \to R$) as $\partial^c (g \circ F)(\mathbf{x}) \subset \partial^c g(F(\mathbf{x})) \nabla F(\mathbf{x})$. For the constraint of the above problem, a part of the function is differentiable in the entire search space and another part which is not differentiable but is a composite function as described above. Using the above chain rule, we obtain the following expression for the derivative of the constraint function:

$$\partial^c g(\mathbf{x}) = \nabla g^1(\mathbf{x}) + s \nabla g^2(\mathbf{x}), \tag{17}$$

where s is the subdifferential $\partial^c |y| = [-1, 1]$ at $y = 0$ and $\nabla g^1(\mathbf{x})$ and $\nabla g^2(\mathbf{x})$ are exact derivatives of the relevant parts of the constraint function. Using this Clarke subdifferential, we have the KKT condition as follows:

$$\begin{array}{rcl} \mathbf{0} & \in & \lambda_1 \nabla f_1(\mathbf{x}) + (1 - \lambda_1) \nabla f_2(\mathbf{x}) + u(\nabla g^1(\mathbf{x}) \\ & & + s \nabla g^2(\mathbf{x})), \\ u g(\mathbf{x}) & = & 0 \ \text{ and } 0 \leq \lambda_1 \leq 1. \end{array}$$

Based on the procedure described in section V, we formulate the following optimization problem:

$$\begin{array}{ll} \text{Minimize} & e(s) = \|A(s)\bar{y} - b\| / \|b\|, \\ \text{Subject to} & s \in [-1, 1]. \end{array} \tag{18}$$

Here, the matrix

$$A(s) = [(\nabla f_1(\mathbf{x}) - \nabla f_2(\mathbf{x})), (\nabla g^1(\mathbf{x}) + s \nabla g^2(\mathbf{x}))]$$

and $b = -\nabla f_2(\mathbf{x})$. The multiplier vector $\bar{y} = (\bar{\lambda}_1, \bar{u})^T$ is found by solving the system $A(s)y = b$, as described in section III. If the optimized solution s^* makes the error function value $e(s^*)$ close to zero, we can argue that we have found a subdifferential value of the absolute function at zero such that the right-hand side is almost close to the zero vector, thereby satisfying the condition given in 18.

Figure 18 shows the reduction in best and median population (non-dominated front) error values with generation using NSGA-II with 200 population members. Although solutions with an error value of zero exists right from the first generation, the figure clearly shows how the median error values reduce and eventually becomes zero at around 10 generations for this problem.

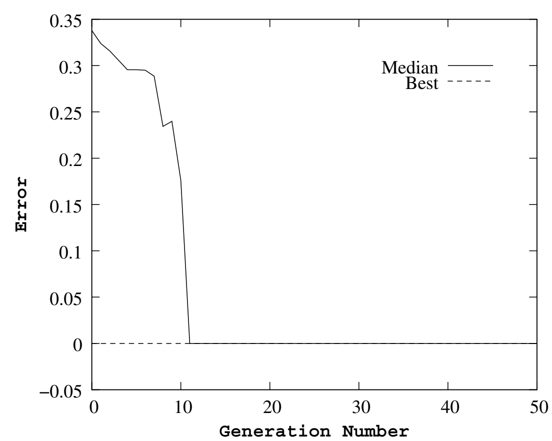

Fig. 18. Error versus generation number for CTP3.

Figure 19 shows the non-dominated solutions are various generations. It is clear how the solutions get closer to the true Pareto-optimal solutions with generation and eventually reach the Pareto-optimal solutions (with an error value of zero). Each Pareto-optimal solution is non-differentiable due to the presence of a 'kink'. But the suggested procedure with Clarke's subdifferential is able to find a valid subdifferential value within $[-1, 1]$ for every optimal solution to satisfy the corresponding KKT conditions. Table II shows the 13

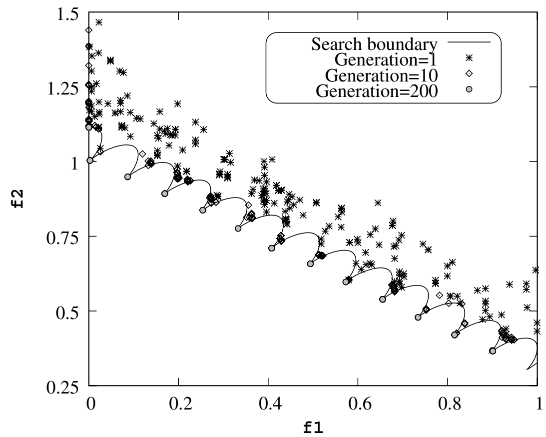

Fig. 19. Nondominated solutions at generations 1, 20, and 200 for CTP3.

solutions found by our approach and corresponding Lagrange multiplier values and subdifferential s.

TABLE II
13 PARETO-OPTIMAL SOLUTIONS.

f_1	f_2	Error	s	λ_1	u	u_{x_1}
0.000	1.114	0.000	0.350	0.692	0.321	0.496
0.003	1.004	0.000	−0.600	1.000	1.000	0.000
0.087	0.949	0.000	0.000	0.385	0.769	
0.170	0.892	0.000	0.000	0.385	0.769	
0.254	0.838	0.000	0.050	0.612	0.716	
0.333	0.776	0.000	0.000	0.385	0.769	
0.408	0.710	0.000	0.000	0.385	0.769	
0.494	0.658	0.000	−0.050	0.679	0.700	
0.573	0.597	0.000	0.050	0.767	0.679	
0.655	0.539	0.000	−0.050	0.711	0.693	
0.734	0.479	0.000	0.000	0.385	0.769	
0.815	0.420	0.000	−0.050	0.775	0.677	
0.901	0.367	0.000	0.050	0.867	0.656	

The above procedure is generic and can be applied to problems facilitating the use of Clarke subdifferential or quasidifferentials [2]. The procedure and the application on a test problem shows that the theoretical optimality conditions can be suitably applied to test optimality of obtained solutions from a numerical optimization algorithm,

VI. CONCLUSIONS

In this paper, we have suggested a post-optimality procedure by which the EA solutions can be verified for being close to theoretical optimal solutions. The systematic procedure suggested here forms a linear set of equations using the KKT conditions of optimality and solves the equations to find a set of Lagrange multipliers for a multi-objective optimization problem. The solution procedure can be a least-square estimate or a minimal-error solution, depending on the dimension of search space and the number of objectives and constraints. On a number of differentiable and non-differentiable test problems, our suggested procedure has been able to demonstrate that solutions obtained by an EA alone or a hybrid EA-local search combination can be very close to being KKT points, thereby providing evidence that EA methodologies coupled with a local search procedure are capable of finding theoretically optimal solutions in multi-objective optimization problems.

We have also argued that such an error metric can be used as a termination criterion for an EA simulation.

Critics of EA approaches should find this paper interesting and hopefully the results should motivate more theoretically oriented researchers to pay further attention to the approaches of evolutionary computation in the coming years.

ACKNOWLEDGEMENT

This study is funded by a gift grant from GM R&D, Bangalore, India.

REFERENCES

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: Theory and algorithms*. Singapore: Wiley, 2004.
[2] C. R. Bector, S. Chandra, and J. Dutta. *Principles of Optimization Theory*. New Delhi: Narosa, 2005.
[3] F. H. Clarke. *Optimization and nonsmooth analysis*. Wiley-Interscience, 1983.
[4] K. Deb. *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: Wiley, 2001.
[5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.
[6] K. Deb and A. Srinivasan. Innovization: Innovating design principles through optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2006)*, pages 1629–1636, New York: The Association of Computing Machinery (ACM), 2006.
[7] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer, Boston, 1999.
[8] G. V. Reklaitis, A. Ravindran, and K. M. Ragsdell. *Engineering Optimization Methods and Applications*. New York : Wiley, 1983.
[9] P. Shukla. On gradient based local search methods in unconstrained evolutionary multi-objective optimization. In *Proceedings of Fourth Conference on Evolutionary Multi-Criterion Optimization (EMO-2007)*, pages 96–110, 2007.