Multimodal Optimization Using a Bi-Objective Evolutionary Algorithm

Kalyanmoy Deb and Amit Saha
Kanpur Genetic Algorithms Laboratory
Indian Institute of Technology Kanpur
PIN 208 016, India
{deb,asaha}@iitk.ac.in

KanGAL Report Number 2009006

January 27, 2010

Abstract

In a multimodal optimization task, the main purpose is to find multiple optimal solutions (global and local), so that the user can have a better knowledge about different optimal solutions in the search space and as and when needed, the current solution may be switched to another suitable optimum solution. To this end, evolutionary optimization algorithms stand as viable methodologies mainly due to their ability to find and capture multiple solutions within a population in a single simulation run. With the preselection method suggested in 1970, there had been a steady suggestion of new algorithms. Most of these methodologies employed a niching scheme in an existing single-objective evolutionary algorithm framework so that similar solutions in a population are de-emphasized in order to focus and maintain multiple distant yet near-optimal solutions. In this paper, we use a completely different strategy in which the single-objective multimodal optimization problem is converted into a suitable bi-objective optimization problem so that all optimal solutions become members of the resulting weak Pareto-optimal set. With the modified definitions of domination and different formulations of an artificially created additional objective function, we present results on as large as 500-optima problems with success. Most past multimodal EA studies considered problems having a few variables. In this paper, we have solved up to 16-variable test-problems having tens of optimal solutions and for the first time suggested multimodal *constrained* test-problems which are scalable in terms of number of optima, constraints, and variables. The concept of using bi-objective optimization for solving single-objective multimodal optimization problems seems novel and interesting and is comparable to a well-known multimodal EA. This paper investigates some useful bi-objective implementations and importantly opens a number of further avenues for research and application.

1 Introduction

Single-objective optimization problems are usually solved for finding a single optimal solution, despite the existence of multiple optima in the search space. In the presence of multiple global and local optimal solutions in a problem, an algorithm is usually preferred if it is able to avoid local optimal solutions and locate the true global optimum.

However, in many practical optimization problems having multiple optima, it is wise to find as many such optima as possible for a number of reasons. First, an optimal solution currently favorable (say, due to availability of some critical resources or satisfaction of some codal principles, or others) may not remain to be so in the future. This would then demand the user to operate at a different solution when such a predicament occur. With the knowledge of another optimal solution for the problem which is favorable to the changed scenario, the user can simply switch to this new optimal solution. Second, the sheer knowledge of multiple optimal solutions in the search space may provide useful insights to the properties of optimal solutions of the problem. A similarity analysis of multiple optimal (high-performing) solutions may bring about useful innovative and hidden principles, similar to that observed in Pareto-optimal solutions in a multi-objective problem solving task (Deb and Srinivasan, 2006).

Thus, in a multimodal optimization task, the goal is to find multiple optimal solutions and not just one single optimum, as it is done in a typical optimization study. If a point-by-point optimization approach is used for this task, the approach must have to be applied many times, every time hoping to find a different optimal solution. The rise of multimodal optimization research in evolutionary computation (EC) field has taken place mainly due to their population approach. By working with a number of population members in each generation, EC algorithms facilitated finding and maintaining multiple optimal solutions from one generation to next. EC researchers employed an additional *niching* operation for this purpose. Instead of comparing population members randomly chosen from the population in their selection operator, a niching method restricts comparison of solutions which are *similar* to each other so that distant good solutions (in the decision variable space) are emphasized. Such efforts were started from 1970 with the concept of preselection (Cavicchio, 1970). Thereafter, crowding, sharing, clearing and many other concepts were suggested to implement the above niching idea.

In this paper, we suggest a different solution methodology based on a bi-objective formulation of the multimodal optimization problem. In addition to the underlying single-objective function, we introduce an additional second objective, the optimum of which will ensure a unique property followed by every optimal solution in the search space which are not shared by any other solution in the search space. Theoretically speaking, one simple strategy would be to use the norm of the gradient vector as the second objective. Since all optimal solutions (global or local) have a zero gradient vector, thereby making the second objective to take its minimum value, the minimization of the original objective function and minimization of the norm of the gradient vector would make only the optimal solutions to lie on the weak Pareto-optimal set of the resulting bi-objective problem. We then suggest a modified evolutionary multi-objective optimization (EMO) procedure to find weak Pareto-optimal solution, thereby finding multiple optimal solutions to the multimodal optimization problem in a single simulation run.

The bi-objective concept is implemented with a couple of pragmatic approaches so that the suggested bi-objective idea is computationally efficient and can also be used in non-differentiable problems. In one of the approaches, we have borrowed a neighborhood evaluation scheme from the classical Hooke-and-Jeeves exploratory search approach (Reklaitis et al., 1983) in order to reduce the computational complexity further. In problems varying from two to 16 variables, having 16 to 500 optima, and having a mix of multiple global and multiple local optima, we demonstrate the working of the proposed procedure. Most existing multimodal EC studies used one to five-variable problems. The success of the proposed procedure to much larger and more complex search spaces remains a hallmark achievement of this paper.

Despite the long history of solving multimodal problems using EC techniques, none of the studies suggested any specific algorithms for handling *constrained* multimodal optimization problems. In this paper, for the first time, we suggest a scalable test-problem in which the number of decision variables, the number of constraints, and the number of optima can all be varied systematically. A slight modification to the proposed unconstrained procedure is able to solve constrained problems having as many as 32 minima, all lying on the intersection surface of multiple constraint boundaries.

In the remainder of the paper, we provide a brief description of past multimodal EC studies in Section 2. The concept of bi-objective optimization for solving multimodal problems is described in Section 3. A couple of gradient-based approaches are presented and simulation results on a simple problem are shown to illustrate the principle of bi-objective solution principle. Thereafter, a number of difficulties in using the gradient approaches in practice are mentioned and a more pragmatic neighborhood based technique is suggested in Section 4. Simulation results on the modified two-variable Rastrigin's functions having as large as 500 optima are shown. Results are compared with that of a popular multimodal optimization procedure. The scalability issue of the neighborhood based technique is discussed and a new Hooke-Jeeves type neighborhood based algorithm is introduced with the proposed bi-objective optimization algorithm in Section 5. Results are shown up to 16-variable problems having as many as 48 optima. Thereafter, a scalable constrained test-problem generator for multimodal optimization is introduced in Section 6. Results up to 10-variable problems and having 16 optima are presented. A repair mechanism shows promise for a computationally faster approach with the proposed bi-objective procedure. Finally, conclusions and a few extensions to this study are highlighted in Section 7.

2 Multimodal Evolutionary Optimization

As the name suggests, a multimodal optimization problem has multiple optimum solutions, of which some can be global optimum solutions having identical objective function value and some can be local optimum solutions having different objective function value. Multimodality in a search and optimization algorithm usually causes difficulty to any optimization algorithm, since there are many attractors, to which the algorithm can become directed to. The task in a multimodal optimization algorithm is to find the multiple optima (global and local) either simultaneously or one after another systematically. This means that, in addition to finding the global optimum solution(s), we are also interested in finding a number of other local optimum solutions. Such information is useful to design engineers and practitioners for choosing an alternate optimum solution, as and when required.

Evolutionary algorithms (EAs) with some changes in the basic framework have been found to be particularly useful in finding multiple optimal solutions simultaneously, simply due to their population approach and their flexibility in modifying the search operators. For making these algorithms suitable for solving multimodal optimization problems, the main challenge has been to maintain an adequate diversity among population members such that multiple optimum solutions can be found and maintained from one generation to another. For this purpose, niching methodologies are employed, in which crowded solutions in the population (usually in the decision variable space) are degraded either by directly reducing the fitness value of neighboring solutions (such as the sharing function approach (Goldberg and Richardson, 1987; Deb and Goldberg, 1989; Mahfoud, 1995; Beasley et al., 1993; Darwen and Yao, 1995)) or by directly ignoring crowded neighbors (crowding approach (DeJong, 1975; Mengsheol and Goldberg, 1999), clearing approach (Pétrowski, 1996; Lee et al., 1999; Im et al., 2004; Singh and Deb, 2006) and others), clustering approach (Yin and Germay, 1993; Streichert et al., 2003) or by other means (Bessaou et al., 2000; Li et al., 2002). A number of recent studies indicates the importance of multimodal evolutionary optimization (Singh and Deb, 2006; Deb and Kumar, 1995; Deb et al., 1993). A number of survey and seminal studies summarize such niching-based evolutionary algorithms to date (Rönkkönen, 2009; Deb, 2001). The niching methods are also used for other non-gradient methods such as particle swarm optimization (Barrera and Coello, 2009; Zhang and Li, 2005; K. Parsapoulos and Vrathis, 2004; Kennedy, 2000; Parrott and Li, 2006), differential evolution (Hendershot, 2004; Li, 2005; Rönkkönen and Lampinen, 2007) and evolution strategies (Shir and Bäck, 2005, 2006). In most of these studies, the selection and/or fitness assignment schemes in a basic EA are modified. In some mating-restriction based EA studies, recombination between similar solutions is also enforced so as to produce less lethal solutions lying on non-optimal regions. The inclusion of a mating restriction scheme to a niching-based EA has been found to improve the on-line performance of the procedure (Deb, 1989).

In this study, we take a completely different approach and employ a bi-objective optimization strategy using evolutionary algorithms to find multiple optimal solutions simultaneously. To the best of our knowledge, such an approach has not been considered yet, but our findings indicate that such a bi-objective approach can be an efficient way of solving multimodal unconstrained and constrained problems.

Despite the long history of systematic studies on unconstrained multimodal optimization, there does not exist too many studies in which constrained multimodal optimization algorithms are specifically discussed. In this paper, we make an attempt to suggest a constrained test-problem generator to create scalable problems having a controlled number of optimum solutions. The proposed algorithm for unconstrained problems is adapted to solve the constrained problems.

3 Multimodal Optimization Using Bi-Objective Optimization

Over the past decade, the principle of evolutionary multi-objective optimization methodologies to find multiple trade-off optimal solutions in a multi-objective optimization problem simultaneously has been extensively used to solve various problem solving tasks. A recent book (Knowles et al., 2008) presents many such ideas and demonstrates the efficacy of such procedures.

In a multimodal optimization problem, we are interested in finding multiple optimal solutions in a single execution of an algorithm. In order to use a multi-objective optimization methodology for this purpose, we need to first identify at least a couple of conflicting (or invariant) objectives for which multiple optimal solutions in a multimodal problem become the trade-off optimal solutions to the corresponding multi-objective optimization problem.

Let us consider the multimodal minimization problem shown in Figure 1 having two minima with different function values:

$$\begin{array}{rl} \text{minimize} & f(x) = 1 - \exp(-x^2)\sin^2(2\pi x), \\ \text{subject to} & 0 \leq x \leq 1. \end{array} \tag{1}$$

If we order the minima according to ascending of their objective function value ($f(x)$), the optimal

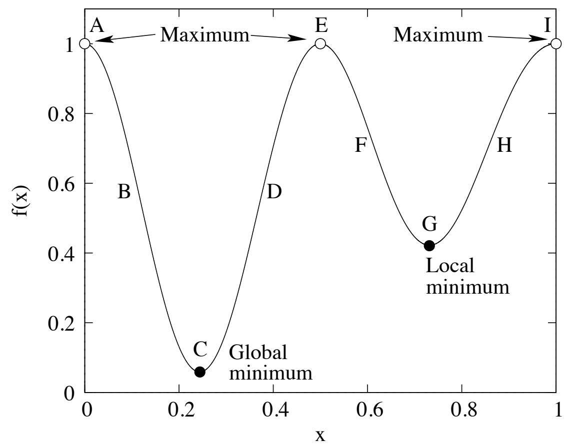


Figure 1: A bi-modal function.

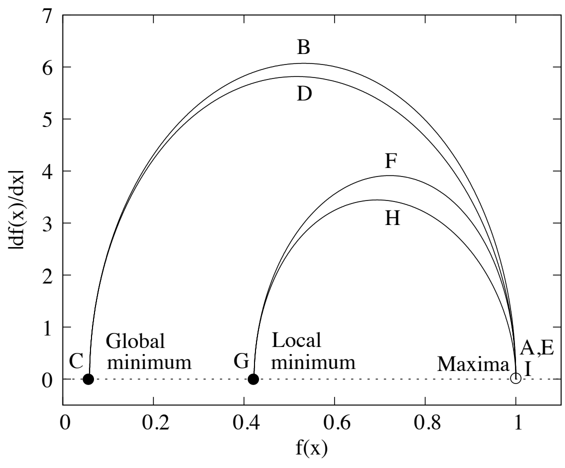


Figure 2: The objective space of the bi-objective problem of minimizing $f(x)$ and $|f'(x)|$.

solutions will line up from the global minimum to the worst local minimum point. In order to have all the minimum points on the trade-off front of a two-objective optimization problem, we need another objective which is either conflicting to $f(x)$ (so they appear on a Pareto-optimal front) or is invariant for all minimum points (so they appear on a weak Pareto-optimal front). We first suggest a gradient-based method and then present a couple of neighborhood based approaches.

3.1 Gradient Based Approach

One property which all minimum points will have in common and which is not shared by other points in the search space (except the maximum an saddle points) is that the derivative of the objective function ($f'(x)$) is zero at these points. We shall discuss about a procedure for distinguishing maximum points from the minimum points later[1], but first we explain the bi-objective concept here. Let us consider the following two objectives:

$$\begin{array}{ll} \text{minimize} & f_1(x) = f(x), \\ \text{minimize} & f_2(x) = |f'(x)| \end{array} \tag{2}$$

in the range $0 \leq x \leq 1$, we observe that the minimum (C and G) and maximum points (A, E and I) of $f(x)$ will correspond to the weak Pareto-optimal points of the above bi-objective problem. Figure 2 shows the corresponding two-dimensional objective space (f_1-f_2). Any point x maps only on the f_1-f_2 curve shown in the figure. Thus, no points other than the points on the shown curve exist in the objective space. It is interesting to observe how f' and f combination makes two different minimum points (C and G) as weak Pareto-optimal solutions of the above bi-objective problem. It is worth reiterating that on the $f_2 = 0$ line, there does not exist any feasible point other than the three points (two corresponding to minimum points and one corresponding to the maximum point) shown in the figure. This observation motivates us to use an EMO procedure in finding all weak Pareto-optimal solutions simultaneously. As a result of the process, we would discover multiple minimum points in a single run of the EMO procedure.

The above bi-objective formulation will also make all maximum points (A, E and I) as weak Pareto-optimal with the minimum points. To avoid this scenario, we can use an alternate second objective, as follows:

$$\begin{array}{ll} \text{minimize} & f_1(x) = f(x), \\ \text{minimize} & f_2(x) = |f'(x)| + (1 - \text{sign}(f''(x))), \end{array} \tag{3}$$

where sign() returns $+1$ if the operand is positive and -1 if the operand is negative. Thus, for a minimum point, $f''(x) > 0$, so the second term in $f_2(x)$ is zero and $f_2(x) = |f'(x)|$. On the other hand, for a maximum point, $f''(x) < 0$, and $f_2(x) = 2 + |f'(x)|$. For $f''(x) = 0$, $f_2(x) = 1 + |f'(x)|$. This modification will make the maximum and $f''(x) = 0$ points dominated by the minimum points. Figure 3 shows the objective space with this modified second objective function on the same problem considered in Figure 1.

3.2 Modified NSGA-II Procedure

The above discussion reveals that the multiple minimum points become different weak Pareto-optimal points of the corresponding bi-objective minimization problem given in equation 2 and 3. However, state-of-the-art EMO algorithms are usually designed to find Pareto-optimal solutions and are not expected to find weak Pareto-optimal solutions. For our purpose here, we need to modify an EMO algorithm to find weak Pareto-optimal points. Here, we discuss the modifications made on a specific EMO algorithm (the NSGA-II procedure (Deb et al., 2002)) to find weak Pareto-optimal solutions:

[1] We have avoided the cases with saddle points in this gradient approach here, but our latter approaches distinguish minimum points from maximum and saddle points.

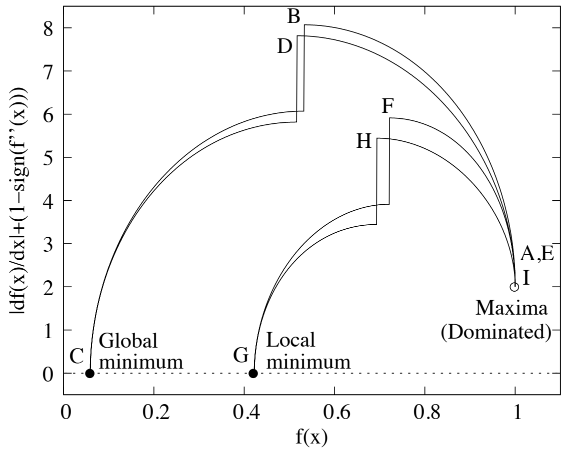


Figure 3: The modified objective space with second-order derivative information makes the maximum points dominated.

1. First, we change the definition of domination between two points a and b. The solution a dominates solution b, if $f_2(a) < f_2(b)$ and $f_1(a) \leq f_1(b)$. Thus, if two solutions have identical f_2 value, they cannot dominate each other. This property will allow two solutions having identical f_2 values to co-survive in the population, thereby allowing us to maintain multiple weak Pareto-optimal solutions, if found in an EMO population.

2. Second, we introduce a *clearing* concept around some selected non-dominated points in the objective space, so as to avoid crowding around minimum points. For this purpose, all points of a non-dominated front (using the above modified domination principle) are first sorted in ascending order of $f(x)$. Thereafter, the point with the smallest $z = f(x)$ solution (having objective values $(f_1, f_2) = (z, z'))$ is kept and all population members (including the current non-dominated front members) in the neighborhood around a box $(f_1 \in (z, z + \delta_f)$ and $f_2 \in (z' + \delta_{f'})$ are cleared and assigned a large non-domination rank. The next solution from the sorted $f(x)$ list is then considered and all solutions around its neighborhood are cleared likewise. After all non-dominated front members are considered, the remaining population members are used to find the next non-dominated set of solutions and the above procedure is repeated. This process continues till all population members are either cleared or assigned a non-domination level. All cleared members are accumulated in the final front and solutions with a larger crowding distance value based on $f(x)$ are preferred.

These two modifications ensure that weak non-dominated solutions with an identical $f_2(x)$ values but different $f(x)$ values are emphasized in the population and non-dominated solutions with smaller $f(x)$ values are preferred and solutions around them are de-emphasized.

3.3 Proof-of-Principle Results with the Gradient Approach

In this subsection, we show the results of the modified NSGA-II procedure on a single-variable problem having five minima:

$$f(x) = 1.1 - \exp(-2x)\sin^2(5\pi x), \quad 0 \leq x \leq 1. \tag{4}$$

Following parameter values are used: population size = 60, SBX probability = 0.9, SBX index = 10, polynomial mutation probability = 0.5, mutation index = 50, $\delta_f = 0.02$, $\delta_{f'} = 0.1$, and maximum number of generations = 100.

Figure 4 shows that the modified NSGA-II is able to find all five minima of this problem. Despite having sparsed points near all five minima, the procedure with its modified domination principle and clearing approach is able to maintain a well-distributed set of points on all five minima.

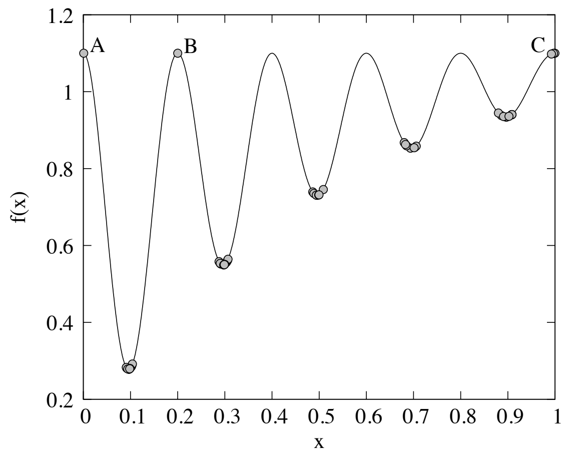


Figure 4: All five minimum points and three maximum points are found by the modified NSGA-II procedure with first derivative information.

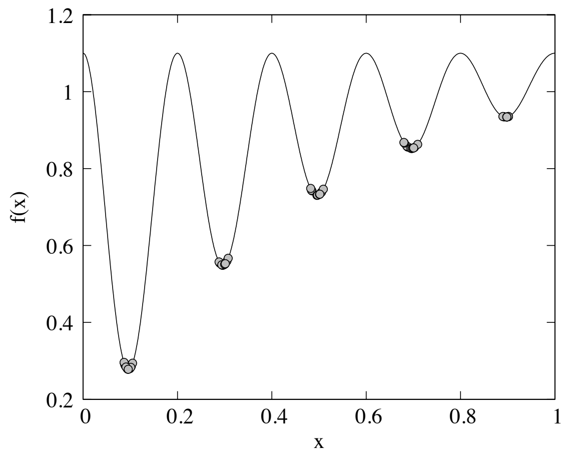


Figure 5: The modified second objective eliminates the maximum points. All five minimum points are found with both first and second derivative information.

In addition to finding all the minimum points, the procedure also finds three of the six maximum points (points A, B, and C, shown in the figure). Since the second objective is the absolute value of $f'(\mathbf{x})$, for both minimum and maximum points, it is zero. Thus, all minimum and maximum points are weak dominated points and become the target of above procedure.

In order to avoid finding the maximum points, next, we use the modified second objective function involving the second derivative $f''(x)$, given in equation 3. Identical parameter settings to that in the above simulation are used here. Since the maximum points now get dominated by any minimum point, this time, we are able to cleanly find all the five minima alone (Figure 5). It is also interesting to note that due to the emphasis on finding the weak Pareto-optimal solutions in the suggested bi-objective approach, some solutions near the optima are found. Since gradient information is used in the second objective, lethal solutions (in non-optimal regions) get dominated by minimum solutions and there is no need for an explicit mating restriction procedure here, which is otherwise recommended in an usual evolutionary multimodal optimization algorithm (Deb and Goldberg, 1989).

3.4 Practical Difficulties with Gradient Based Methods

Although the concept of first and second-order gradient information allows all minimum points to have a common property which enabled us to find multiple minimum points using a bi-objective procedure, the gradient based approaches have certain well-known difficulties.

First, such a method has restricted applications, because the method cannot be applied in problems which do not have gradient at some intermediate points or at the minimum points. The gradient information drives the search towards the minimum points and if gradient information is not accurate, such a method is not likely to work well.

Second, the gradient information is specific to the objective function and the range of gradient values near one optimum point may vary from another, depending on flatness of the function values near optimum points. A problem having a differing gradient values near different optima will provide unequal importance to different optima, thereby causing an algorithm difficulty in finding all optimum points in a single run.

Third and a crucial issue is related to the computational complexity of the approach. Although the first-order derivative approach requires $2n$ function evaluations for each population member, this comes with the additional burden of finding the maximum points. To avoid finding unwanted maximum points, the second-derivative approach can be used. But the computation of all second derivatives numerically takes $2n^2 + 1$ function evaluations (Deb, 1995), which is prohibitory for large-sized problems.

To illustrate the above difficulties, we consider a five-variable modified Rastrigin's function having eight minimum points (function described later in equation 6 with $k_1 = k_2 = k_3 = 1$, $k_4 = 2$ and $k_5 = 4$). The norm of the gradient vector ($\|\nabla f(x)\|$) is used as the second objective. When we apply our proposed gradient-based approach with identical parameter setting as above to solve this problem, out of eight minimum points only two are found in the best of 10 runs. Figure 6 shows the objective space enumerated with about 400,000 points created by equispaced points in each variable within $x_i \in [0, 1]$. The eight minimum points are shown on the $f_2 = 0$ line and paths leading to these minimum points obtained by varying x_5 and by fixing x_1 to x_4 to their theoretical optimal values. The obtained two minimum points are shown with a diamond. It is interesting to note from the figure that points near the minimum solutions are scarce. On the other hand, there are many maximum points lying on the $f_2 = 0$ line, which are accessible from the bulk of the objective space. With the first derivative information, an algorithm can get lead to these maximum points.

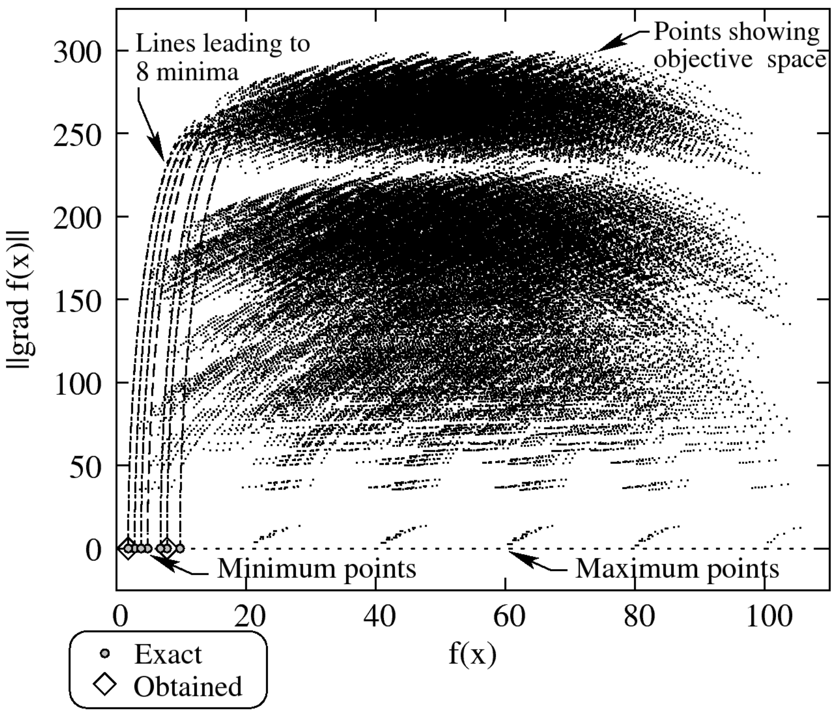


Figure 6: The two-objective problem is illustrated for the five-variable modified Rastrigin's function. Variation of x_5 by fixing all other variables to their optimal values show eight paths (two almost coincide having almost equal function value) culminating in eight multimodal minimum points.

The above study is pedagogical and demonstrates that a multimodal optimization problem can be in principle converted to an equivalent bi-objective problem by using first and second-order derivative based optimality conditions, however there are implementation issues which will restrict the use of the gradient-based methods to reasonably higher-dimensional problems. Nevertheless, the idea of using a bi-objective optimization technique to find multiple optimum points in a multimodal optimization problems is interesting and next we suggest a more pragmatic approach.

4 Multimodal Optimization Using Neighboring Solutions

Instead of checking the gradients for establishing optimality of a solution ($\mathbf{x}$), we can simply compare a sample of neighboring solutions with the current solution $\mathbf{x}$. The second objective function $f_2(\mathbf{x})$ can be assigned as the count of the number of neighboring solutions which are better than the current solution $\mathbf{x}$ in terms of their objective function ($f(\mathbf{x})$) values. Figure 7 illustrates the idea on a single-variable hypothetical problem. Point A is the minimum point of this function and both neighboring solutions are worse than point A, thereby making $f_2(A) = 0$. On the other hand, any other point (which is not a minimum point) will have at least one neighboring point worse than that point. For example, point B has one worse solution, thereby making $f_2(B) = 1$, and point C has two points worse having $f_2(C) = 2$. To implement the

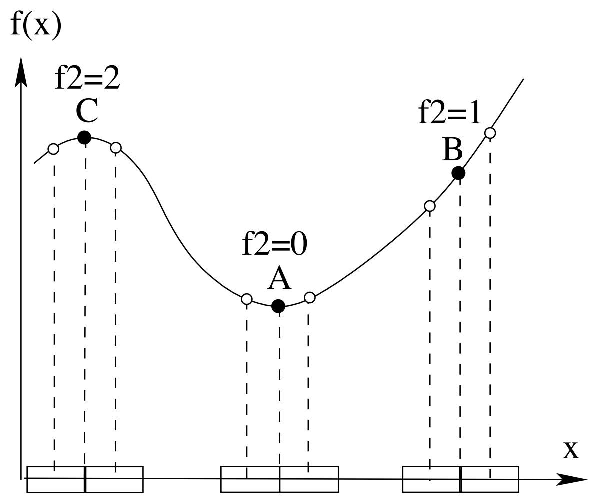


Figure 7: The second objective is obtained by counting number of worse neighboring solutions. The minimum solution will correspond to $f_2 = 0$.

idea, a set of H solutions in the neighborhood of $\mathbf{x}$ can be computed systematically. One simple approach would be to use the Latin hypercube sampling (LHS) method with H divisions in each dimension and pick a combination of divisions such that a representative from each division in every dimension is present. For a large-dimensional problem and with a reasonably large H, this procedure may not provide a good sampling of the neighborhood. Another idea would be to make k divisions ($k = 2$, for example) along each dimension and sample a point in each division, thereby constructing a total of $H = k^n$ points to compute the second objective. Clearly, other computationally more efficient ideas are possible and we shall suggest one such approach in a later section.

4.1 Modified NSGA-II with Neighboring Point Count

As mentioned, the second objective $f_2(\mathbf{x})$ is computed by counting the number of neighboring solutions better in terms of objective function value ($f(\mathbf{x})$) than the current point. Thus, only solutions close to a minimum solution will have a zero count on f_2, as locally there does not exist any neighboring solution smaller than the minimum solution. Since f_2 now takes integer values only, the f_2-space is discrete.

To introduce the niching idea as before on the objective space, we also consider all solutions having an identical f_2 value and clear all solutions within a distance δ_f from the minimum $f(\mathbf{x})$ value. Figure 8 illustrates the modified non-domination procedure. A population of 40 solutions are plotted on a f_1-f_2 space. Three non-dominated fronts are marked in the figure to make the

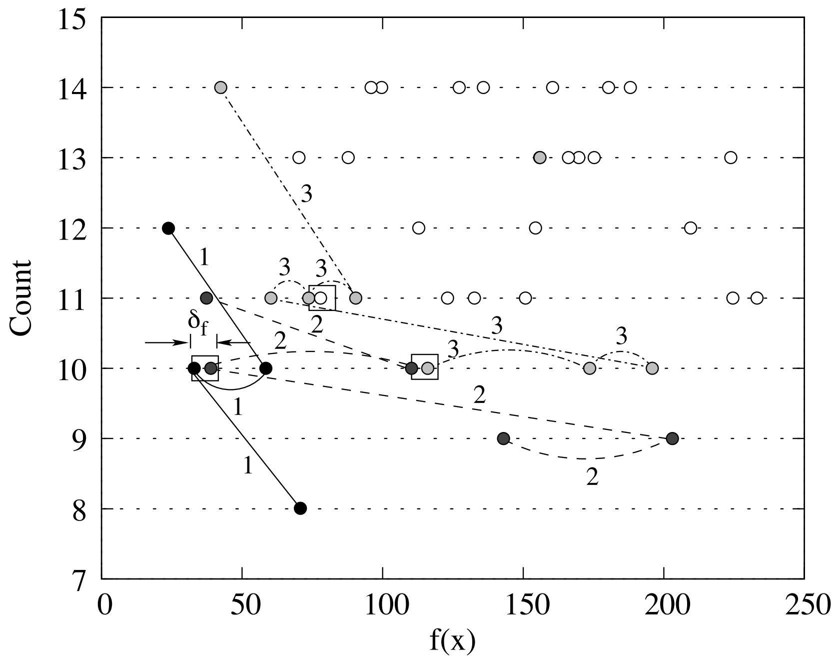


Figure 8: A set of 40 members are ranked according to increasing level of non-domination based on the modified principle and niching operation. Three non-dominated fronts are marked to illustrate the niched domination principle.

domination principle along with the niching operation clear. The population has one point with $f_2 = 8$, two points with $f_2 = 9$, seven points with $f_2 = 10$, and so on, as shown in the figure. The point with $f_2 = 8$ dominates (in the usual sense) both $f_2 = 9$ points and the right-most four $f_2 = 10$ points. In an usual scenario, the minimum f_1 point among $f_2 = 10$ points would have dominated other six $f_2 = 10$ points. But due to the modified domination principle, three left-most points with $f_2 = 10$ are non-dominated to each other and are non-dominated with the $f_2 = 8$ point as well. However, the niching consideration makes the second point on $f_2 = 10$ line from left dominated by the first point. This is because the second point is within δ_f distance from the first point. The third point on $f_2 = 10$ line is more than δ_f distance away from the left-most point on $f_2 = 10$ line and hence the third point qualifies to be on the first non-dominated front. Since all points with $f_2 = 11$ get dominated by the left-most point on the $f_2 = 10$ line, none of them qualifies to be on the first non-dominated front. However, the left-most point with $f_2 = 12$ is non-dominated (in the usual sense) with the other first front members. Thus, the best non-dominated front is constituted with one point having $f_2 = 8$, two points having $f_2 = 10$ and one point having $f_2 = 12$. These points are joined a solid line in the figure to show that they belong to the same front.

Similarly, the second non-dominated front members are identified by using the modified domination principle and the niching operation and are marked in the figure with a dashed line. It is interesting to note that these front classification causes a different outcome from that the usual domination principle would have produced. Points lying sparsely in smaller f_2 lines and having smaller f_1 value are emphasized. Such an emphasis will eventually lead the EMO procedure to solutions on $f_2 = 0$ line and the niching procedure will help maintain multiple solutions in the population.

4.2 Proof-of-Principle Results

In this subsection, we present simulation results of the above algorithm, first on a number of two-variable test-problems and then on higher-dimensional problems. We consider the following two-variable problem:

$$\begin{array}{rl} \text{minimize} & f(x_1, x_2) = (x_1^2 + x_1 + x_2^2 + 2.1x_2) + \sum_{i=1}^{2} 10\,(1 - \cos(2\pi x_i))\,, \\ \text{subject to} & 0.5 \le x_1 \le (K_1 + 0.5), \quad 0.5 \le x_2 \le (K_2 + 0.5). \end{array} \tag{5}$$

There is a minimum point close to every integer value of each variable within the lower and upper bounds. Since there are K_i integers for each variable within $x_i \in [0.5, K_i + 0.5]$, the total number of minima in this problem are $M = K_1K_2$. First, we consider $K_1 = K_2 = 4$, so that there are 16 minima. We use a population of size of 160 (meaning an average of 10 population members are expected around each minimum point), SBX probability of 0.9, SBX index of 10, polynomial mutation probability of 0.5, mutation index of 20, and run each NSGA-II for 100 generations. The parameter $\delta_f = 0.1$ is used here. To make proof-of-principle runs, 4 neighboring points are created within ± 0.1 of each variable value. Figure 9 shows the outcome of the proposed procedure: All 16 minima are found by the procedure in a single simulation run. Since the best non-dominated front members are declared as outcome of the proposed procedure, the optimum solutions are cleanly found. To have a better understanding of the working principle of the proposed procedure, we

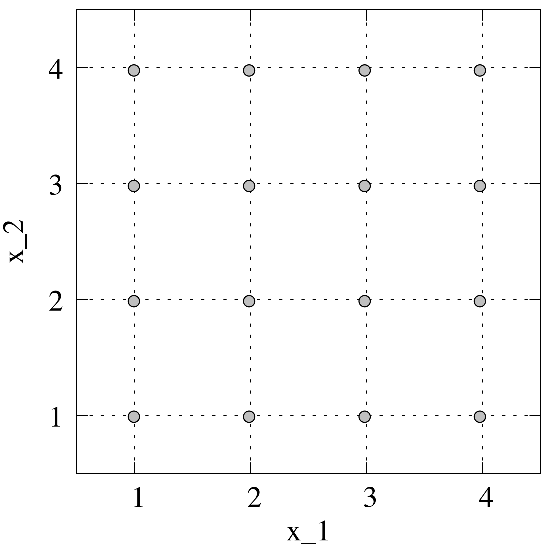


Figure 9: All 16 minimum points are found by the proposed procedure.

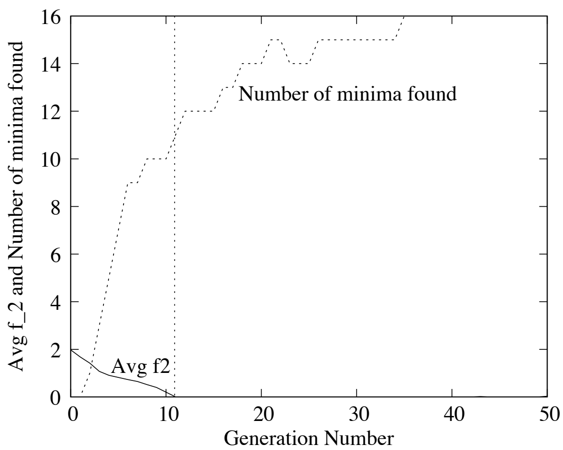


Figure 10: Variation of population average f_2 and the number of minima found with generation counter.

record the population average f_2 value and the number of minima found in every generation, and plot them in Figure 10. It is interesting to note that on an average, the random initial population members have about 2 (out of 4) neighboring points better than them. As generations increase, points with lesser number of better neighboring points are discovered and eventually at generation 12, all population members have no better neighboring points. Since neighboring points are created around a finite neighborhood, a point with $f_2 = 0$ does not necessarily mean that the point is a minimum point (it can be a point close to the true minimum point), but the emphasis on minimum $f_1(\mathbf{x})$ provided in the domination principle allows solutions close to the true minimum points to be found with subsequent generations.

At each generation, the number of population members which are within a normalized Euclidean distance of 0.03 from the true minimum in each variable dimension are noted and plotted in the figure. Interestingly, it took initial three generations for the proposed procedure to find a single solution close to a minimum point and thereafter increasing number of points are found close to true minimum points with generations. At generation 12 when all population members have $f_2 = 0$, 12 out of 16 minima are found. But eventually at generation 35, all 16 minimum points are found by NSGA-II operators. We would like stress the fact that the proposed modified domination principle and niching operation together are able to emphasize and maintain various minimum points and also able to accommodate increasing number of minimum points, as and when they are discovered. This is a typical performance of the proposed NSGA-II procedure, which we have observed over multiple runs.

4.3 Scalability Results up to 500-Optima Problems

Staying with the same two-variable problem, we now perform a scalability study with number of optimal points. We vary the population size as $N = 10m$ (where m is the number of minima) and other parameter values same as in the case of 16-minima problem presented above. We construct a 20-minima problem ($K_1 = 5$, $K_2 = 4$), a 50-minima problem ($K_1 = 10$, $K_2 = 5$), a 100-minima problem ($K_1 = 10$, $K_2 = 10$), a 200-minima problem ($K_1 = 20$, $K_2 = 10$), and a 500-minima problem ($K_1 = 25$, $K_2 = 20$). Final population members for the 100-minima and the 500-minima problems are shown in Figures 11 and 12, respectively. In the case of 100-minima problem, 97 minimum points are found and in the case of 500-minima problem, 483 minimum points are found.

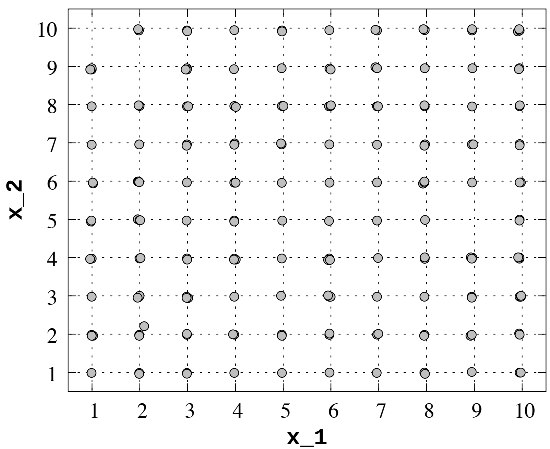


Figure 11: 97 out of 100 minimum points are found by the proposed procedure.

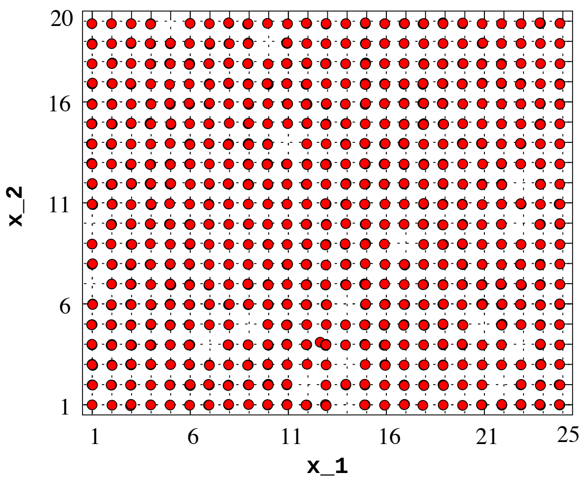


Figure 12: 483 out of 500 minimum points are found by the proposed procedure.

Figure 13 shows the number of minima found in each of the five problems studied with 10 runs for each problem. The average number of obtained minima are shown with circles and worst and best numbers are shown with bars (which can be hardly seen in the figure). The proposed algorithm is found to be reliably scalable in finding up to 500 optimal points in a single simulation.

The modified domination and the niching strategy seems to work together in finding almost all available minima (and as large as 500) for the chosen problems.

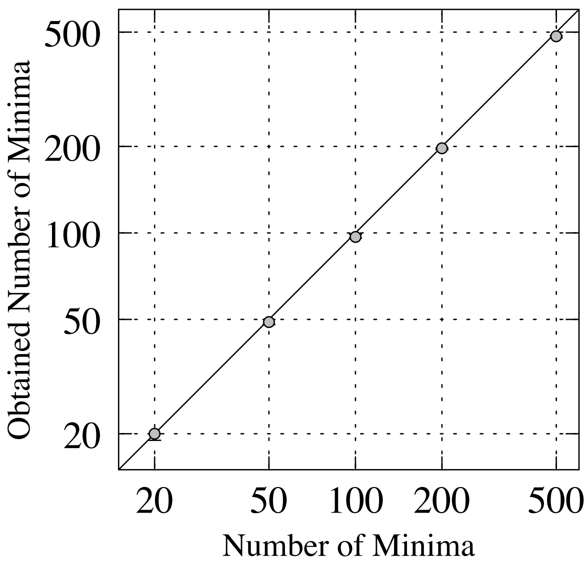


Figure 13: Almost all minimum points are found on two-variable problems by the proposed procedure.

4.4 Comparison with the Clearing Approach

Before we leave this two-variable multiple-minima problem, we would like to test the working of an existing state-of-the-art multimodal evolutionary algorithm. In some earlier studies (Pétrowski, 1996; Singh and Deb, 2006), it has been reported that the clearing approach performs the better in most occasions compared to other evolutionary multimodal approaches. Here, we implement the original clearing approach (Pétrowski, 1996) on a real-parameter genetic algorithm using the stochastic remainder roulette-wheel selection operator (Goldberg, 1989), SBX recombination operator, and the polynomial mutation operator (Deb, 2001). Since proportionate selection operator is used in the clearing approach, only maximization problems (and problems having non-negative objective values) can be solved. In the clearing approach, population members are first sorted based on their objective function value from best to worst. Then, except the first κ sorted members, all subsequent members around the niche radius distance from the best population member is forced to have an objective value zero. Thereafter, the next κ solutions are retained and all other solutions from the niche radius distance are *cleared* and assigned a value of zero. The stochastic remainder selection operator is then applied on the modified objective values.

We have used different values of κ (the parameter signifying the number of top solutions emphasized within each niche) and niche radius for clearing to determine which combinations works the best for this problem. We have attempted to solve the 100-peaked problem given in equation 5 by converting the problem to a maximization problem as follows (with $K_1 = K_2 = 10$):

$$f(x_1, x_2) = 275 - (x_1^2 + x_1 + x_2^2 + 2.1x_2) - \sum_{i=1}^{2} 10\,(1 - \cos(2\pi x_i))\,.$$

After a number of simulations with different κ and niche radius values, we found that $\kappa = 3$ and niche radius of 0.7 worked the best for this problem.

Figure 14 shows the population members having non-zero function values obtained using the clearing GA with $N = 1,000$. Other GA parameters are the same as used in our proposed approach. Note that to solve the minimization version of 100-optima problem (Figure 11), our proposed approach required a population of size 1,000. It is clear that the clearing approach is not able to cleanly find a number of optimal points. When we increase the population size by 1,000 at a time, we have observed that with $N = 5,000$ the clearing approach is able to find all 100 maximum points adequately. Some unwanted boundary points are also found by this algorithm. The corresponding points are shown in Figure 15. Our bi-objective approach requires four evaluations around each population members, thereby implying that on a similar 100-minima problem the bi-objective approach is equivalent to the clearing approach in finding multiple optimal solutions. Our approach requires comparatively lesser population size but evaluation of additional neighboring solutions, whereas the clearing approach does not require any neighboring solution evaluation but requires a bigger population size. Motivated by this performance of our proposed approach, we now discuss computationally faster methodology with our proposed approach.

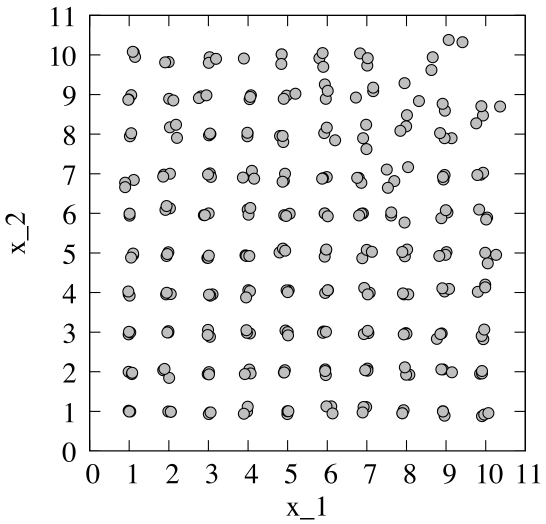


Figure 14: The clearing based GA with 1,000 population members does not perform well.

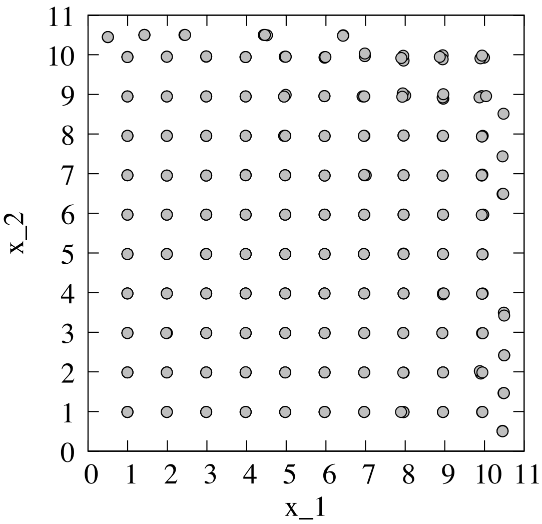


Figure 15: The clearing based GA with 5,000 population members find all minima adequately.

4.5 Difficulties in Handling a Large Number of Variables

In the above simulations with the proposed bi-objective procedure, two points were created around the current point in each variable dimension to determine the second objective function. We realize that this strategy is not scalable to many variables, as the required number of neighboring points will grow exponentially (2^n) with the number of variables. In the following section, we suggest a computationally quick procedure motivated by a classical optimization principle which requires linearly increasing evaluation points with the number of variables.

5 A Faster Approach with Hooke-Jeeve's Exploratory Search

Instead of generating H solutions at random around a given solution point ($\mathbf{x}$), we can choose the points judiciously by using the way an exploratory search is performed in a Hooke-Jeeve's classical optimization algorithm (Reklaitis et al., 1983). The procedure starts with the first variable ($i = 1$)

dimension and creates two extra points $x_i^c \pm \delta_i$ around the current solution $\mathbf{x}^c = \mathbf{x}$. Thereafter, three solutions $(\mathbf{x}^c - \delta_i\mathbf{e}_i, \mathbf{x}^c, \mathbf{x}^c + \delta_i\mathbf{e}_i)$ (where $\mathbf{e}_i$ is the unit vector along i-th variable axis in the n-dimensional variable space) are compared with their function values and the best is chosen. The current solution $\mathbf{x}^c$ is then moved to the best solution. Similar operations are done for $i = 2$ and continued for the remaining variables. Every time a solution having a better objective value than the original objective value $(f(\mathbf{x}))$ is encountered, the second objective value $f_2(\mathbf{x})$ is incremented by one. This procedure requires a total of $2n$ function evaluations to compute $f_2(\mathbf{x})$ for every solution $\mathbf{x}$ in solving an n-dimensional problem. Although this procedure requires a similar amount of additional function evaluations $(2n)$ to that in a numerical implementation of the first-derivative approach described earlier, the idea here is more generic and does not have the difficulties associated with the first-derivative approach. Later in Section 6.5, we shall discuss a repair procedure which is potential for a further reduction in function evaluations.

5.1 Results with H-J Approach

The past multimodal evolutionary algorithms studies handled problems having a few variables. In this section, we present results on a number of multi-dimensional multimodal test-problems. First, we define a scalable n-variable test-problem, as follows:

$$\text{MMP}(n): \begin{array}{rl} \text{minimize} & f(\mathbf{x}) = \sum_{i=1}^{n} 10(1+\cos(2\pi k_i x_i)) + 2k_i x_i^2, \\ \text{subject to} & 0 \le x_i \le 1, \quad i = 1, 2, \ldots, n. \end{array} \tag{6}$$

This function is similar to the Rastrigin's function. Here, the total number of global and local minima are $M = \Pi_{i=1}^{n} k_i$. We use the following k_i values for the three MPP problems, each having 48 minimum points:

$$\begin{array}{rlllll} \text{MPP(4):} & k_1 = 2 & k_2 = 2 & k_3 = 3 & k_4 = 4. & \\ \text{MPP(8):} & k_2 = 2 & k_4 = 2 & k_6 = 3 & k_8 = 4 & k_i = 1. \ \text{for} \ i = 1,3,5,7, \\ \text{MPP(16):} & k_4 = 2 & k_8 = 2 & k_{12} = 3 & k_{16} = 4 & k_i = 1, \ \text{for} \ i = \text{1-3}, \text{5-7}, \text{9-11}, \text{13-15}. \end{array}$$

All of these problems have one global minimum and 47 local minimum points. The theoretical minimum variable values are computed by using the first and second-order optimality conditions and solving the resulting root finding problem numerically:

$$x_i - 5\pi \sin(2\pi k_i x_i) = 0.$$

In the range $x_i \in [0, 1]$, the above equation has k_i roots. Different k_i values and corresponding k_i number of x_i variable values are shown in Table 1. In all runs, we use $\delta_i = 0.005n$. For the

Table 1: Optimal x_i values for different k_i values. There are k_i number of solutions for each k_i.

	k_i			
	1	2	3	4
x_i	0.49498	0.24874	0.16611	0.12468
		0.74622	0.49832	0.37405
			0.83053	0.62342
				0.87279

domination definition, we have used $\delta_f = 0.5$. For all problems, we have used a population size of $N = 15\max(n, M)$, where n is the number of variables and M is the number of optima. Crossover and mutation probabilities and their indices are the same as before.

First, we solve the four-variable ($n = 4$) problem. Figure 16 shows the theoretical objective value ($f(\mathbf{x})$) of all 48 minimum points with diamonds. The figure also shows the minimum points found by the proposed H-J based NSGA-II procedure with circles. Only 14 out of 48 minimum points are found. We explain this (apparently poor) performance of the proposed procedure here.

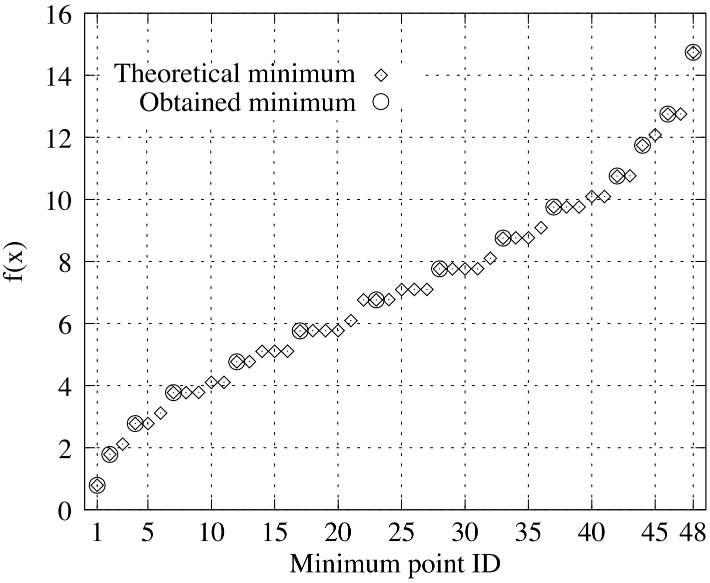


Figure 16: Only 14 of the 48 minimum points are found for the MMP(4) problem by the objective-space niching algorithm.

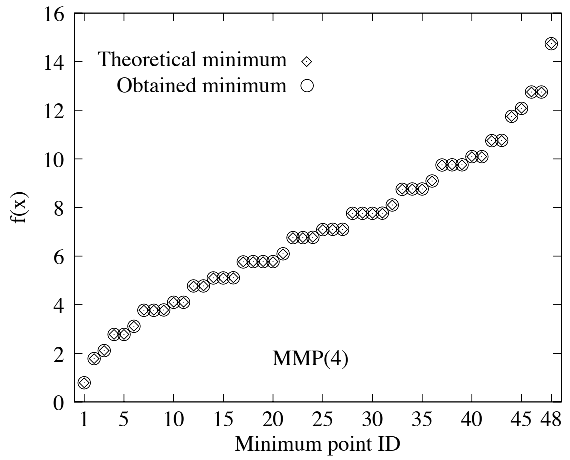
The figure shows that although there are 48 minima, a number of minima have identical objective values. For example, 14, 15, and 16-th minimum points have identical function value of 5.105. Since the modified domination procedure de-emphasizes all solutions having a difference of $\delta_f = 0.2$ function value from each other, from a cluster of multiple minimum points having identical function value only one of the minimum points is expected be found; other minimum solutions will be cleared. This scenario can be observed from the figure. Once a point is found, no other minimum point having a similar value is obtained by the algorithm.

5.2 Multiple Global Optimal Solutions

In order to find multiple minimum points having identical or almost equal function values (multiple global minima, for example), we need to perform an additional variable-space niching among closer objective solutions by emphasizing solutions having widely different variable vectors. To implement the idea, we check the normalized Euclidean distance (in the variable space) of any two solutions having function values within δ_f. If the normalized distance is greater than another parameter δ_x, we assign both solutions an identical non-dominated rank, otherwise both solutions are considered arising from the same optimal basin and we assign a large dominated rank to the solution having the worse objective value $f(\mathbf{x})$. In all runs here, we use $\delta_x = 0.2$ here.

Figure 17 summarizes the result obtained by this modified procedure. All 48 minimum points are discovered by the modified procedure. All 10 runs starting from different initial random populations find 48 minima in each case.

Next, we solve the $n = 8$ and 16-variable problems using the modified procedure. Figures 18 and 19 show the corresponding results. The modified procedure is able to find all 48 minimum points on each occasion and in each of the 10 independent runs.



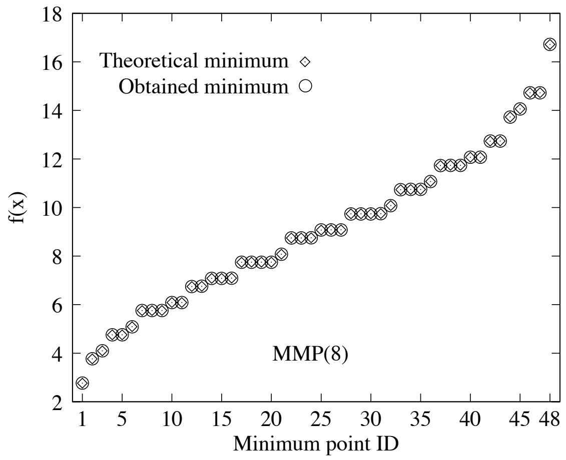


Figure 17: All 48 minimum points are found for the four-variable MMP(4) problem.

Figure 18: All 48 minimum points are found for the eight-variable MMP(8) problem.

There are a few multimodal problems having 8 or 16 variables that have been attempted in the existing multimodal evolutionary optimization studies. Here, we demonstrate successful working of our procedure on such relatively large-sized problems and also having as many as 48 minimum points.

5.3 48 Global Optimal Solutions

Next, we modify the objective function as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{n} 10 + 9\cos(2\pi k_i x_i).$$

In this problem, we use $n = 16$ and k_i values are chosen as in MMP(16) above, so that there are 48 minima having an identical objective value of $f = 16$. Interestingly, there is no locally minimum solution in this problem. We believe this problem will provide a stiff challenge to the variable-space niching procedure, as all 48 global minimum points will lie on an identical point on the objective space and the only way to distinguish them from each other would be to investigate the variable space and emphasize distant solutions.

Figure 20 shows that the proposed niching-based NSGA-II procedure is able to find all 48 minimum points for the above 16-variable modified problem. Exactly the same results are obtained in nine other runs, each starting with a different initial population. It is interesting that a simple variable-space niching approach described above is adequate to find all 48 minimum points of the problem.

Most multimodal EAs require additional parameters to identify niches around each optimum. Our proposed H-J based procedure is not free from additional parameters. However, both δ_f and δ_x parameters directly control the differences between any two optima in objective and decision variable spaces which are desired. Thus, the setting of these two parameters can be motivated from a practical standpoint. However, we are currently pursuing ways of self-adapting these two parameters based on population statistics and the desired number of optima in a problem. Nevertheless, the demonstration of successful working of the proposed approach up to 16-variable problems and having as many as 48 optima amply indicates its efficacy and should motivate further future studies in favor of the proposed bi-objective concept.

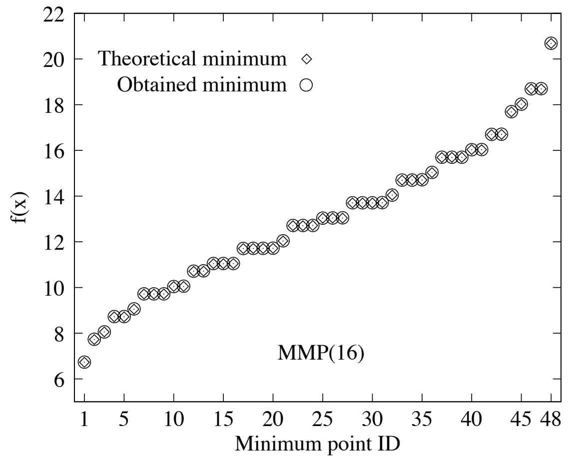


Figure 19: All 48 minimum points are found for the 16-variable MMP(16) problem.

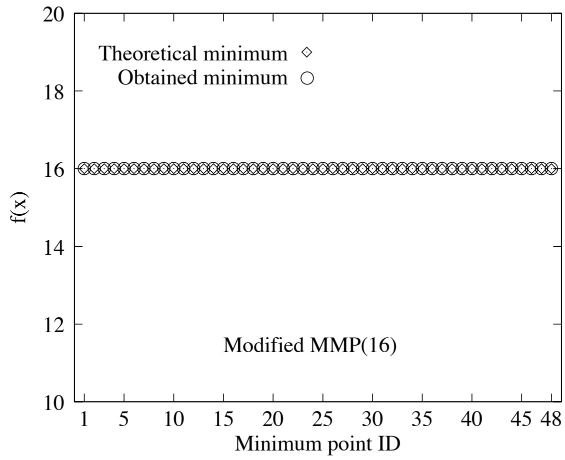


Figure 20: All 48 global minimum points are found for the modified MMP(16) problem.

6 Constrained Multimodal Optimization

Most studies on evolutionary multimodal optimization concentrated on solving unconstrained multimodal problems and some unconstrained test-problem generators (Rönkkönen et al., 2008) now exist. To our knowledge, a systematic study on suggesting test-problems on *constrained* multimodal optimization problems and algorithms to find multiple optima which lie on constraint boundaries do not exist. However, practical optimization problems most often involve constraints and in such cases multiple optima are likely to lie on one or more constraint boundaries. It is not clear (and has not demonstrated well yet) whether existing evolutionary multimodal optimization algorithms are adequate to solve such constrained problems. In this paper, we first design scalable test-problems for this purpose and then extend our proposed H-J based evolutionary multimodal optimization algorithm to handle constraints. Simulation results are then presented on the test-problems.

6.1 Constrained Multimodal Test-Problems

Here, we suggest a simple test-problem construction procedure which is extendable to an arbitrary dimension of the search space and have a direct control on the number of global and local optima the test-problem will have. To illustrate the construction procedure, we first consider a two-variable problem. There are two non-convex constraints and the minimum solutions lie on the intersection on the constraint boundaries. We call these problems CMMP(n, G, L), where n is the number of variables, G is the number of global minima and L is the number of local minima in the problem.

$$\begin{array}{lrl} \textsc{Problem} & \text{minimize} & f(x_1,x_2) = x_1^2 + x_2^2, \\ \text{CMMP}(2,4,0): & \text{subject to} & g_1(x_1,x_2) \equiv 4x_1^2 + x_2^2 \geq 4, \\ & & g_2(x_1,x_2) \equiv x_1^2 + 4x_2^2 \geq 4, \\ & & -3 \leq x_1 \leq 3, \quad -3 \leq x_2 \leq 3. \end{array} \tag{7}$$

Figure 21 shows the constraint boundaries and the feasible search space (shaded region). Since the square of the distance from the origin is minimized, there are clearly four *global* minima in this problem. These points are intersection points of both ellipses (denoting the constraint boundaries) and are shown in the figure with circles. The optimal points are at $x_1^* = \pm\sqrt{0.8}$ and $x_2^* = \pm\sqrt{0.8}$

with a function value equal to $f^* = 1.6$. Contour lines of the objective function shows that all four minimum points have the same function value.

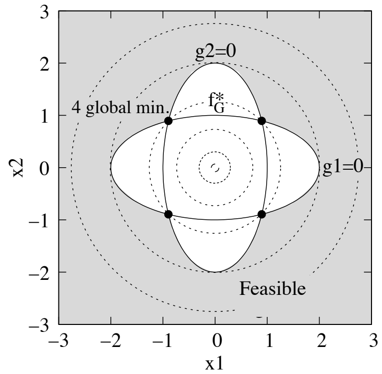


Figure 21: Four global minimum points for the two-variable, two constraint CMMP(2,4,0) problem.

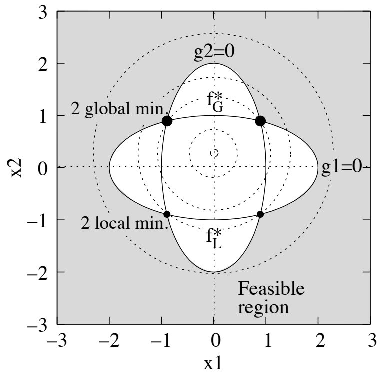


Figure 22: Two global and two local minimum points for the two-variable, two constraint CMMP(2,2,2) problem.

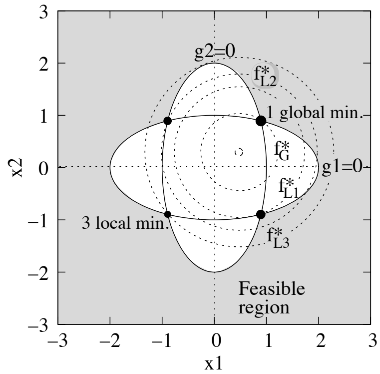


Figure 23: One global and three local minimum points for the two-variable, two constraint CMMP(2,1,3) problem.

Interestingly, if we modify the objective function as $f(x_1, x_2) = x_1^2 + (x_2 - 0.2)^2$ (problem CMMP(2,2,2)), there are two global minima: $(\sqrt{0.8}, \sqrt{0.8})$, $(-\sqrt{0.8}, \sqrt{0.8})$ with a function value $f_G^* = 1.282$ and there are two local minima: $(\sqrt{0.8}, -\sqrt{0.8})$, $(-\sqrt{0.8}, -\sqrt{0.8})$ with the function value $f^* = 1.998$. Figure 22 shows that the contour of the objective function starting with a value zero at $(0, 0.2)$ and then increasing with a point's distance from $(0, 0.2)$. The global and local minima are also marked in the figure.

If the objective function is modified as $f(x_1, x_2) = (x_1 - 0.3)^2 + (x_2 - 0.2)^2$ (the problem CMMP(2,1,3)), there are three different local minima and one global minimum, as shown in Figure 23. The global minimum is at $(\sqrt{0.8}, \sqrt{0.8})$ with a function value $f^* = 0.836$. The three local minima are $(\sqrt{0.8}, -\sqrt{0.8})$ with a function value $f^* = 1.551$, $(-\sqrt{0.8}, \sqrt{0.8})$ with a function value $f^* = 1.909$ and $(-\sqrt{0.8}, -\sqrt{0.8})$ with a function value $f^* = 2.624$.

6.2 Higher-Dimensional Constrained Test-Problems

For higher-dimensional (with n variables) search spaces, the above concept can be extended with higher-dimensional ellipses as constraint boundaries and with more constraints. The following construction can have $J = 1$ to at most $J = n$ constraints. For J constraints, there are 2^J minima. In the following problem, we have 2^J global minima:

$$\text{PROBLEM CMMP}(n, 2^J, 0): \quad \begin{array}{rl} \text{minimize} & f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2, \\ \text{subject to} & g_1(\mathbf{x}) \equiv x_1^2 + 4x_2^2 + 9x_3^2 + \cdots + n^2 x_n^2 \geq n^2, \\ & g_2(\mathbf{x}) \equiv n^2 x_1^2 + x_2^2 + 4x_3^2 + \cdots + (n-1)^2 x_n^2 \geq n^2, \\ & \vdots \\ & g_J(\mathbf{x}) \equiv C_{J,1}^2 x_1^2 + C_{J,2}^2 x_2^2 + C_{J,3}^2 x_3^2 + \cdots + C_{J,n}^2 x_n^2 \geq n^2, \\ & -(n+1) \leq x_i \leq (n+1), \quad \text{for } i = 1, 2, \ldots, n, \end{array} \tag{8}$$

where

$$C_{j,k} = \begin{cases} (n - j + k + 1) \mod n, & \text{if } (n - j + k + 1) \mod n \neq 0, \\ n, & \text{otherwise.} \end{cases}$$

In order to find the minimum points, we realize that they will lie on the intersection of constraint boundaries. Thus, we construct the Lagrange function:

$$L(\mathbf{x}, \mathbf{v}) = f(\mathbf{x}) - \sum_{j=1}^{J} v_j (g_j(\mathbf{x}) - n^2). \tag{9}$$

and set partial derivatives to zero: $\partial L/\partial x_i = 0$ and $\partial L/\partial v_j = 0$. The second set of conditions result in the constraint conditions themselves. The first set of n equations are as follows:

$$x_i^*(1 + \sum_{j=1}^{J} v_j C_{j,i}^2) = 0, \quad \text{for } i = 1, \ldots, n. \tag{10}$$

For $J < n$, $(n - J)$ variables (x_i) will take a value zero to satisfy the above equations. From the remaining J linear equations, we can compute J optimal Lagrange multipliers v_j. To find the remaining J optimal variable values x_i^*, we set $x_i = 0$ for $(n - J)$ chosen variables in constraint conditions. This process will lead to J equations with J variables (x_i), which can be solved to find x_i^* values. It turns out that by choosing different combinations of $(n - J)$ variables to take a value zero, there will be multiple solutions to the above set of equations. We then compute the function values of each solution and choose the solution(s) which make the function value the smallest. It turns out that the smallest objective value occurs for $x_i^* = 0$, $i = J, J+1, \ldots, (n-1)$. The remaining x_i^* variables for $i = \{1, \ldots, (J-1), n\}$ take non-zero values. We have computed these values for $J = 1$ to $J = 4$ and present the results below.

1. For one constraint ($J = 1$) leading to the problem CMMP$(n, 2, 0)$, there are two global minima: $(0, 0, \ldots, 1)$ and $(0, 0, \ldots, -1)$ having a function value of $f^* = 1$.

2. For two constraints ($J = 2$) leading to CMMP$(n, 4, 0)$, there are four global minima:

$$\begin{aligned} x_1^* &= \pm\sqrt{\frac{n^2(2n-1)}{(n^2+n-1)(n^2-n+1)}}, \\ x_i^* &= 0, \quad \text{for } i = 2, \ldots, (n-1), \\ x_n^* &= \pm\sqrt{\frac{n^2(n^2-1)}{(n^2+n-1)(n^2-n+1)}}, \\ f^* &= \frac{n^2(n^2+2n-2)}{(n^2+n-1)(n^2-n+1)}. \end{aligned}$$

   For example, for $n = 3$ variable $J = 2$ constraint CMMP$(3, 4, 0)$ problem, $x_1^* = \pm\sqrt{45/77}$, $x_2^* = 0$, and $x_3 = \pm\sqrt{72/77}$. Each of the four minimum points has a function value 117/77 or 1.519. Two constraints and the corresponding four minimum points are shown in Figure 24 for the three-variable problem. As the number of variables (n) increase, x_1^* approaches zero, x_n^* approaches one, and the optimal function value approaches one.

3. For three constraints ($J = 3$) leading to CMMP$(n, 8, 0)$, there are eight global minima:

$$\begin{aligned} x_1^* &= \pm\sqrt{\frac{n^2(2n^3-n^2+4n-8)}{n^6-2n^4+4n^3+7n^2-20n+8}}, \\ x_2^* &= \pm\sqrt{\frac{n^2(2n^3+n^2-6n+4)}{n^6-2n^4+4n^3+7n^2-20n+8}}, \\ x_i^* &= 0, \quad \text{for } i = 3, \ldots, (n-1), \\ x_n^* &= \pm\sqrt{\frac{n^2(n^4-2n^2-6n+4)}{n^6-2n^4+4n^3+7n^2-20n+8}}, \\ f^* &= \frac{n^3(n^3+4n^2-2n-8)}{n^6-2n^4+4n^3+7n^2-20n+8}. \end{aligned}$$

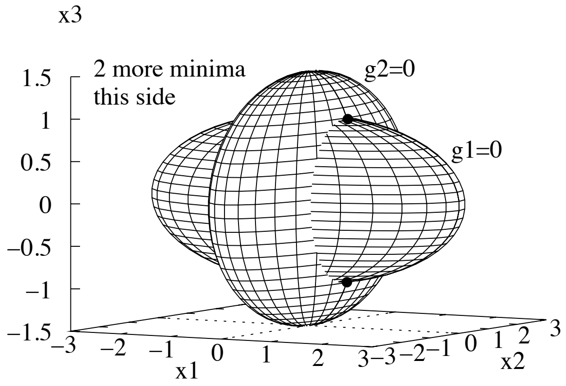


Figure 24: All 48 minimum points are found for the 16-variable MMP(16) problem.

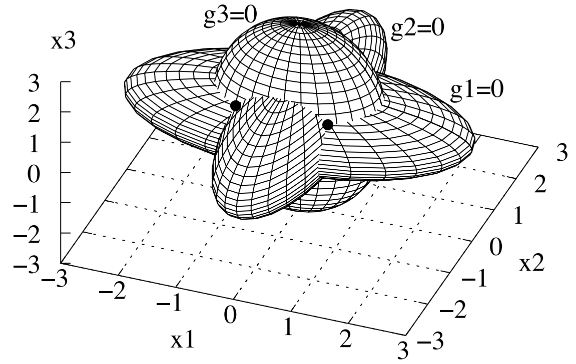


Figure 25: All 48 global minimum points are found for the modified MMP(16) problem.

Figure 25 shows three constraints and location of eight minimum points for the three-variable problem.

4. For $J = 4$ (CMMP$(n, 16, 0)$), there are 16 global minima:

$$\begin{aligned}
x_1^* &= \pm\sqrt{\frac{n^2(2n^4 - n^3 + 12n^2 - 24n - 24)}{n^7 - 3n^5 + 6n^4 + 34n^3 - 80n^2 - 60n + 96}},\\
x_2^* &= \pm\sqrt{\frac{n^2(2n^4 + n^3 - 2n^2 - 24)}{n^7 - 3n^5 + 6n^4 + 34n^3 - 80n^2 - 60n + 96}},\\
x_3^* &= \pm\sqrt{\frac{n^2(2n^4 + 3n^3 - 12n^2 - 4n + 24)}{n^7 - 3n^5 + 6n^4 + 34n^3 - 80n^2 - 60n + 96}},\\
x_i^* &= 0, \quad \text{for } i = 4, \ldots, (n-1),\\
x_n^* &= \pm\sqrt{\frac{n^2(n^5 - 3n^3 - 22n^2 + 4n + 24)}{n^7 - 3n^5 + 6n^4 + 34n^3 - 80n^2 - 60n + 96}},\\
f^* &= \frac{n^3(n^4 + 6n^3 - 24n - 24)}{n^7 - 3n^5 + 6n^4 + 34n^3 - 80n^2 - 60n + 96}.
\end{aligned}$$

For a 10-variable, 4-constraint problem (CMMP$(10, 16, 0)$), the minimum points are $(\pm 0.4514, \pm 0.4608, \pm 0.4718, 0, 0, 0, 0, 0, 0, \pm 0.9846)$ having a function value 1.6081.

5. Similarly, the minima for other $J (\geq 5)$ values can also be computed. Finally, for n constraints ($J = n$) leading to CMMP$(n, 2^n, 0)$ problem, there are 2^n global minima:

$$\begin{aligned}
x_i^* &= \pm\sqrt{\frac{6n}{(n+1)(2n+1)}}, \quad \text{for all } i = 1, 2, \ldots, n,\\
f^* &= \frac{6n^2}{(n+1)(2n+1)}.
\end{aligned}$$

For example, for a 5-variable, 5-constraint problem (CMMP$(5, 32, 0)$), $x_i^* = \pm 0.6742$ for $i = 1, \ldots, 5$, with a function value 2.2727.

Like in the case of two-variable problem described in the previous subsection, the objective function in a higher-dimensional problem can also be modified by shifting the zero of the function away from the origin to introduce different types of local minima.

6.3 Constrained Handling Procedure for Multimodal Optimization

To take care of feasible and infeasible solutions present in a population, we use the constraint handling strategy of the NSGA-II procedure (Deb, 2001; Deb et al., 2002). When two solutions are compared for domination, one of the following three scenarios can happen: (i) when one solution is feasible and other is infeasible, we choose the feasible solution, (ii) when two infeasible solutions are being compared, we simply choose the solution having smaller overall normalized constraint violation. and (iii) when both solutions are feasible, we follow the niching comparison method used before for unconstrained problems. Instead of using a procedure which is completely different from the unconstrained method, we use a simple modification to our proposed approach and importantly without the need of any additional constrained handling parameter. To get an accurate evaluation of f_2 for solutions close to a constraint boundary, we ignore any infeasible solution in the count of better neighboring solutions. The H-J exploratory search is used for evaluating the second objective function.

6.4 Constraint Handling Results

In this subsection, we construct a number of constrained multi-objective optimization problems using the above procedure and attempt to solve them using the proposed procedure. Identical NSGA-II parameter values as those used in the unconstrained cases are used here, except we increase the population size to $N = 25\max(n, M)$, due to a better handling of non-linear and non-convex constraints here. For all constrained problems, we use $\delta_f = 0.2$ and $\delta_x = 0.08$. In all cases, we perform 10 runs and show results from a typical run.

6.4.1 Problem CMMP(2,4,0)

Table 2 presents the four minimum points found by the proposed procedure. By comparing with Figure 21, it can be observed that the obtained points are identical to the four minimum points.

Table 2: Multimodal points found by constraint handling procedure for different problems.

Problem		Solution 1	Solution 2	Solution 3	Solution 4
CMMP(2,4,0)	$\mathbf{x}$	(0.894, 0.895)	(-0.895, 0.895)	(0.894, -0.896)	(-0.894, -0.897)
	f	1.601	1.601	1.602	1.603
CMMP(2,2,2)	$\mathbf{x}$	(0.894, 0.895)	(0.895, -0.894)	(0.894, -0.895)	(-0.894, -0.896)
	f	1.583	1.584	1.619	1.620
CMMP(2,1,3)	$\mathbf{x}$	(0.896, 0.894)	(-0.897, 0.894)	(0.894, -0.895)	(-0.894, -0.895)
	f	1.548	1.587	1.619	1.655

6.4.2 Problem CMMP(5,4,0)

Figure 26 presents four minimum points found by the proposed procedure for five-variable, two-constraint CMMP(5,4,0) problem. This figure allows an interesting way to visualize multiple solutions for more than two variables. The x-axis indicates the variable number (1 to 5, in this case) and the y-axis denotes the value of the variable. Thus, a polyline joining the variable values at different variable IDs represent one of the solutions found by the proposed procedure. It is interesting to note that the three intermediate variables (x_2 to x_4) take a value zero for all four solutions. Four solutions are formed from two values each from x_1 and x_5. The four solutions have following objective values: 1.355, 1.356, 1.355, and 1.356. Although theoretically all these values should be identical, the obtained solutions differ only in their third decimal places in their function values.

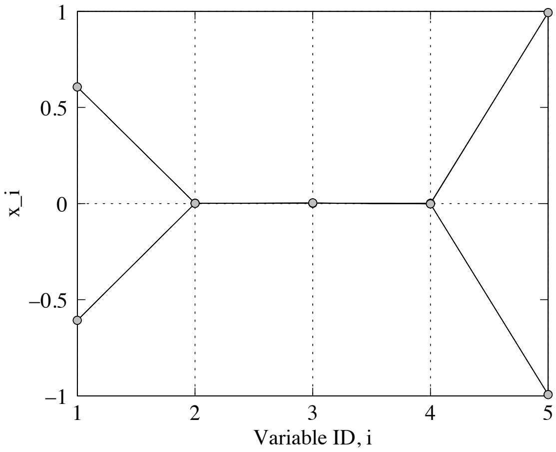


Figure 26: Four global minimum points for the five-variable, two-constraint CMMP(5,4,0) problem.

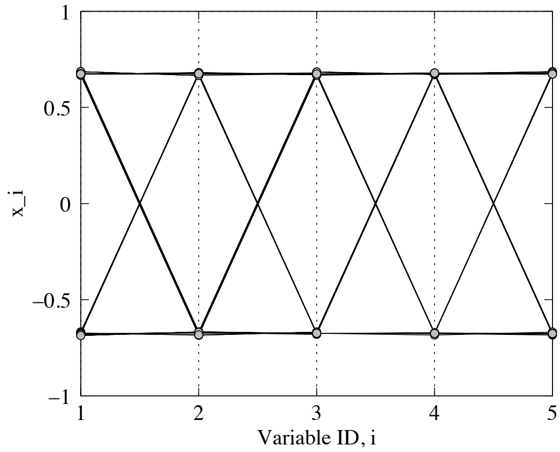


Figure 27: 32 global minimum points for the five-variable, five-constraint CMMP(5,32,0) problem.

6.4.3 Problem CMMP(5,32,0)

Next, we consider the five-variable, five-constraint problem, which has 32 minimum points having an identical objective value. Figure 27 represents all 32 solutions arising from two different values of each variable. The objective values vary in the range [2.277, 2.290], having a maximum of 0.57% difference in all 32 function values.

6.4.4 Problem CMMP(5,1,31)

Next, we include a problem having a combination of global and local constrained minima. The objective function in CMMP(5,32,0) is modified to construct CMMP(5,1,31), as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{5} (x_i - 0.05i)^2. \tag{11}$$

In this problem, we have only one global minimum and 31 other local minima, all lying on intersection on constraint boundaries. Figure 28 plots the function value of the obtained minimum points and the corresponding theoretical minimum function values. It is clear that out of the 32 minimum points, the proposed procedure is able to find best 31 of them (based on objective function value) and misses the local minimum point having the worst objective value. The figure represents a typical run obtained with 10 different initial populations. This study shows that despite the presence of multiple constrained local minimum points, the proposed procedure is able to locate all but one minima.

6.4.5 Problem CMMP(10,16,0)

Figure 29 shows the obtained solutions on the 10-variable, four-constraint CMMP(10,16,0) problem. The intermediate six variables (x_4 to x_9) take a value close to zero and 16 solutions come from each of two values of other four variables. The objective function values among 16 solutions vary only a maximum of 0.43% in the range [1.614, 1.621].

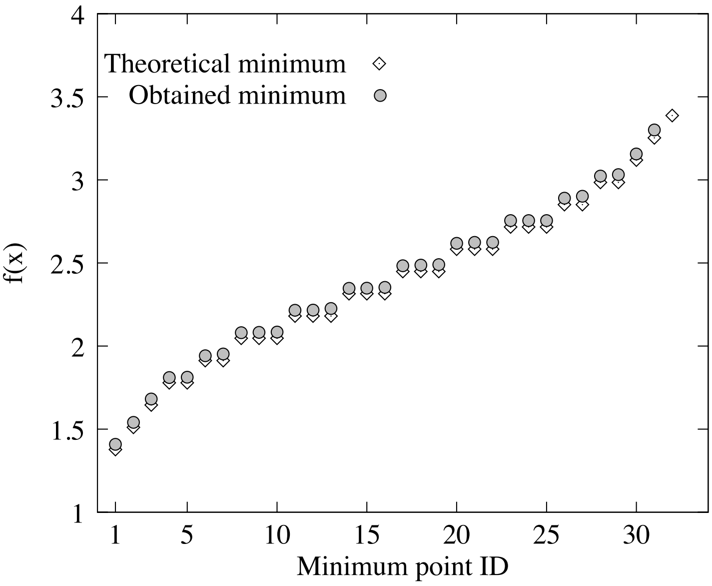


Figure 28: The CMMP(5,1,31) problem having one global and 31 local minimum points.

6.5 Repairing Procedure and Results

The H-J exploratory search around each population members helped us count the number of neighboring solutions that are better than the population members. However, if any better point is found by the exploratory search (final $\mathbf{x}^c$ is better than $\mathbf{x}$, as described in Section 5), the population member can be replaced with that point, as a repair operator (in the sense of Lamarckian repairing (Whitley et al., 1994; Orvosh and Davis, 1994)). Since this does not cost any extra computation, in general, this may be a better approach. In this

section, we show simulation results on 10-variable, 16-minima constrained problem. Figure 30 shows that the repair mechanism enables finding all 16 minimum points in 32 generations lesser than that needed without the repair mechanism. All parameters are kept identical between the two runs, except that in the repair case we increase the population size from 240 to 500 (to introduce diversity lost due to greediness introduced by accepting better solutions). The result in the figure is a typical performance we have observed in multiple runs and in solving other problems.

The success of the repair mechanism reveals an important computational aspect. Although $2n$ neighboring solutions are needed for each population members to evaluate the second objective, the repair mechanism also facilitates in advancing the current population members closer to their respective minimum point. This helps in reducing the number of generations needed in arriving at multiple optima simultaneously, as evident from the above figure. This implies that a repair procedure is able to reduce the number of neighboring solution evaluation effectively, thereby constituting a parallel and computationally efficient optimization procedure. However, the requirement of an increase in population size is a concern and we are currently investigating ways to modify our algorithm to reduce the effect of added selection pressure caused by the Lamarckian repair approach. Nevertheless, this first result is motivating from the point of view of the reduced computational requirement.

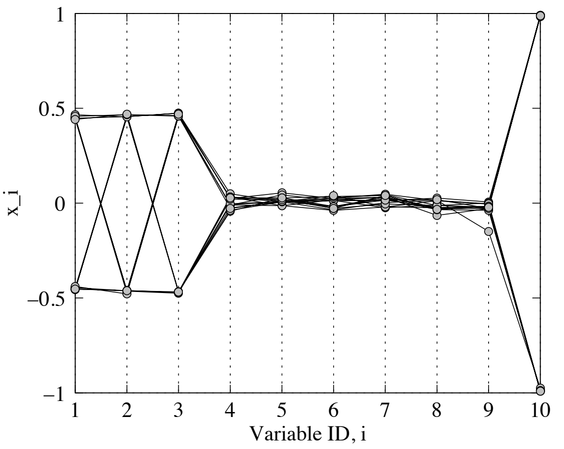


Figure 29: 16 global minimum points for the 10-variable, four-constraint CMMP(10,16,0) problem.

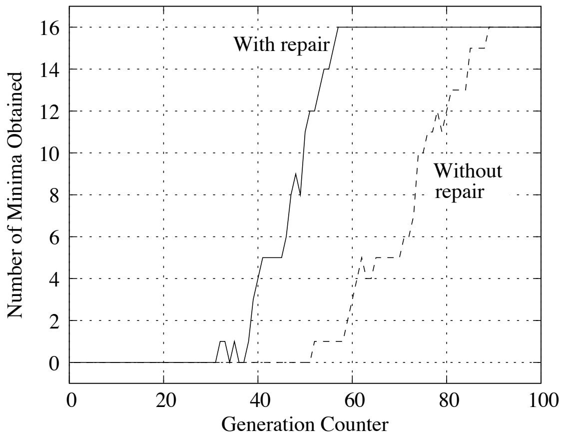


Figure 30: Inclusion of repair operation helps solve the CMMP(10,16,0) problem quicker.

7 Conclusions

In this paper, we have suggested a bi-objective formulation of a multimodal optimization problem so that multiple global and local optimal solutions become the only candidate weak Pareto-optimal set. While the objective function of the multimodal optimization problem is one of the objectives, a number of suggestions of the second objective function have been made here. Starting with the gradient-based approaches (demonstrating the foundation of the bi-objective approach), more pragmatic neighborhood count based approaches have been systematically developed for this purpose. Modifications to an existing EMO procedure are made in order to find weak non-dominated solutions. Using the Hooke-Jeeves based neighborhood count method for the second objective, the proposed EMO procedure has been able to solve 16-variable, 48-optima problems having a combination of global and local optima.

Another hallmark of this study is the suggestion of a multimodal constrained test-problem generator which is scalable in terms of the number of variables, the number of constraints, and the number of optima. The unconstrained multimodal optimization procedure developed in this study has been adapted to solve 10-variable, 16-optima constrained problems with success. The existing multimodal optimization literature avoided a systematic study of constrained problems. The construction procedure of the test-problem generator and the development of the proposed algorithm should motivate researchers to pursue constrained multimodal studies further.

This study explores a few forms of an additional objective, which, in conjunction with the original objective function of the multimodal problem, uniquely maps the optimal solutions on the weak Pareto-optimal front. The idea is interesting and other more efficient forms of the additional objective function can be designed and explored.

The modified domination definition used in this study requires two niching parameters – one for the objective space and another for the decision variable space. In another algorithm (Deb and Tiwari, 2008), we have suggested an agglomerate parameter by combining the two distance measures with an appropriate normalization procedure. Similar ideas can be employed here as well to reduce the number of parameters. In fact, these parameters can be adaptively sized depending on the population diversity and the desired number of optima.

The Hooke-Jeeves neighborhood count procedure requires $2n$ evaluations around each solution. But the best neighboring point around a solution can be used to replace the current solution in a repair operator, which we briefly presented in Section 6.5. Our initial results on the 10-variable constrained problem show that this Lamarckian repair approach can result in a faster discovery of multiple optima. Despite the use of $2n$ solution evaluations for each population member, lesser number of generations are needed, thereby making the procedure computationally efficient. Thus, in effect, on an average, less than $2n$ neighboring solutions are evaluated for each population member. This requires further investigation along the recent findings on repair based EA studies. Instead of repairing every population member, it can be performed probabilistically to reduce the *greediness* of the procedure. Further reduction in computational complexity may be possible by stopping the Hooke-Jeeves's exploratory search when it fails to find an adequate number of better neighboring solutions within a fraction of the total $2n$ allocations.

No mating restriction scheme is employed here. Like in the existing multimodal optimization studies, a mating restriction between similar solutions (in the variable space) may make the search more efficient and a study in this direction is worth performing.

In summary, we have suggested and demonstrated a novel bi-objective optimization procedure for finding multiple optimal solutions in solving multimodal problems. Niching tasks in both variable and objective spaces are introduced through redefining the domination principle so that an evolutionary multi-objective optimization procedure can be used to find multiple optima reliably to problems having a reasonably large dimension and having a large number of optima. This study has also introduced a multimodal constrained test-problem generator and, to our knowledge for the first time, has made a systematic study in finding multiple constrained optimum solutions. The multimodal solution methodology described here is different from the usual approach and borrows ideas from a fast-growing field in evolutionary computation. Besides the scalable and successful results on unconstrained and constrained problems, the study has also raised a number interesting issues which must be pursued in the near future.

Acknowledgments

Authors wish to thank Prof. Juergen Branke for sharing some initial thoughts with the first author during his Bessel research award visit to University of Karlsruhe in 2003.

References

Barrera, J. and Coello, C. A. C. (2009). *A Review of Particle Swarm Optimization Methods Used for Multimodal Optimization*, pages 9–37. Berlin: Springer.

Beasley, D., Bull, D. R., and Martin, R. R. (1993). A sequential niche technique for multimodal function optimization. *Evolutionary Computation*, 1(2):101–125.

Bessaou, M., Petrowski, A., and Siarry, P. (2000). Island model cooperating with speciation for multimodal optimization. In *PPSN VI: Proceedings of the Sixth International Conference on Parallel Problem Solving from Nature (PPSN-6)*, pages 437–446. Springer-Verlag.

Cavicchio, D. J. (1970). *Adaptive Search Using Simulated Evolution.* PhD thesis, Ann Arbor, MI: University of Michigan.

Darwen, P. J. and Yao, X. (1995). A dilemma for fitness sharing with a scaling function. In *Proceedings of the Second IEEE International Conference on Evolutionary Computation*, pages 166–171. IEEE Press.

Deb, K. (1989). Genetic algorithms in multi-modal function optimization. Master's thesis, Tuscaloosa, AL: University of Alabama.

Deb, K. (1995). *Optimization for Engineering Design: Algorithms and Examples.* New Delhi: Prentice-Hall.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms.* Wiley, Chichester, UK.

Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

Deb, K. and Goldberg, D. E. (1989). An investigation of niche and species formation in genetic function optimization. In *Proceedings of the Third International Conference on Genetic Algorithms*, pages 42–50.

Deb, K., Horn, J., and Goldberg, D. E. (1993). Multi-modal deceptive functions. *Complex Systems*, 7(2):131–153.

Deb, K. and Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multi-modal and multi-objective problems. *Complex Systems*, 9(6):431–454.

Deb, K. and Srinivasan, A. (2006). 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).

Deb, K. and Tiwari, S. (2008). Omni-optimizer: A generic evolutionary algorithm for global optimization. *European Journal of Operations Research (EJOR)*, 185(3):1062–1087.

DeJong, K. A. (1975). *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, Ann Arbor, MI: University of Michigan. Dissertation Abstracts International 36(10), 5140B (University Microfilms No. 76-9381).

Goldberg, D. E. (1989). *Genetic Algorithms for Search, Optimization, and Machine Learning.* Reading, MA: Addison-Wesley.

Goldberg, D. E. and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*, pages 41–49.

Hendershot, Z. (2004). A differential evolution algorithm for automatically discovering multiple global optima in multidimensional, discontinuous spaces. In *Proceedings of Fifteenth Midwest Artificial Intelligence and Cognitive Sciences Conference*, pages 92–97.

Im, C., Kim, H., Jung, H., and Choi, K. (2004). A novel algorithm for multimodal function optimization based on evolution strategy. *IEEE Transactions on Magnetics*, 40(2).

K. Parsapoulos, K. and Vrathis, M. (2004). On the computation of all minimizers through particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8(3).

Kennedy, J. (2000). Stereotyping: Improving particle swarm performance with cluster analysis. In *Proceedings of the Congress on Evolutionary Computation (CEC-2000)*, pages 1507–1512.

Knowles, J. D., Corne, D. W., and Deb, K., editors (2008). *Multiobjective problem solving from nature.* Springer Natural Computing Series, Springer-Verlag.

Lee, C., Cho, D., and Jung, H. (1999). Niching genetic algorithm with restricted competition selection for multimodal function optimization. *IEEE Transactions on Magnetics*, 35(3).

Li, J.-P., Parks, M. E. B. G. T., and Clarkson, P. J. (2002). A species conserving genetic algorithm for multimodal function optimization. *Evolutionary Computation*, 10(3):207–234.

Li, X. (2005). Efficient differential evolution using speciation for multimodal function optimization. In *Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO-2005)*, pages 873–880.

Mahfoud, S. W. (1995). A comparison of parallel and sequential niching methods. In *Proceedings of 6th International Conference on Genetic Algorithms*, pages 136–143.

Mengsheol, O. and Goldberg, D. E. (1999). Probabilistic crowding: deterministic crowding with probabilistic replacement. In *Proceedings of genetic and Evolutionary computation conference (GECCO-1999)*, pages 409–416.

Orvosh, D. and Davis, L. (1994). Using a genetic algorithm to optimize problems with feasibility constraints. In *Proceedings of the first IEEE conference on evolutionary computation*, volume 2, pages 548–553. New York, WA, USA: IEEE Service Center.

Parrott, D. and Li, X. (2006). Locating and tracking multiple dynamic optima by a particle swarm model using speciation. *IEEE Transactions on Evolutionary Computation*, 10(4):440–458.

Pétrowski, A. (1996). A clearing procedure as a niching method for genetic algorithms. In *IEEE 3rd International Conference on Evolutionary Computation (ICEC'96)*, pages 798–803.

Reklaitis, G. V., Ravindran, A., and Ragsdell, K. M. (1983). *Engineering Optimization Methods and Applications*. New York : Wiley.

Rönkkönen, J. (2009). Continuous Multimodal Global Optimization with Dierential Evolution-Based Methods. Unpublished.

Rönkkönen, J. and Lampinen, J. (2007). On determining multiple global optima by differential evolution. In *Proceedings of Evolutionary and Deterministic Methods for Design, optimization and Control (Eurogen 2007)*, pages 146–151.

Rönkkönen, J., Li, X., Kyrki, V., and Lampinen, J. (2008). A generator for multimodal test functions with multiple global optima. In *Proceedings of the Seventh International Conference on Simulated Evolution And Learning (SEAL-08)*, pages 239–248.

Shir, O. and Bäck, T. (2005). Niching in evolution strategies. In *Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO-2005)*, pages 915–916.

Shir, O. and Bäck, T. (2006). Niche radius adaptation in the CMA-ES niching algorithm. In *Parallel Problem Solving from Nature (PPSN-9)*, pages 142–151.

Singh, G. and Deb, K. (2006). Comparison of multi-modal optimization algorithms based on evolutionary algorithms. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation, July*, pages 08–12.

Streichert, F., Stein, G., Ulmer, H., and Zell, A. (2003). A clustering based niching ea for multimodal search spaces. In *Proceedings of the International Conference Evolution Artificielle*, pages 293–304. LNCS 2936.

Whitley, D., Gordon, V. S., and Mathias, K. (1994). Lamarckian evolution, the baldwin effect and function optimization. In *Parallel Problem Solving from Nature (PPSN-III)*, pages 6–15. Springer-Verlag.

Yin, X. and Germay, N. (1993). A fast genetic algorithm with sharing scheme using clustering analysis methods in multimodal function optimization. In *Proceedings of International Conference on Artificial Neural Networks and Genetic Algorithms*, pages 450–457.

Zhang, J. and Li, K. (2005). A sequential niching technique for particle swarm optimization. In *Proceedings of 2005 international conference on intelligent computing (ICIC-2005)*, pages 390–399.