one such a circle. The points marked on each line can be joined with straight lines to form a polygon. Thus, for N solutions, there will be N such circles. Visually, they will convey the convergence and diversity in obtained solutions.

8.1.5 Visual Method

This method is particularly suitable for design related problems. The obtained non-dominated solutions can be shown side-by-side with a tag of objective function values. In this way, a user can evaluate and compare different trade-off solutions. This will also allow the user to have a feel of how solutions change when a particular objective function value is changed. Figure 180 shows nine non-dominated solutions obtained for the shape design of a simply supported beam with a central-point loading (described below in Section 9.6). The figure shows solutions for two objectives: the weight of the beam and the maximum deflection of the beam. Starting from a low-weight beam, the obtained set of solutions shows how beams become more stiffened, incurring less deflection but with more weight (or cost) of the beam.

If possible, a pseudo-weight vector can also be given along with each solution. Since the true minimum and maximum objective function values of the Pareto-optimal set are usually not known, the pseudo-weights can be derived as follows:

$$w_i = \frac{f_i^{\max} - f_i}{f_i^{\max} - f_i^{\min}} \bigg/ \sum_{j=1}^{M} \frac{f_j^{\max} - f_j}{f_j^{\max} - f_j^{\min}}, \tag{8.1}$$

where $f_i^{\min}$ and $f_i^{\max}$ denote the minimum and maximum values of the i-th objective function among the obtained solutions (or Pareto-optimal solutions, if known). This weight vector for each solution provides a relative importance factor for each objective corresponding to the solution. The figure also shows the pseudo-weights calculated for all nine solutions.

Besides the obvious visual appearance, this method has another advantage. The decision variables and the corresponding objective function values (in terms of a pseudo-weight vector) are all shown in one set of plots. This provides a user with a plethora of information, which would hopefully make the decision-making easier.

8.2 Performance Metrics

When a new and innovative methodology is initially discovered for solving a search and optimization problem, a visual description is adequate to demonstrate the working of the proposed methodology. In such pioneering studies, it is important to establish, in the mind of the reader, a picture of the new suggested procedure. However, when the methodology becomes popular and a number of different implementations exist, it becomes necessary to compare these in terms of their performance on various test problems. This has been a common trend in the development of many successful solution methodologies, including multi-objective evolutionary algorithms.

Most earlier MOEAs demonstrated their working by showing the obtained non-

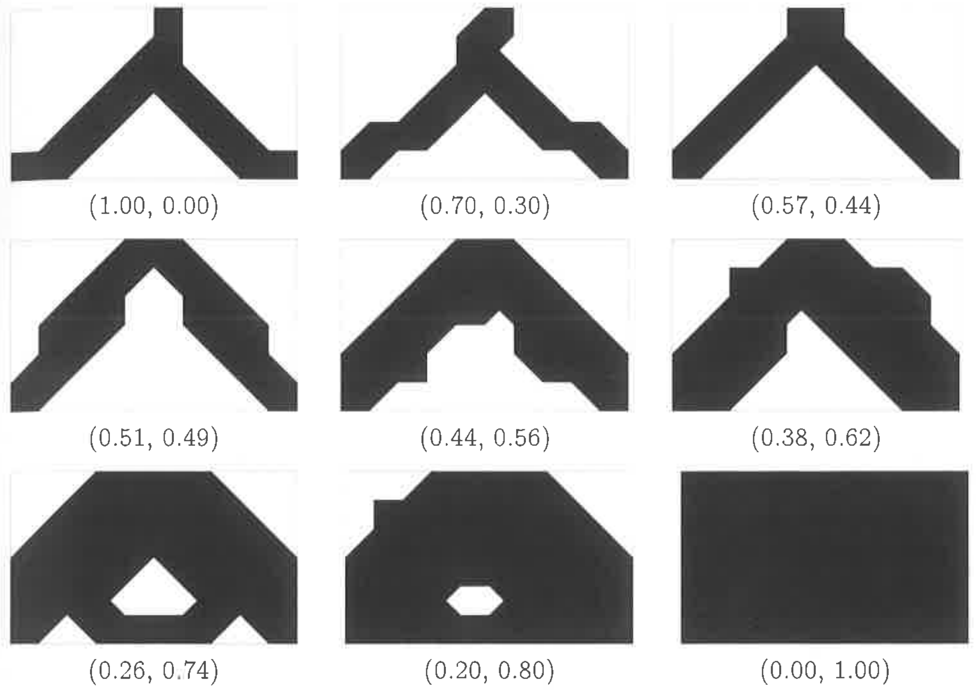

Figure 180 Nine non-dominated solutions obtained using the NSGA-II for a simply supported beam design problem (see page 475). Starting from the top and moving towards the right, solutions are presented as increasing magnitude of weight (one of the objectives). This is a reprint of Figure 13 from Deb and Goel (2001b) (© Springer-Verlag Berlin Heidelberg 2001).

dominated solutions along with the true Pareto-optimal solutions in the objective space. In those studies, the emphasis has been given to demonstrate how closely the obtained solutions have converged to the true Pareto-optimal front. With the existence of many different MOEAs, it is necessary that their performances be quantified on a number of test problems. Before we discuss the performance metrics used in MOEA studies, we would like to highlight that similar to the importance of a choice of performance metric, there is a need to choose appropriate test problems for a comparative study. Test problems might be known for their nature of difficulties, the extent of difficulties and the exact location of the Pareto-optimal solutions (both in the decision variable and in the objective space). We shall discuss this important issue of test problem development for multi-objective optimization in Section 8.3 below.

It is amply mentioned in this book that there are two distinct goals in multi-objective optimization: (i) discover solutions as close to the Pareto-optimal solutions as possible, and (ii) find solutions as diverse as possible in the obtained non-dominated front. In some sense, these two goals are *orthogonal* to each other. The first goal requires a search *towards* the Pareto-optimal region, while the second goal requires a search *along* the Pareto-optimal front, as depicted in Figure 181. In this book, a

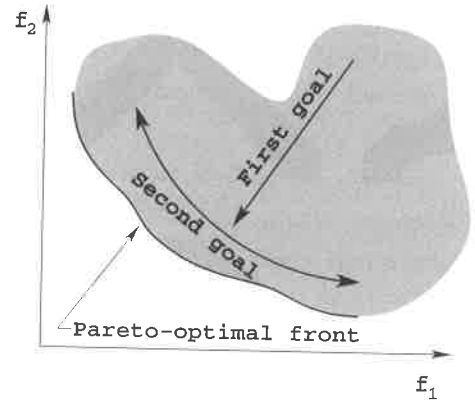

Figure 181 Two goals of multi-objective optimization.

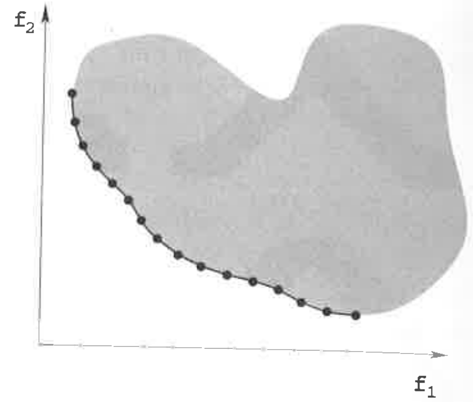

Figure 182 An ideal set of non-dominated solutions.

diverse set of solutions is meant to represent a set of solutions covering the entire Pareto-optimal region uniformly. The measure of diversity can also be separated in two different measures of *extent* (meaning the spread of extreme solutions) and *distribution* (meaning the relative distance among solutions) (Zitzler et al., 2000).

An MOEA will be termed a good MOEA, if both goals are satisfied adequately. Thus, with a good MOEA, a user is expected to find solutions close to the true Pareto-optimal front, as well as solutions that span the entire Pareto-optimal region uniformly. In Figure 182, we show the performance of an ideal MOEA on a hypothetical problem. It is clear that all of the obtained non-dominated solutions lie on the Pareto-optimal front and they also maintain a uniform-like distribution over the entire Pareto-optimal region. However, because of the different types of difficulties associated with a problem or the inherent inefficiencies associated with the chosen algorithm, such a well-converged and well-distributed non-dominated set of solutions may not always be found by an MOEA in solving any arbitrary problem. Take two sets of non-dominated solutions obtained by using two algorithms on an identical problem, as depicted in Figures 183 and 184. With the first algorithm (Algorithm 1), the obtained solutions converge fairly well on the Pareto-optimal front, but clearly there is lack of diversity among them. This algorithm has failed to provide information about the intermediate Pareto-optimal region. On the other hand, the latter algorithm (Algorithm 2) has obtained a good diverse set of solutions, but unfortunately the solutions are not close to the true Pareto-optimal front. Although this latter set of solutions can provide a rough idea of different trade-off solutions, the exact Pareto-optimal solutions are not discovered. With such sets of obtained solutions, it is difficult to conclude which set is better in an absolute sense.

Since convergence to the Pareto-optimal front and the maintenance of a diverse set of solutions are two distinct and somewhat conflicting goals of multi-objective optimization, no single metric can decide the performance of an algorithm in an absolute sense. Algorithm 1 fairs well with respect to the first task of multi-objective

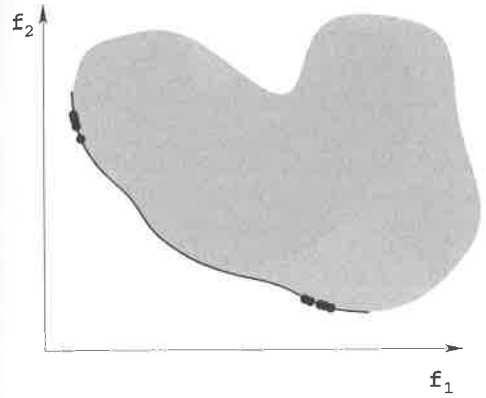

Figure 183 Convergence is good, but distribution is poor (by Algorithm 1).

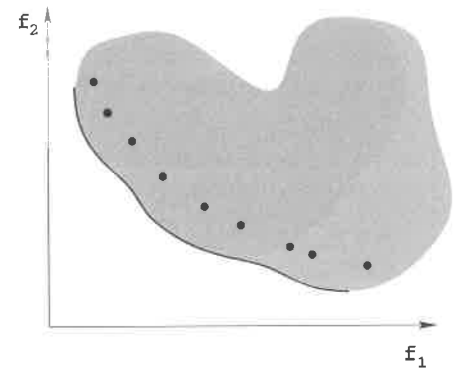

Figure 184 Convergence is poor, but distribution is good (by Algorithm 2).

optimization, while Algorithm 2 fairs well with respect to the second task. We are again faced here with a two-objective scenario. If we define a metric responsible for finding the closeness of the obtained set of solutions to the Pareto-optimal front and another metric responsible for finding the spread of the solutions, the performance of the above described two algorithms are *non-dominated* to each other. There is a clear need of having at least two performance metrics for adequately evaluating both goals of multi-objective optimization.

Besides the two extreme cases of an ideal convergence with a bad diversity (Algorithm 1 in the previous example) and a bad convergence with an ideal diversity (Algorithm 2), there could be other scenarios. Figure 185 shows that the non-dominated set of solutions obtained by using Algorithm A dominates the non-dominated set of solutions obtained by using Algorithm B. In this case, Algorithm A has clearly performed better than Algorithm B. However, there could be a more confusing scenario (Figure 186) where a part of the solutions obtained by using Algorithm A dominates a part of the solutions obtained by using Algorithm B, and vice versa. This scenario introduces a third dimension of difficulty in designing a performance metric for multi-objective optimization. In this case, both algorithms have similar convergence and diversity properties. The outcome of the comparison will largely depend on the exact definition of the metrics used for these measures.

Based on these discussions, we realize that while comparing two or more algorithms, at least two performance metrics (one evaluating the progress towards the Pareto-optimal front and the other evaluating the spread of solutions) need to be used and the exact definitions of the performance metrics are important. In the following three subsections, we will categorize some of the performance metrics commonly used in the literature. The first type discusses metrics that can be used to measure the progress towards the Pareto-optimal front explicitly. The second type discusses metrics that can be used to measure the diversity among the obtained solutions explicitly. The

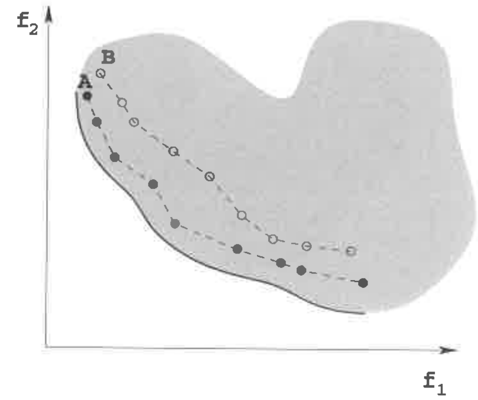

Figure 185 Algorithm A performs better than Algorithm B.

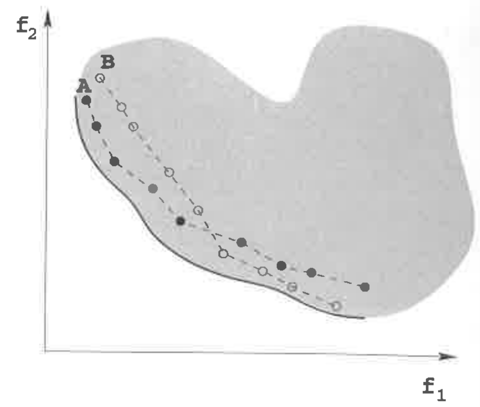

Figure 186 Algorithms A and B are difficult to compare.

third type uses two metrics which measure both goals of multi-objective optimization in an implicit manner.

8.2.1 Metrics Evaluating Closeness to the Pareto-Optimal Front

This metric explicitly computes a measure of the closeness of a set Q of N solutions from a known set of the Pareto-optimal set P^*. In some test problems, the set P^* may be known as a set of infinite solutions (for example, the case where an equation describing the Pareto-optimal relationship among the decision variables is known) or a set of finite solutions (only a few solutions are known or possible to compute). In order to find the proximity between two sets of different sizes, a number of metrics can be defined. The following metrics are already used for this purpose in different MOEA studies. They provide a good estimate of convergence if a large set for P^* is chosen.

Error Ratio

This metric (ER) simply counts the number of solutions of Q which are not members of the Pareto-optimal set P^* (Veldhuizen, 1999), or mathematically,

$$\mathrm{ER} = \frac{\sum_{i=1}^{|Q|} e_i}{|Q|}, \tag{8.2}$$

where $e_i = 1$ if $i \notin P^*$ and $e_i = 0$, otherwise. Figure 187 shows the Pareto-optimal set as filled circles and the obtained non-dominated set of solutions as open squares. In this case, the error ratio $\mathrm{ER} = 3/5 = 0.6$, since there are three solutions which are not members of the Pareto-optimal set. Equation (8.2) reveals that a smaller value of ER means a better convergence to the Pareto-optimal front. The metric ER takes a value between zero and one. An $\mathrm{ER} = 0$ means all solutions are members of the

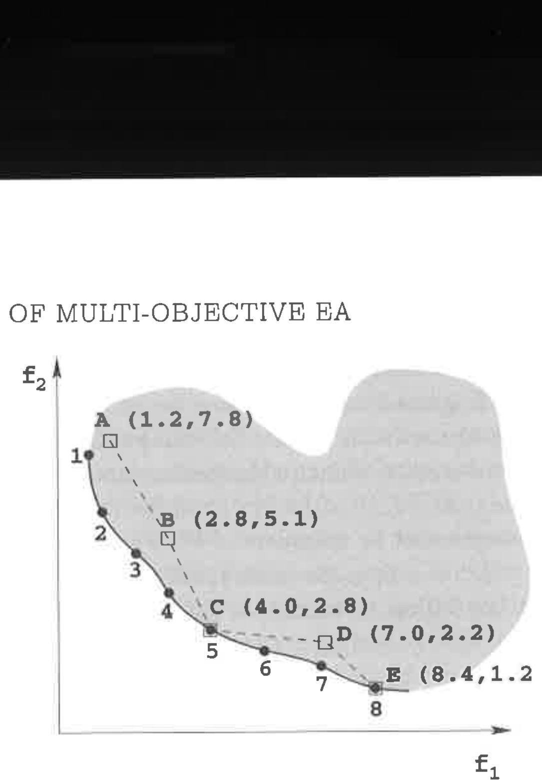

Figure 187 The set of non-dominated solutions Q are shown as open squares, while a set of the chosen Pareto-optimal set P^* is shown as filled circles.

Pareto-optimal front P^*, and an $\mathrm{ER} = 1$ means no solution is a member of the P^*.

It is worth mentioning here that although a member of Q is Pareto-optimal, if that solution does not exist in P^*, it may be counted in equation (8.2) as a non-Pareto-optimal solution. Thus, it is essential that a large set for P^* is used in the above equation. Another drawback of this metric is that if no member of Q is in the Pareto-optimal set, it does not distinguish the relative closeness of any set Q from P^*. Because of this discreteness in the values of ER, this metric is not popularly used. The metric can be made more useful, by redefining e_i as follows. For each solution $i \in Q$, if the minimum Euclidean distance (in the objective space) between i and P^* is larger than a threshold value δ, the parameter e_i is set to one. By using a suitable value for δ, this modified metric would then represent a measure of the proportion of the solutions close to the Pareto-optimal front.

Set Coverage Metric

A similar metric is suggested by Zitzler (1999). However, the metric can also be used to get an idea of the relative spread of solutions between two sets of solution vectors A and B. The set coverage metric $\mathcal{C}(A, B)$ calculates the proportion of solutions in B, which are weakly dominated by solutions of A:

$$\mathcal{C}(A, B) = \frac{|\{b \in B | \exists a \in A : a \preceq b\}|}{|B|}. \tag{8.3}$$

The metric value $\mathcal{C}(A, B) = 1$ means all members of B are weakly dominated by A. On the other hand, $\mathcal{C}(A, B) = 0$ means that no member of B is weakly dominated by A. Since the domination operator is not a symmetric operator (refer to Section 2.4.3), $\mathcal{C}(A, B)$ is not necessarily equal to $1 - \mathcal{C}(B, A)$. Thus, it is necessary to calculate both $\mathcal{C}(A, B)$ and $\mathcal{C}(B, A)$ to understand how many solutions of A are covered by B and vice versa. It is interesting to note that the cardinality of both vectors A and B need not be the same while using the above equation.

Although Zitzler (1999) used this metric for comparing the performance of two algorithms, it can also be used to evaluate the performance of an algorithm by using $A = P^*$ and $B = Q$. The metric $\mathcal{C}(P^*, Q)$ will determine the proportion of solutions in Q, which are weakly dominated by members of P^*. For the sets P^* and Q shown in Figure 187, $\mathcal{C}(P^*, Q) = 3/5 = 0.6$, since three solutions (A, B and D) are dominated by a member of P^*. It is needless to write that $\mathcal{C}(Q, P^*)$ is always zero.

Generational Distance

Instead of finding whether a solution of Q belongs to the set P^* or not, this metric finds an average distance of the solutions of Q from P^*, as follows (Veldhuizen, 1999):

$$GD = \frac{(\sum_{i=1}^{|Q|} d_i^p)^{1/p}}{|Q|}. \tag{8.4}$$

For $p = 2$, the parameter d_i is the Euclidean distance (in the objective space) between the solution $i \in Q$ and the *nearest* member of P^*:

$$d_i = \min_{k=1}^{|P^*|} \sqrt{\sum_{m=1}^{M} (f_m^{(i)} - f_m^{*(k)})^2}, \tag{8.5}$$

where $f_m^{*(k)}$ is the m-th objective function value of the k-th member of P^*. For the obtained solutions shown in Figure 187, solution A is closest to the Pareto-optimal solution 1, solution B is closest to solution 3, solution C is closest to solution 5, solution D is closest to solution 7 and solution E is closest to 8. If the Pareto-optimal solutions have the following objective function values:

Solution	f_1	f_2	Solution	f_1	f_2
1	1.0	7.5	5	4.0	2.8
2	1.1	5.5	6	5.5	2.5
3	2.0	5.0	7	6.8	2.0
4	3.0	4.0	8	8.4	1.2

the corresponding Euclidean distances are as follows:

$$\begin{aligned}
d_{A1} &= \sqrt{(1.2-1.0)^2 + (7.8-7.5)^2} = 0.36,\\
d_{B3} &= \sqrt{(2.8-2.0)^2 + (5.1-5.0)^2} = 0.81,\\
d_{C5} &= \sqrt{(4.0-4.0)^2 + (2.8-2.8)^2} = 0.00,\\
d_{D7} &= \sqrt{(7.0-6.8)^2 + (2.2-2.0)^2} = 0.28,\\
d_{E8} &= \sqrt{(8.4-8.4)^2 + (1.2-1.2)^2} = 0.00.
\end{aligned}$$

Thus, the generational distance (GD) calculated using equation (8.4) with $p = 2$ is $GD = 0.19$. Intuitively, an algorithm having a small value of GD is better.

The difficulty with the above metric is that if there exists a Q for which there is a large fluctuation in the distance values, the metric may not reveal the true distance. In such an event, the calculation of the variance of the metric GD is necessary. Furthermore, if the objective function values are of differing magnitude, they should be normalized before calculating the distance measure. In order to make the distance calculations reliable, a large number of solutions in the P^* set is recommended.

Because of its simplicity and average characteristics (with $p = 1$), other researchers have also suggested (Zitzler, 1999) and used this metric (Deb et al., 2000a). The latter investigators have used this metric in a recent comparative study. For each computation of this metric (they called it Υ), the standard deviations of Υ among multiple runs are also reported. If a small value of the standard deviation is observed, the calculated Υ can be accepted with confidence.

Maximum Pareto-Optimal Front Error

This metric (MFE) computes the worst distance d_i among all members of Q (Veldhuizen, 1999). For the example problem shown in Figure 187, the worst distance is caused by the solution B (referring to the calculations above). Thus, $MFE = 0.81$. This measure is a conservative measure of convergence and may provide incorrect information about the distribution of solutions. In this connection, a 'χ' percentile (where $\chi = 25$ or 50) of the distances d_i among all solutions of Q can be used as a metric. We will discuss more about such percentile measures later in this section.

8.2.2 Metrics Evaluating Diversity Among Non-Dominated Solutions

There also exists a number of metrics to find the diversity among obtained non-dominated solutions. In the following, we describe a few of them.

Spacing

Schott (1995) suggested a metric which is calculated with a relative distance measure between consecutive solutions in the obtained non-dominated set, as follows:

$$S = \sqrt{\frac{1}{|Q|} \sum_{i=1}^{|Q|} (d_i - \overline{d})^2}, \tag{8.6}$$

where $d_i = \min_{k \in Q \wedge k \neq i} \sum_{m=1}^{M} |f_m^i - f_m^k|$ and $\overline{d}$ is the mean value of the above distance measure $\overline{d} = \sum_{i=1}^{|Q|} d_i/|Q|$. The distance measure is the minimum value of the sum of the absolute difference in objective function values between the i-th solution and any other solution in the obtained non-dominated set. Notice that this distance measure is different from the minimum Euclidean distance between two solutions.

The above metric measures the standard deviations of different d_i values. When the solutions are near uniformly spaced, the corresponding distance measure will be small. Thus, an algorithm finding a set of non-dominated solutions having a smaller spacing (S) is better.

For the example problem presented in Figure 187, we show the calculation procedure for d_A:

$$\begin{aligned} d_A &= \min\left((1.6+2.7),(2.8+5.0),(5.8+5.6),(7.2+6.6)\right) \\ &= 4.3. \end{aligned}$$

Similarly, $d_B = 3.5$, $d_C = 3.5$, $d_D = 2.4$ and $d_E = 2.4$. Figure 187 shows that the solutions A to E are almost uniformly spaced in the objective space. Thus, the standard deviation in the corresponding d_i values would be large. We observe that $\bar{d} = 3.22$ and the metric $S = 0.73$. For a set of non-dominated solutions which are randomly placed in the objective space, the standard deviation measure would be much small.

Conceptually, the above metric provides useful information about the spread of the obtained non-dominated solutions. However, if proper bookkeeping is not used, the implementational complexity is $O(|Q|^2)$. This is because for each solution i, all other solutions must be checked for finding the minimum distance d_i. Although by using the symmetry in distance measures, half the calculations can be avoided, the complexity is still quadratic to the number of obtained non-dominated solutions. Deb et al. (2000a) suggested calculating d_i between consecutive solutions in each objective function independently. The procedure is as follows.

First, sort the obtained non-dominated front in ascending order of magnitude in each objective function. Now, for each solution, sum the difference in objective function values between two nearest neighbors in each objective. For a detailed procedure, refer to Section 6.2 above. Since the sorting has a complexity $O(|Q|\log|Q|)$, this distance metric would be quicker to compute than the above distance metric. Since different objective functions are added together, normalizing the objectives before using equation (8.6) is essential. Moreover, the above metric does not take into account the extent of spread. As long as the spread is uniform within the range of obtained solutions, the metric S produces a small value. The following metric takes care of the extent of spread in the obtained solutions.

Spread

Deb et al. (2000a) suggested the following metric to alleviate the above difficulty:

$$\Delta = \frac{\sum_{m=1}^{M} d_m^e + \sum_{i=1}^{|Q|} |d_i - \bar{d}|}{\sum_{m=1}^{M} d_m^e + |Q|\bar{d}}, \tag{8.7}$$

where d_i can be any distance measure between neighboring solutions and $\bar{d}$ is the mean value of these distance measures. The Euclidean distance, the sum of the absolute differences in objective values or the crowding distance (defined earlier on page 248) can be used to calculate d_i. The parameter d_m^e is the distance between the extreme solutions of P^* and Q corresponding to m-th objective function.

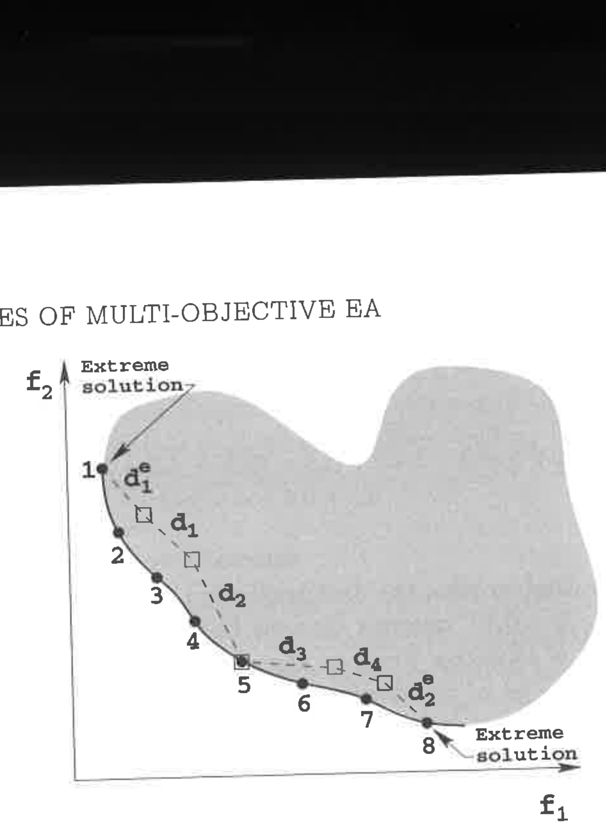

Figure 188 Distances from the extreme solutions.

For a two-objective problem, the corresponding d_1^e and d_2^e are shown in Figure 188 and d_i can be taken as the consecutive Euclidean distance between i-th and $(i+1)$-th solutions. Thus, the term $|Q|$ in the above equation may be replaced by the term $(|Q|-1)$. The metric takes a value zero for an ideal distribution, only when $d_j^e = 0$ and all d_i values are identical to their mean $\bar{d}$. The first condition means that in the obtained non-dominated set of solutions, the true extreme Pareto-optimal solutions exist. The second condition means that the distribution of intermediate solutions is uniform. Such a set is an ideal outcome of any multi-objective EA. Thus, for an ideal distribution of solutions, $\Delta = 0$. Consider another case, where the distribution of the obtained non-dominated solutions is uniform but they are clustered in one place. Such a distribution will make all $|d_i - \bar{d}|$ values zero, but will cause non-zero values for d_m^e. The corresponding Δ becomes $\sum_{m=1}^{M} d_m^e / (\sum_{m=1}^{M} d_m^e + (|Q|-1)\bar{d})$. This quantity lies within $[0,1]$. Since the denominator measures the length of the piece-wise approximation of the Pareto-front, the Δ value increases with d_m^e. Thus, as the solutions get clustered more and more closer from the ideal distribution, the Δ value increases from zero towards one. For a non-uniform distribution of the non-dominated solutions, the second term in the numerator is not zero and, in turn, makes the Δ value more than that with a non-uniform distribution. Thus, for bad distributions, the Δ values can be more than one as well.

For the example problem in Figure 187, the extreme right Pareto-optimal solution (solution 8) is the same as the extreme non-dominated solution (solution E). Thus, $d_2^e = 0$ here. Since solution 1 (the extreme left Pareto-optimal solution) is not found, we calculate $d_1^e = 0.5$ (using Schott's difference distance measure). We calculate the d_i values accordingly:

$$d_1 = 4.3, \quad d_2 = 3.5,$$
$$d_3 = 3.5, \quad d_4 = 2.4.$$

The average of these values is $\overline{d} = 3.43$. We can now use these values to calculate the spread metric:

$$\begin{aligned}\Delta &= \frac{0.5 + 0.0 + |4.3 - 3.43| + |3.5 - 3.43| + |3.5 - 3.43| + |2.4 - 3.43|}{0.5 + 0.0 + 4 \times 3.43},\\ &= 0.18.\end{aligned}$$

Since this value is close to zero, the distribution is not bad. If solution A were the same as solution 1, $\Delta = 0.15$, meaning that the distribution would have been better than the current set of solutions. Thus, an algorithm finding a smaller Δ value is able to find a better diverse set of non-dominated solutions.

Maximum Spread

Zitzler (1999) defined a metric measuring the length of the diagonal of a hyperbox formed by the extreme function values observed in the non-dominated set:

$$D = \sqrt{\sum_{m=1}^{M} \left(\max_{i=1}^{|Q|} f_m^i - \min_{i=1}^{|Q|} f_m^i \right)^2}. \tag{8.8}$$

For two-objective problems, this metric refers to the Euclidean distance between the two extreme solutions in the objective space, as shown in Figure 189.

In order to have a normalized version of the above metric, it can be modified as follows:

$$\overline{D} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(\frac{\max_{i=1}^{|Q|} f_m^i - \min_{i=1}^{|Q|} f_m^i}{F_m^{\max} - F_m^{\min}} \right)^2}. \tag{8.9}$$

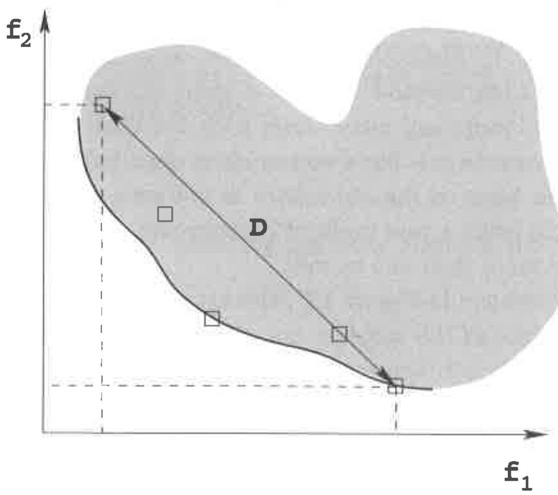

Figure 189 The maximum spread does not reveal true distribution of solutions.

Here, $F_m^{\max}$ and $F_m^{\min}$ are the maximum and minimum value of the m-th objective in the chosen set of Pareto-optimal solutions, P^*. In this way, if the above metric is one, a widely spread set of solutions is obtained. However, neither $\overline{D}$ nor D can evaluate the exact distribution of intermediate solutions.

Chi-Square-Like Deviation Measure

We proposed this metric for multi-modal function optimization and later used it to evaluate the distributing ability of multi-objective optimization algorithms (Deb, 1989). A brief description of this measure is also presented on page 158. In this metric, a neighborhood parameter ϵ is used to count the number of solutions, n_i, within each chosen Pareto-optimal solution (solution $i \in P^*$). The distance calculation can be made in either the objective space or in the decision variable space. The deviation between this counted set of numbers with an ideal set is measured in the chi-square sense:

$$\iota = \sqrt{\sum_{i=1}^{|P^*|+1} \left(\frac{n_i - \overline{n}_i}{\sigma_i} \right)^2}, \tag{8.10}$$

Since there is no preference to any particular Pareto-optimal solution, it is customary to choose a uniform distribution as an ideal distribution. This means that there should be $\overline{n}_i = |Q|/|P^*|$ solutions allocated in the niche of each chosen Pareto-optimal solution. The parameter $\sigma_i^2 = \overline{n}_i(1 - \overline{n}_i/|Q|)$ is suggested for $i = 1, 2, \ldots, |P^*|$. However, the index $i = |P^*|+1$ represents all solutions which do not reside in the ϵ-neighborhood of any of the chosen Pareto-optimal solutions. For this index, the ideal number of solutions and its variance are calculated as follows:

$$\overline{n}_{|P^*|+1} = 0, \quad \sigma^2_{|P^*|+1} = \sum_{i=1}^{|P^*|} \sigma_i^2 = |Q| \left(1 - \frac{1}{|P^*|} \right).$$

Let us illustrate the calculation procedure for the scenario depicted in Figure 190. There are five chosen Pareto-optimal solutions, marked as 1 to 5. The obtained non-dominated set of $|Q| = 10$ solutions are marked as squares. Thus, the expected number of solutions near each of the five Pareto-optimal solutions ($|P^*| = 5$) is $\overline{n} = 10/5$ or 2 and $\overline{n}_6 = 0$ (the expected number of solutions away from these five Pareto-optimal solutions). The corresponding standard deviations can be calculated by using these numbers: $\sigma_i^2 = 1.6$ for $i = 1$ to 5 and $\sigma_6^2 = 5 \times 1.6 = 8.0$. Now, we count the actual number of non-dominated solutions present near each Pareto-optimal solution (the neighborhood is shown by dashed lines). We observe from the figure that

$$n_1 = 1, \quad n_2 = 3, \quad n_3 = 1, \quad n_4 = 2, \quad n_5 = 1, \quad n_6 = 2.$$

With these values, the metric is:

$$\begin{aligned}\iota &= \sqrt{\left(\frac{1-2}{\sqrt{1.6}}\right)^2 + \left(\frac{3-2}{\sqrt{1.6}}\right)^2 + \left(\frac{1-2}{\sqrt{1.6}}\right)^2 + \left(\frac{2-2}{\sqrt{1.6}}\right)^2 + \left(\frac{1-2}{\sqrt{1.6}}\right)^2 + \left(\frac{2-0}{\sqrt{8.0}}\right)^2},\\ &= 1.73.\end{aligned}$$

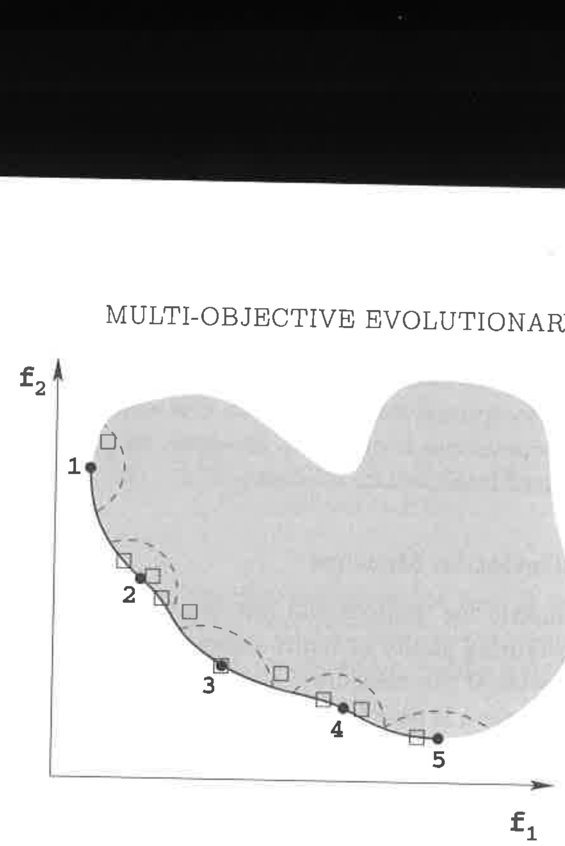

Figure 190 Non-dominated solutions in each niche of the five Pareto-optimal solutions.

Since the above metric is a deviation measure, an algorithm finding a smaller value of this metric is better able to distribute its solutions near the Pareto-optimal region. If the ideal distribution is found, this metric will have a value zero. If most solutions are away from the Pareto-optimal solutions, this metric cannot evaluate the spread of solutions adequately.

8.2.3 Metrics Evaluating Closeness and Diversity

There exist some metrics where both tasks have been evaluated in a combined sense. Such a metric can only provide a qualitative measure of convergence as well as diversity. Nevertheless, they can be used along with one of the above metrics to get a better overall evaluation.

Hypervolume

This metric calculates the volume (in the objective space) covered by members of Q (the region shown hatched in Figure 191) for problems where all objectives are to be minimized (Veldhuizen, 1999; Zitzler and Thiele, 1998b). Mathematically, for each solution $i \in Q$, a hypercube v_i is constructed with a reference point W and the solution i as the diagonal corners of the hypercube. The reference point can simply be found by constructing a vector of worst objective function values. Thereafter, a union of all hypercubes is found and its hypervolume (HV) is calculated:

$$\mathrm{HV} = \text{volume}\left(\cup_{i=1}^{|Q|} v_i\right). \tag{8.11}$$

Figure 191 shows the chosen reference point W. The hypervolume is shown as a hatched region. Obviously, an algorithm with a large value of HV is desirable. For

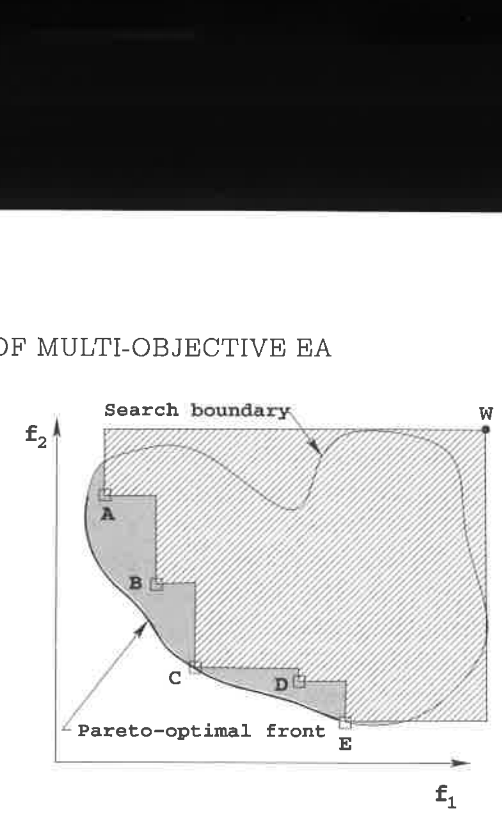

Figure 191 The hypervolume enclosed by the non-dominated solutions.

the example problem shown in Figure 187, the hypervolume HV is calculated with $W = (11.0, 10.0)^T$ as:

$$\begin{aligned}\mathrm{HV} &= (11.0-8.4)\times(10.0-1.2)+(8.4-7.0)\times(10.0-2.2)\\ &\quad +(7.0-4.0)\times(10.0-2.8)+(4.0-2.8)\times(10.0-5.1)\\ &\quad +(2.8-1.2)\times(10.0-7.8)\\ &= 64.80.\end{aligned}$$

This metric is not free from arbitrary scaling of objectives. For example, if the first objective function takes values an order of magnitude more than that of the second objective, a unit improvement in f_1 would reduce HV much more than that a unit improvement in f_2. Thus, this metric will favor a set Q which has a better converged solution set for the least-scaled objective function. To eliminate this difficulty, the above metric can be evaluated by using normalized objective function values.

Another way to eliminate the bias to some extent and to be able to calculate a normalized value of this metric is to use the metric HVR which is the ratio of the HV of Q and of P^*, as follows (Veldhuizen, 1999):

$$\mathrm{HVR} = \frac{\mathrm{HV}(Q)}{\mathrm{HV}(P^*)}. \tag{8.12}$$

For a problem where all objectives are to be minimized, the best (maximum) value of the HVR is one (when $Q = P^*$). For the Pareto-optimal solutions shown in Figure 187, $\mathrm{HV}(P^*) = 71.53$. Thus, $\mathrm{HVR} = 64.80/71.53 = 0.91$. Since this value is close to one, the obtained set is near the Pareto-optimal set. It is clear that values of both HV and HVR metrics depend on the chosen reference point W.

Attainment Surface Based Statistical Metric

Fonseca and Fleming (1996) suggested the concept of an attainment surface in the context of multi-objective optimization. In many studies, the obtained non-dominated solutions are usually shown by joining them with a curve. Although such a curve provides a better illustration of a front, there is no guarantee that any intermediate solution lying on the front is feasible, nor is there any guarantee that intermediate solutions are Pareto-optimal. Fonseca and Fleming argued that instead of joining the obtained non-dominated solutions by a curve, an envelope can be formed marking all those solutions in the search space which are sure to be dominated by the set of obtained non-dominated solutions. Figure 192 shows this envelope for a set of non-dominated solutions. The generated envelope is called an *attainment surface* and is identical to the surface used to calculate the hypervolume (discussed above). Like the hypervolume metric, an attainment surface also signifies a combination of both convergence and diversity of the obtained solutions.

The metric derived from the concept of attainment surface is a practical one. In practice, an MOEA will be run multiple times, each time starting the MOEA from a different initial population or parameter setting. Once all runs are over, the obtained non-dominated solutions can be used to find an attainment surface for each run. Figure 193 shows the non-dominated solutions for three different MOEA runs for the same problem with the same algorithm, but with different initial populations. With a stochastic search algorithm, such as an evolutionary algorithm, it is expected that variations in its performance over multiple runs may result. Thus, such a plot does not provide a clear idea of the true non-dominated front. When two or more algorithms are to be compared, the cluttering of the solutions near the Pareto-optimal front may not provide a clear idea about which algorithm performed better. Figure 194 shows the corresponding attainment surfaces, which can be used to define a metric for a reliable

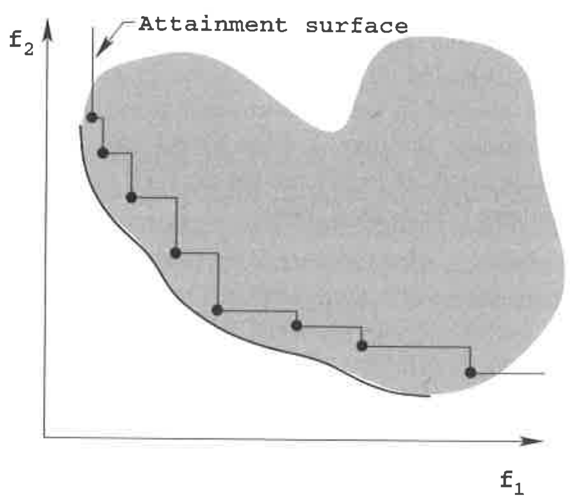

Figure 192 The attainment surface is created for a number of non-dominated solutions.

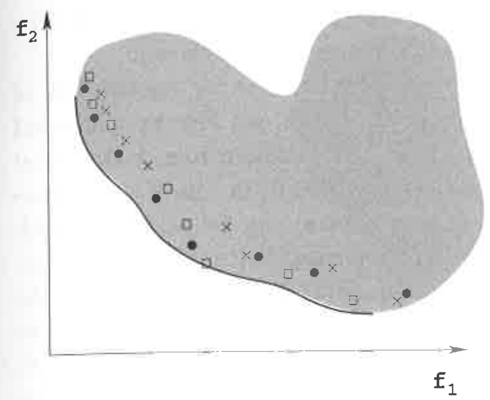

Figure 193 Non-dominated solutions obtained using three different runs of an MOEA.

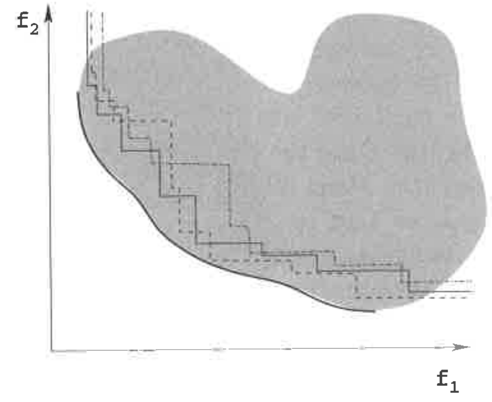

Figure 194 Corresponding attainment surfaces provide a clear understanding of the obtained fronts.

comparison of two or more algorithms or for a clear understanding of the obtained non-dominated front.

First, a number of diagonal imaginary lines, running in the direction of the improvement in all objectives, are chosen. For each line, the intersecting points of all attainment surfaces for an algorithm are calculated. These points will lie on the chosen line and, thus will follow a frequency distribution. Using these points, a number of statistics, such as 25, 50 or 75% attainment surfaces, can be derived. Figure 195 shows an arbitrary cross-line AB and the corresponding intersecting points by the three attainment surfaces. The frequency distribution along the cross-line for a large

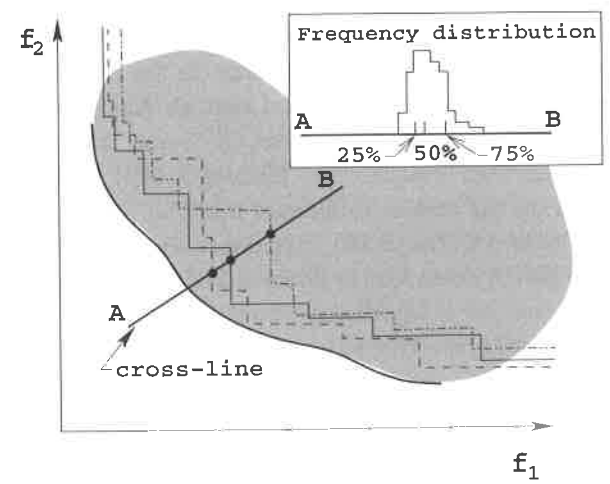

Figure 195 Intersection points on a typical cross-line. A frequency distribution or a histogram can be created from these points.

number of hypothetical attainment surfaces is also marked on the cross-line AB. The points corresponding to 25, 50 or 75% attainment surfaces are also shown.

The frequency distribution and different attainment surfaces for another set of non-dominated solutions obtained using a second algorithm can also be computed likewise. Once the frequency distributions are found on a chosen line, a statistical test (the Mann–Whitney U test (Knowles and Corne, 2000), or the Kolmogorov–Smirnov type test (Fonesca and Fleming, 1996), or others) can be performed with a confidence level β to conclude which algorithm performed better along that line. The same procedure can be repeated for other cross-lines (at different locations in the trade-off regions with different slopes). On every line, one of three decisions can be made with the chosen confidence level: (i) Algorithm A is better than Algorithm B, (ii) Algorithm B is better than Algorithm A, or (iii) no conclusion can be made with the chosen confidence level. With a total of L lines chosen, Knowles and Corne (2000) suggested a metric $[a, b]$ with a confidence limit β, where a is the percentage of times that Algorithm A was found better and b is the percentage of times that Algorithm B was found better. Thus, $100 - (a + b)$ gives the percentage of cases the results were statistically inconclusive. Thus, if two algorithms have performed equally well or equally bad, the percentage values a and b will be small, such as $[2.8, 3.2]$, meaning that, in 94% of the region found by two algorithms no algorithm is better than the other. However, if the metric returns values such as $[98.0, 1.8]$, it can be said that Algorithm A has performed better than Algorithm B. Noting the lines where each algorithm performed better or where the results were inconclusive, the outcome of this statistical metric can also be shown visually on the objective space. Figure 196 shows the regions (by continuous lines) where Algorithm A performed better and regions where the algorithm B performed better. The figure also shows the regions where inconclusive results were found (marked with dashed curves). The results on such a plot can be shown on the *grand* 50% attainment surface, which may be computed as the 50% attainment surface generated by using the combined set of points of two algorithms along any cross-line. In the absence of any preference to any objective function, it can be concluded that an Algorithm A has performed better than Algorithm B if $a > b$, and vice versa. Of course, the outcome will depends on the chosen confidence level. Fonseca and Fleming (1996) nicely showed how such a comparison depends on the chosen confidence limit. In most studies (Fonesca and Fleming, 1996; Knowles and Corne, 2000), confidence levels of 95 and 99% were used.

Knowles and Corne (2000) extended the definition of the above metric for comparing more than two algorithms. For K algorithms, the above procedure can be repeated for $\binom{K}{2}$ distinct pairs of algorithms. Thereafter, for each algorithm k the following two counts can be made:

1. The percentage (a_k) of the region where one can be statistically confident with the chosen confidence level that the algorithm K was not *beaten* by any other algorithm.
2. The percentage (b_k) of the region where one can be statistically confident with the chosen confidence level that the algorithm *beats* all other $(K - 1)$ algorithms.

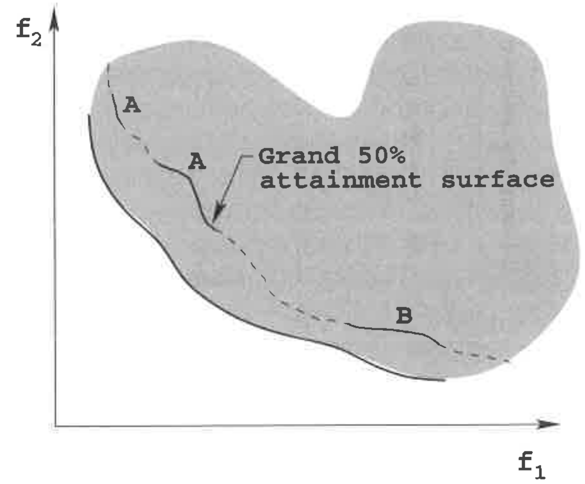

Figure 196 The regions where either Algorithm A or B performed statistically better are shown by continuous lines, while the regions where no conclusion can be made are marked with dashed lines.

It is interesting to note that for every algorithm, $a_k \geq b_k$, since the event described in item 2 is included in item 1 above. An algorithm with large values of a_k and b_k is considered to be good. Two or more algorithms having more or less the same values of a_k should be judged by the b_k value. In such an event, the algorithm having a large value of b_k is better.

Weighted Metric

A simple procedure to evaluate both goals would be to define a weighted metric of combining one of the convergence metrics and one of the diversity measuring metric together, as follows:

$$W = w_1 \mathrm{GD} + w_2 \Delta, \tag{8.13}$$

with $w_1 + w_2 = 1$. Here, we have combined the generational distance (GD) metric for evaluating the converging ability and Δ to measure the diversity-preserving ability of an algorithm. We have seen in the previous two subsections that the GD takes a small value for a good converging algorithm and Δ takes a small value for a good diversity-preserving algorithm. Thus, an algorithm having an overall small value of W means that the algorithm is good in both aspects. The user can choose appropriate weights (w_1 and w_2) for combining the two metrics. However, if this metric is to be used, it is better that a normalized pair of metrics is employed.

Non-Dominated Evaluation Metric

Since both metrics evaluate two conflicting goals, it is ideal to pose the evaluation of MOEAs as a two-objective evaluation problem. If the metric values for one algorithm

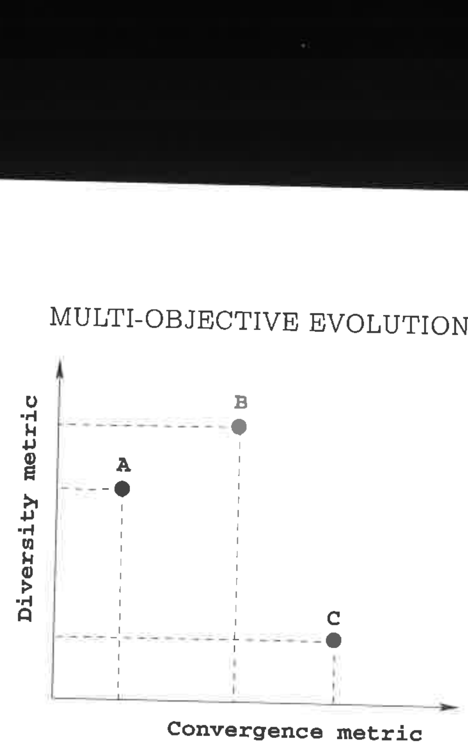

Figure 197 Algorithms A and C produce a non-dominated outcome.

dominate that of the other algorithm, then the former is undoubtedly better than the latter. Otherwise, no affirmative conclusion can be made about the two algorithms. Figure 197 shows the performance of three algorithms on a hypothetical problem. Clearly, Algorithm A dominates Algorithm B, but it cannot be said which is better between Algorithms A and C.

A number of other metrics and guidelines for comparing two non-dominated sets are discussed elsewhere (Hansen and Jaskiewicz, 1998).

8.3 Test Problem Design

In multi-objective evolutionary computation, researchers have used many different test problems with known sets of Pareto-optimal solutions. Veldhuizen (1999) in his doctoral thesis outlined many such problems. Here, we present a number of such test problems which are commonly used. Later, we argue that most of these test problems are not tunable and it is difficult to establish what feature of an algorithm has been tested by these problems. Based on these arguments, we shall present a systematic procedure of designing test problems for unconstrained and constrained multi-objective evolutionary optimization. Moreover, although many test problems were used in earlier studies, the exact locations of the Pareto-optimal solutions were not clearly shown. Here, we make an attempt to identify their exact location using the optimality conditions described in Section 2.5.

Although simple, the most studied single-variable test problem is Schaffer's two-objective problem (Schaffer, 1984):

$$\mathsf{SCH1}: \begin{cases} \text{Minimize} & f_1(x) = x^2, \\ \text{Minimize} & f_2(x) = (x-2)^2, \\ & -A \le x \le A. \end{cases} \tag{8.14}$$

This problem has Pareto-optimal solutions $x^* \in [0, 2]$ and the Pareto-optimal set is a convex set:

$$f_2^* = (\sqrt{f_1^*} - 2)^2.$$

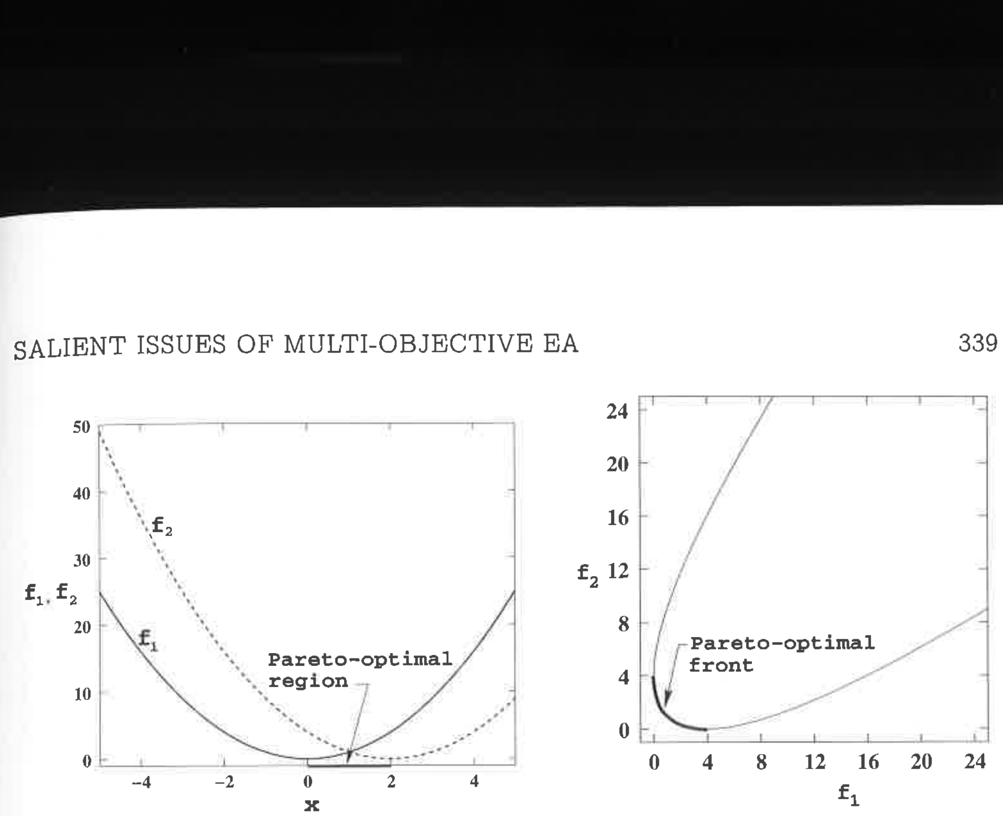

Figure 198 Decision variable and objective space in Schaffer's function SCH1.

in the range $0 \le f_1^* \le 4$. Figure 198 shows the objective space for this problem and the Pareto-optimal front. Different values of the bound-parameter A are used in different studies. Values as low as $A = 10$ to values as high as $A = 10^5$ have been used. As the value of A increases, the difficulty in approaching towards the Pareto-optimal front is enhanced.

Schaffer's second function, SCH2, is also used in many studies (Schaffer, 1984):

$$\mathsf{SCH2}: \begin{cases} \text{Minimize} & f_1(x) = \begin{cases} -x & \text{if } x \le 1, \\ x-2 & \text{if } 1 < x \le 3, \\ 4-x & \text{if } 3 < x \le 4, \\ x-4 & \text{if } x > 4, \end{cases} \\ \text{Minimize} & f_2(x) = (x-5)^2, \\ & -5 \le x \le 10. \end{cases} \tag{8.15}$$

The Pareto-optimal set consists of two discontinuous regions: $x^* \in \{[1, 2] \cup [4, 5]\}$. Figure 199 shows both functions and the objective space. The Pareto-optimal regions are shown by bold curves. Note that although the region BC produces a conflicting scenario with the two objectives (as f_1 increases, f_2 decreases, and vice versa), the scenario for DE is better than that for BC. Thus, the region BC does not belong to the Pareto-optimal region. The corresponding Pareto-optimal regions (AB and DE) are also shown in the objective space. The main difficulty an algorithm may face in solving this problem is that a stable subpopulation on each of the two disconnected Pareto-optimal regions may be difficult to maintain.

Fonseca and Fleming (1995) used a two-objective optimization problem having n variables:

$$\mathsf{FON}: \begin{cases} \text{Minimize} & f_1(\mathbf{x}) = 1 - \exp\left(-\sum_{i=1}^{n}(x_i - \frac{1}{\sqrt{n}})^2\right), \\ \text{Minimize} & f_2(\mathbf{x}) = 1 - \exp\left(-\sum_{i=1}^{n}(x_i + \frac{1}{\sqrt{n}})^2\right), \\ & -4 \le x_i \le 4 \quad i = 1, 2, \ldots, n. \end{cases} \tag{8.16}$$