Project Report

Md Sahil
BCSE-IV
Roll-001710501029

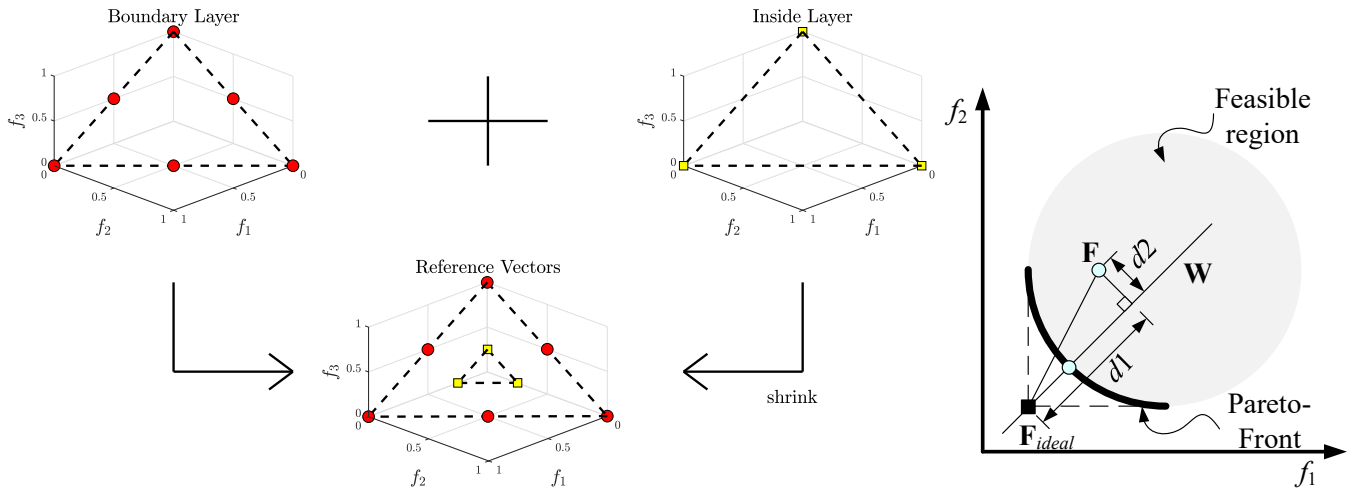


Figure 1: (a) Two-layered approach of generating the set $\mathcal{W}$ of n_{dir} reference points with $n_{div}^{out} = 2$ and $n_{div}^{in} = 1$ divisions in the boundary (outside) and the inside layers, respectively, (b) Illustration of the relation between $\mathbf{F}$, $d1$ and $d2$.

1 Introduction

The project is based around modifying the simulated annealing algorithm knows as AMOSA in order to improve its performance for higher objective problems.

2 Background

2.1 Key Concepts

For decomposition of the objective space into multiple sub-spaces, a set $\mathcal{W} = [\mathbf{W}_1, \cdots, \mathbf{W}_{n_{dir}}]$ of reference points is generated using the Das and Dennis's method. In this method, the reference points are sampled from a unit hyper-plane. If there are n_{div}^{out} uniform divisions per objective, the number reference points generated will be $n_{dir} = \binom{n_{div}^{out}+M-1}{M-1}$. However, with $M = 7$ and $n_{div}^{out} = 7$, as many as $n_{dir} = 1716$ reference points will be required. Moreover, when $n_{div}^{out} < M$, the reference points will only be located at the boundary of the unit simplex.

To avoid this unchecked growth of reference points and to have reference points at non-boundary regions of the simplex, a two-layered strategy is followed when $M \geq 7$. An outside layer with n_{div}^{out} divisions and another inside layer with n_{div}^{in} divisions per objective is considered as shown in Fig. 1a. In the inside layer, the j^{th} element of the i^{th} reference point (w_{ij}) is shrunk according to the following rule

$$w'_{ij} = \frac{1-\tau}{M} + \tau \times w_{ij}, \text{ with } \tau = 0.5. \tag{1}$$

Thus, with the two-layered approach, the total number of reference points become

$$n_{dir} = \binom{n_{div}^{out} + M - 1}{M-1} + \binom{n_{div}^{in} + M - 1}{M-1}. \tag{2}$$

To simplify the optimization of many objectives, within each sub-space, the MaOP is transformed into a single objective optimization problem using the Penalty-based Boundary Intersection (PBI) approach. For a sub-space defined by $\mathbf{W}$, the PBI-based problem is defined as

$$\text{Minimize } PBI(\mathbf{F}, \mathbf{W}, \mathbf{F_{ideal}}) = d1 + \theta d2, \tag{3}$$

$$\text{where, } d1 = \frac{\|(\mathbf{F} - \mathbf{F}_{ideal})^T \times \mathbf{W}\|}{\|\mathbf{W}\|} \text{ and } d2 = \left\| \mathbf{F} - \left(d1 * \frac{\mathbf{W}}{\|\mathbf{W}\|} + \mathbf{F}_{ideal} \right) \right\|. \tag{4}$$

Here, $\mathbf{F}_{ideal}$ denotes the ideal objective vector for the given problem and θ is a user-defined penalty parameter trading-off between convergence (characterized by $d1$) and diversity (characterized by $d2$). An illustration of $d1$ and $d2$ is shown in Fig. 1b such that the PBI function aims to bring $\mathbf{F}$ to the intersection of PF and $\mathbf{W}$.

The population is partitioned in accordance to the sub-spaces $\mathcal{S}$ defined by the reference vectors $\mathcal{W}$ in order to reduce problem complexity. Objective vectors are associated with sub-spaces based on their $d2$ values (Eq. 4). For each objective vector, its $d2$ is calculated against all the sub-spaces. The vector gets associated with the sub-space corresponding to the lowest $d2$ value as follows.

$$\text{associate } \mathbf{F} \text{ with } \mathcal{S}_i \ni d2(\mathbf{W}_i, \mathbf{F}) = min(d2(\mathbf{W}_j, \mathbf{F})) \text{ where } j \in \{1, \cdots, n_{dir}\} \tag{5}$$

The Goal of the optimization problem is to generate objective vectors that are evenly distributed throughout the objective hyperspace. The inclusion of the concept of sub-spaces ensures that the algorithm doesn't just focus on convergence but also diversity. But in order to achieve diversity, decomposition of objective hyperspace alone is not enough to guarantee desired results. In order to ensure the exploration of the whole objective hyperspace the point generation algorithm must also be able to generate diverse points. Thus in addition to decomposition of objective space, an improved point generation algorithm is also introduced as explained in section 3.3.

2.2 Framework of AMOSA and its Drawbacks

AMOSA starts by selecting a unique solution as the current point ($\mathbf{X}_{cur}$). A new solution ($\mathbf{X}_{new}$) is generated by perturbing $\mathbf{X}_{cur}$. An archive ($\mathcal{A}$) is maintained which stores all the non-dominating solutions generated so far. The new solution ($\mathbf{X}_{new}$) is added to $\mathcal{A}$ if it is a non-dominating solution. Also, $\mathbf{X}_{new}$ may be selected as $\mathbf{X}_{cur}$ with a certain probability P that is dependent on the current temperature T. Mathematically,

$$P = \frac{1}{1 + \exp(\Delta dom \times T^n)}. \tag{6}$$

In Eq. (6), n is a non-negative integer and Δdom is a measure of domination of one point over other Mathematically

$$\Delta dom_{\mathbf{F}_i, \mathbf{F}_j} = \prod_{k=1}^{N} \frac{|F_{ik} - F_{jk}|}{R_k} \tag{7}$$

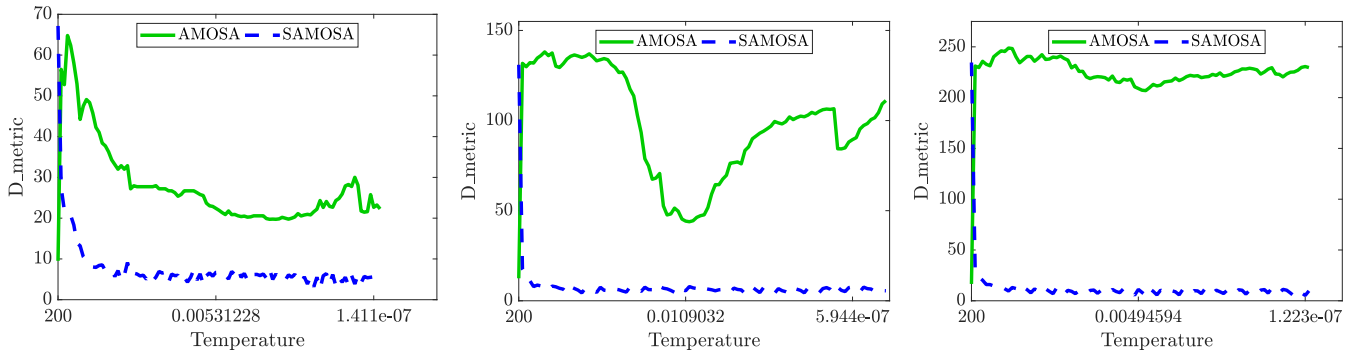
Based on the relationship between the current point $\mathbf{F}_{cur}$, the new generated point $\mathbf{F}_{new}$, and the archive $\mathcal{A}_{\mathbf{F}}$ three situations may arise.

1. The new point $\mathbf{F}_{new}$ is dominated by the current point $\mathbf{F}_{cur}$. In such cases the new point is assigned as the current point with a probability P as stated in equation 6. Here domination of the point with respect to all other points in the archive that dominates the new point is considered. Thus here average domination $\Delta dom_{avg} = \frac{(\sum_K^{i=1} \Delta dom_{\mathbf{F}_i, \mathbf{F}_{new}}) + \Delta\ dom_{\mathbf{F}_{cur}, \mathbf{F}_{new}}}{K+1}$ is used.

2. The new point $\mathbf{F}_{new}$ and the current point $\mathbf{F}_{cur}$ are non-dominating with respect to each other. Here if the new point is non dominating with respect to all other points in the archive then the new point is an addition to the current pareto set. Thus it is added to the archive and assigned as the current point. Otherwise, just like in the previous case, the new point is assigned as the current point with the probability P, were average domination ($\Delta\ dom_{avg}$) with respect to the dominating points in the archive is considered. Also points in the archive dominated by the new point are removed.

3. The new point $\mathbf{F}_{new}$ dominates the current point $\mathbf{F}_{cur}$. If the new point is dominated by points in the archive then the point corresponding to the minimum domination value is set as new current point with a probability $P = \frac{1}{1+\exp(-\Delta dom_{min})}$, where $\Delta\ dom_{min}$ corresponds to the minimum domination value when compared with all dominating points in the archive. Accordingly the new point is set as current point with a probability of $1-P$. If the new point is not dominated by any point in the archive then, the new point is added to the archive and is set as the current point. Also points in the archive dominated by the new point are removed.

Eq. (6) ensures that even dominated solutions are given the chance to be selected as $\mathbf{X}_{cur}$. This probability decreases with decreasing temperature and dominance. The fact that even a dominated solution may also be selected as $\mathbf{X}_{cur}$ and be used to generate $\mathbf{X}_{new}$ helps in escaping local optima.



(a) (a) DTLZ4 ($M = 3$) (b) (b) DTLZ4 ($M = 8$) (c) (c) DTLZ4 ($M = 10$)

Figure 2: Evolution of solution diversity with decrease in temperature for AMOSA variants.

- Where and why is AMOSA lacking? Even though AMOSA converges some solutions to the Pareto-Set, it fails to deliver the ideal solution diversity. AMOSA is majorly based on pareto dominance and doesn't take into consideration the distribution of solutions over the objective space. For example, in problems with biased solution density (like DTLZ4), AMOSA terminates with a very high D_metric (ideally, D_metric should reach zero. As seen from Fig. 2, AMOSA's solution diversity across different number of objectives for DTLZ4 does not have a tendency to converge towards ideal scenario. AMOSA uses Laplacian mutation algorithm to generate new solutions. Laplacian mutate operation generates a new vector from a single parent vector. Thus, this mutation strategy isn't the best approach to generate diverse points throughout the objective space.

2.3 Contributions of the Proposed Work

In order to present SAMOSA, the following modifications are performed to extend the framework of AMOSA:

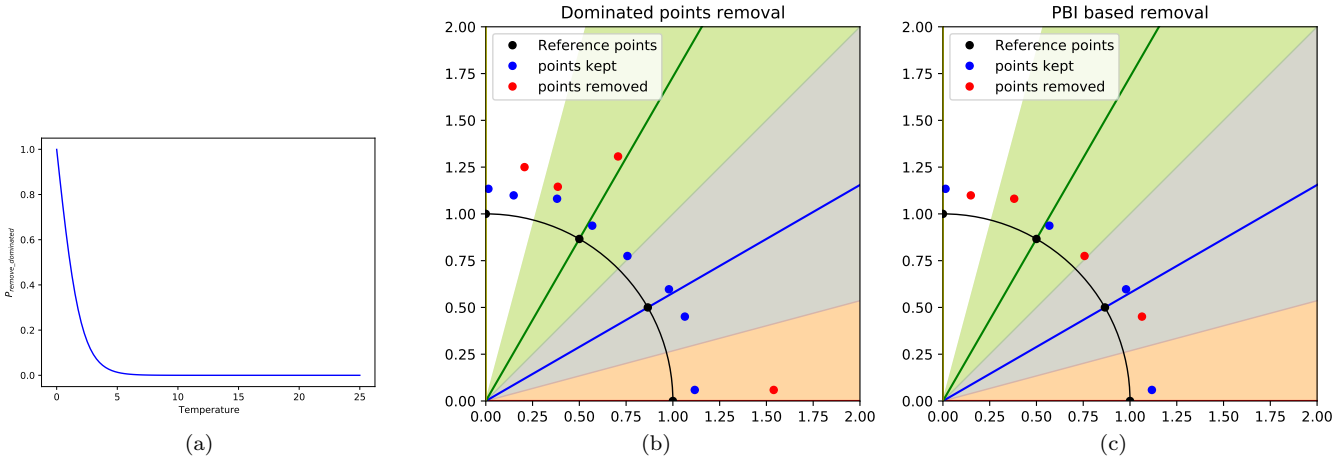
1. *Limiting the archive size*: When newly generated non-dominated solutions are repeatedly added to the archive, its size grows without any limit. In order to avoid such boundless growth, when the archive size exceeds a soft limit (l_{soft}), AMOSA performs Single Linkage Clustering and eliminates solutions from each cluster to reduce the archive size to a hard limit (l_{hard}). In SAMOSA, as reference vector based decomposition already partitions the archive in the objective space, solutions from different sub-spaces are eliminated to reduce the archive size to l_{hard}. This change in archive partitioning enhances the solution diversity in the archive. Moreover, in SAMOSA the solution elimination is dictated by minimizing Penalty-based Boundary Intersection (PBI) which along with Pareto-dominance enhances the selection pressure as recommended in. The solution elimination also ensures that if a sub-space ever gets associated with a solution, it will continue to have at least one associated solution throughout the evolutionary process.

2. *Selecting the solution to be perturbed*: In AMOSA, the selection of the solution to be perturbed ($\mathbf{X}_{cur}$) is only guided by the position of the solution generated in immediately previous perturbation ($\mathbf{X}_{new}$) with respect to $\mathcal{A}$. However, in SAMOSA, if $\mathbf{X}_{cur}$ (as per AMOSA's strategy) is from a sub-space already considered in the last few iterations, then $\mathbf{X}_{cur}$ randomly chosen from a non-empty unvisited sub-spaces. This modification ensures that perturbations occur in all those sub-spaces where at least one solution currently exists. Thus, all the sub-spaces are equally likely to generate a non-dominated solution.

3. *Mating strategy to generate new solutions*: In AMOSA, $\mathbf{X}_{cur}$ is always perturbed in a random direction. However, due to the neighborhood property of sub-spaces, in SAMOSA, mating among solutions from neighboring sub-spaces can perturb $\mathbf{X}_{cur}$. This strategy can have huge benefits in generating a solution in the empty (unexplored) sub-spaces.

3 Algorithmic Framework

The overall framework of Subspace-based AMOSA (SAMOSA) is presented in Algorithm 2. SAMOSA takes the name of the test function, l_{hard}, l_{soft}, T_{max}, T_{min}, and $MaxFES$ as input. The archive $\mathcal{A}$ is initialized with l_{hard}



(a) (b) (c)

Figure 3: (a) Variation in probability of removing dominated solutions with temperature, (b) Dominated solutions for removal, (c) Diversity preserving selection.

number of randomly generated points within the bounds of the decision variables (line 1). These points are then associated with the sub-spaces (line 2). The main loop of the algorithm (lines 6 to 28) starts with temperature $T = T_{max}$ and ends at $T = T_{min}$, where at each iteration the temperature is reduced as $T = T \times \alpha$ (line 27).

The parameter α, i.e., the cooling rate, depends on $MaxFES$. For each candidate, $\mathbf{F}_{new}$ is evaluated once. Thus, to maintain the standard fitness evaluation budget of the problem, α must satisfy

$$\log_{\alpha}\left(\frac{T_{max}}{T_{min}}\right) \times n_{iter} \leq MaxFES. \tag{8}$$

Thus, the maximum value of alpha can be calculated as

$$\alpha = \left(\frac{T_{min}}{T_{max}}\right)^{\frac{n_{iter}}{MaxFES}}. \tag{9}$$

In Algorithm 2, α is set during initialization on line 1.

At each temperature, the internal loop (lines 7 to 26) runs n_{iter} times, where n_{iter} is set as a multiple of n_{dir}. During each iteration, a point is assigned as $\mathbf{X}_{cur}$ (as explained in Section 3.2), $\mathbf{X}_{cur}$ is perturbed (as explained in Section 3.3) to generate $\mathbf{X}_{new}$ and corresponding $\mathbf{F}_{new}$ (line 14). Depending on the value of $\mathbf{F}_{new}$, the archive $\mathcal{A}$ is updated.

3.1 Decomposition of the Objective Space

The objective space is partitioned into n_{dir} sub-spaces by defining $\mathcal{W}$ (Algorithm 2, line 4). The i^{th} sub-space ($\mathcal{S}_i$) is defined by the i^{th} reference vector $\mathbf{W}_i$ as described in Section 2.1. For defining the set $\mathcal{W}$ of reference vectors, the two-layered approach of Das and Dennis is followed as it is the standard approach of decomposition in most of the algorithms such as NSGA-III, MOEA/DD, NAEMO, ESOEA and AR-MOEA.

3.2 Subspace based Current Point Selection

SAMOSA's current point ($\mathbf{X}_{cur}$) selection is dependent on whether AMOSA's $\mathbf{X}_{cur}$ (using Algorithm 1) is from a sub-spaces considered in the recent iterations. All non-empty unvisited sub-spaces are maintained in a list $\mathbf{V}$ (Algorithm 2, line 4). If $\mathbf{X}_{cur}$ is not from an unvisited non-empty subspace, i.e., the subspace associated with $\mathbf{X}_{cur}$ is not in $\mathbf{V}$, a random sub-space from a $\mathbf{V}$ is assigned as $\mathbf{X}_{cur}$ (Algorithm 2, line 22). Once all the non-empty sub-spaces are visited, i.e., $\mathbf{V}$ is empty, $\mathbf{V}$ is refilled with all the non-empty sub-spaces (Algorithm 2, line 9).

Algorithm 1 AMOSA_SELECT function

Require: $\mathcal{A}$: the current archive in decision space; $\mathbf{X}_{cur}$: current point; $\mathbf{X}_{new}$: new perturbed point; T: current temperature

Ensure: Updated $\mathbf{X}_{cur}, \mathcal{A}$

1: **procedure** AMOSA_SELECT($\mathcal{A}, \mathbf{X}_{cur}, \mathbf{X}_{new}, T$)
2: Check domination status of $\mathbf{X}_{new}$ and $\mathbf{X}_{cur}$
3: **if** $\mathbf{X}_{new}$ is dominated by $\mathbf{X}_{cur}$ **then**
4: K = total number of points in $\mathcal{A}$ that dominates $\mathbf{X}_{new}$
5: $\Delta dom_{avg} = \frac{(\sum_{K}^{i=1} \Delta dom_{\mathbf{F}_i,\mathbf{F}_{new}}) + \Delta\ dom_{\mathbf{F}_{cur},\mathbf{F}_{new}}}{K+1}$
6: $P = \frac{1}{1+\exp(\Delta dom_{avg} * T)}$
7: set $\mathbf{X}_{cur}$ as $\mathbf{X}_{new}$ with probability P
8: **else if** $\mathbf{X}_{new}$ and $\mathbf{X}_{cur}$ are non-dominating w.r.t each-other **then**
9: check domination status of $\mathbf{X}_{new}$ and points in $\mathcal{A}$
10: **if** $\mathbf{X}_{new}$ is dominated by K ($K >= 1$) points in $\mathcal{A}$ **then**
11: $\Delta dom_{avg} = \frac{(\sum_{K}^{i=1} \Delta dom_{\mathbf{F}_i,\mathbf{F}_{new}})}{K}$
12: $P = \frac{1}{1+\exp(\Delta dom_{avg} * T)}$
13: set $\mathbf{X}_{cur}$ as $\mathbf{X}_{new}$ with probability P
14: **else**
15: $\mathcal{A} \leftarrow \mathcal{A} \cup \mathbf{X}_{new}$
16: $\mathbf{X}_{cur} \leftarrow \mathbf{X}_{new}$
17: **end if**
18: **else if** $\mathbf{X}_{new}$ dominates $\mathbf{X}_{cur}$ **then**
19: check domination status of $\mathbf{X}_{new}$ and points in $\mathcal{A}$
20: **if** $\mathbf{X}_{new}$ is dominated by K ($K >= 1$) points in $\mathcal{A}$ **then**
21: Δdom_{min} = minimum of the distance of domination between $\mathbf{F}_{new}$ and K points in $\mathcal{A}$
22: $P = \frac{1}{1+\exp(-\Delta dom_{min})}$
23: $\mathbf{X}_{cur} \leftarrow \mathbf{X}_{new}$
24: set the point in $\mathcal{A}$ which corresponds to Δdom_{min} as the $\mathbf{X}_{cur}$ with probability P
25: **else**
26: $\mathcal{A} \leftarrow \mathcal{A} \cup \mathbf{X}_{new}$
27: $\mathbf{X}_{cur} \leftarrow \mathbf{X}_{new}$
28: **end if**
29: **end if**
30: **return** $\mathbf{X}_{cur}, \mathcal{A}$
31: **end procedure**

3.3 Perturbation Scheme

For perturbing the chosen $\mathbf{X}_{cur}$ to generate a new solution $\mathbf{X}_{new}$, Algorithm 3 is followed. It consists of probabilistic switching between Differential Evolution (DE) and Simulated Binary Crossover (SBX) reproduction schemes followed by laplacian mutation strategy. DE requires three more parent points other than $\mathbf{X}_{cur}$. These three points are sampled from the non-empty sub-space nearest to the sub-space that $\mathbf{X}_{cur}$ belongs to. Similarly, SBX requires one more parent point other than $\mathbf{X}_{cur}$ which is also sampled int the same manner.

3.4 Candidate Selection Strategy

SAMOSA uses a subspace based solution selection. The aim of this selection stage (Algorithm 4) is restricting the size of the archives and at the same time improving solution diversity. Algorithm 4 is triggered when the archive size exceeds l_{soft} (Algorithm 2, line 16).

The first step involves a probabilistic removal of dominated solution for the archive. The probability that the dominated solution gets removed is a function of the current temperature T (Algorithm 4, line 2). As shown in Fig. 3a at high temperatures the probability is very low, thus dominated solutions remain in the archive. This ensures

Algorithm 2 Basic framework of SAMOSA

Require: $prob(M, N)$: problem specifications; l_{hard}: hard limit; l_{soft}: soft limit; n_{dir}: number of sub-spaces; T_{max}, T_{min}: maximum and minimum temperatures; $MaxFES$: function evaluation budget
Ensure: $\mathcal{A}$: Estimated Pareto-optimal Set

1: $\mathcal{A}, \alpha, n_{iter}, \mathcal{W} \leftarrow$ INITIALIZE($prob, M, N, l_{soft}, n_{dir}, MaxFES$)
2: ASSOCIATE-ARCHIVE($\mathcal{A}$, $\mathcal{W}$)
3: $T = T_{max}$; $i = 0$ (Initialize temperature and iteration counter)
4: $\mathbf{V}$ = GET_ASSOCIATED_SUBSPACES($\mathcal{S}$) $\leftarrow$ list of non-empty unvisited sub-spaces
5: $\mathbf{X}_{cur} \leftarrow$ random point from $\mathcal{A}$
6: **while** $T > T_{min}$ **do**
7: **while** $i < n_{iter}$ **do**
8: **if** mod $(i, n_{dir} + 1) = 0$ or $\mathbf{V}$ is empty **then**
9: $\mathbf{V}$ = GET_ASSOCIATED_SUBSPACES($\mathcal{S}$)
10: **end if**
11: $\mathcal{S}_r \leftarrow$ Sub-space with which $\mathbf{X}_{cur}$ is associated
12: **if** $\mathcal{S}_r$ in $\mathbf{V}$ ($\mathcal{S}_r$ is unvisited) **then**
13: remove $\mathcal{S}_r$ from $\mathbf{V}$ (Set $\mathcal{S}_r$ as visited)
14: $\mathbf{X}_{new}$ = PERTURB_ASSOCIATE($\mathbf{X}_{cur}, \mathcal{A}, \mathbf{W}, \mathcal{S}$)
15: $\mathbf{X}_{cur}, \mathcal{A} \leftarrow$ AMOSA_SELECT($\mathcal{A}, \mathbf{X}_{cur}, \mathbf{X}_{new}, T$) (Revise $\mathcal{A}, \mathcal{A}_{\mathbf{F}}$ and $\mathbf{X}_{cur}$)
16: **if** $|\mathcal{A}| > l_{soft}$ **then**
17: $\mathcal{A}, \mathcal{A}_{\mathbf{F}} \leftarrow$RESIZE($\mathcal{A}, \mathcal{A}_{\mathbf{F}}, \mathcal{S}, l_{hard}, T$) (To avoid unchecked growth of $\mathcal{A}$)
18: **end if**
19: **else**
20: Randomly select another sub-space $\mathcal{S}_r$ from $\mathbf{V}$
21: Remove $\mathcal{S}_r$ from $\mathbf{V}$
22: $\mathbf{X}_{cur} \leftarrow$ a random point from $\mathcal{A}$ associated with $\mathcal{S}_r$
23: Same as steps 14 to 18
24: **end if**
25: $i = i + 1$ (Update iteration counter)
26: **end while**
27: $T = T \times \alpha$ (Update temperature)
28: **end while**

Algorithm 3 Algorithm to mutate points and generate new point

Require: $\mathbf{X}_{cur}$: The current point; $\mathcal{A}$: the solution archive; $\mathbf{W}$: the set of reference points; $\mathcal{S}$: the sub-space set
Ensure: $\mathbf{X}_{new}$: generated new point

1: **procedure** PERTURB_ASSOCIATE($\mathbf{X}_{cur}, \mathcal{A}, \mathbf{W}, \mathcal{S}$)
2: $P_{perturb} = 0.5$
3: **if** RAND( ) ¡ $P_{perturb}$ **then**
4: $\mathbf{X}_{new} \leftarrow$ DE($\mathbf{X}_{cur}, \mathcal{A}, \mathbf{W}, \mathcal{S}$)
5: **else**
6: $\mathbf{X}_{new} \leftarrow$ SBX($\mathbf{X}_{cur}, \mathcal{A}, \mathbf{W}, \mathcal{S}$)
7: **end if**
8: $\mathbf{X}_{new} \leftarrow$ LAPLACIAN_MUTATE($\mathbf{X}_{new}$)
9: **return** $\mathbf{X}_{new}$
10: **end procedure**

better exploration of the objective space at high temperatures. At low temperatures the probability is high, resulting removal of dominated solutions. This ensures that the solutions converge better at lower temperatures.

After removal of dominated solutions (Fig. 3b), if the archive size is still greater than l_{hardl}, the candidate selection from different sub-spaces reduces the archive size to l_{hard}. The while loop (lines 6 to 14) keeps on removing points with the worst PBI values from sub-spaces with the highest number of points till the desired archive size is

Algorithm 4 Algorithm to reduce size of archive to l_{hard}

Require: $\mathcal{A}$: current set of solution points; $\mathcal{S}$: set of sub-spaces; l_{hard}: hard limit of archive size; T: current temperature

Ensure: $\mathcal{A}$, $\mathcal{A}_{\mathbf{F}}$: archives with size reduced to l_{hard}

procedure RESIZE($\mathcal{A}, \mathcal{A}_{\mathbf{F}}, \mathcal{S}, l_{hard}, T$)
  $P_{remove\_dominated} = 2 - \frac{2}{1+\exp(-T)}$
  if RAND( )$< P_{remove\_dominated}$ **then**
    $\mathcal{A}, \mathcal{A}_{\mathbf{F}}$ = REMOVE_DOMINATED($\mathcal{A}, \mathcal{A}_{\mathbf{F}}$)
  end if
  while size of $\mathcal{A} > l_{hard}$ **do**
    $\mathcal{S}_r \leftarrow$ sub-space with highest number of points
    $\mathbf{X}_r \leftarrow$ point in $\mathcal{S}_r$ with worst PBI value
    $\mathbf{F}_r \leftarrow$ point in objective space corresponding to $\mathbf{X}_r$
    if number of points in $\mathcal{S} > 1$ **then**
      remove $\mathbf{X}_r$ from $\mathcal{A}$
      remove $\mathbf{F}_r$ from $\mathcal{A}_{\mathbf{F}}$
    end if
  end while
  return $\mathcal{A}$, $\mathcal{A}_{\mathbf{F}}$
end procedure

achieved, as shown in Fig. 3c.