Static and Dynamic Multimodal Optimization by Improved Covariance Matrix Self-Adaptation Evolution Strategy with Repelling Subpopulations

Ali Ahrari*, *Member, IEEE* Saber Elsayed*, *Member, IEEE,* Ruhul Sarker*, *Member, IEEE,* Daryl Essam*, *Member, IEEE,* and Carlos A. Coello Coello+, *Fellow, IEEE*

***Abstract*—The covariance matrix self-adaptation evolution strategy with repelling subpopulations (RS-CMSA-ES) is one of the most successful multimodal optimization methods currently available. However, some of its components may become inefficient in certain situations. This study introduces the second variant of this method, called RS-CMSA-ESII. It improves the adaptation schemes for the normalized taboo distances of the archived solutions and the covariance matrix of the subpopulation, the termination criteria for the subpopulations, and the way in which the infeasible solutions are treated. It also improves the time complexity of RS-CMSA-ES by updating the initialization procedure of a subpopulation and developing a more accurate metric for determining critical taboo regions. The effects of these modifications are illustrated by designing controlled numerical simulations. RS-CMSA-ESII is then compared with the most successful and recent niching methods for multimodal optimization on a widely adopted test suite. The results obtained reveal the superiority of RS-CMSA-ESII over these methods, including the winners of the competition on niching methods for multimodal optimization in previous years. Besides, this study extends RS-CMSA-ESII to dynamic multimodal optimization and compares it with a few recently proposed methods on the modified moving peak benchmark functions.**

***Index Terms*—Continuous optimization, niching, evolutionary algorithm, dynamic optimization,**

I. Introduction

AN optimization problem has been traditionally perceived as that of finding a single best solution (the global optimum) given an objective function, decision parameters, and possibly, problem constraints. There are real-world problems with features that are not easy to formulate, e.g. a product's aesthetics, or a company's preferences towards specific suppliers. In these situations, the decision-maker may be interested or even need a set of diverse optimal solutions instead of one. The best solution to the actual problem may then be selected from this diverse set of near-optimal solutions considering hard-to-formulate aspects of the actual problem. Additionally, knowing all near-optimal solutions of a problem might be crucial in certain problems. For example, a design engineer may need to know all the resonance frequencies of a mechanical system [1].

*Authors are with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT, Australia (E-mail: {a.ahrari; s.elsayed; r.sarker; d.essam}@adfa.edu.au, aliahrari1983@gmail.com)

+The author is with is with CINVESTAV-IPN, Departamento de Computación, Mexico City, Mexico; Basque Center for Applied Mathematics (BCAM) & Ikerbasque, Spain (e-mail: ccoello@cs.cinvestav.mx)

Finding multiple global optima, and in some cases, good local optima, can be achieved by multimodal optimization (MMO), which can be perceived as an extension of global optimization. Methods tailored for MMO employ *niching*, a diversity preservation strategy which makes it possible to converge to distinct global optima. The field of MMO has attracted a lot of interest in the recent decade. A growing number of niching strategies have been proposed [1] and combined with a variety of population-based optimization methods such as differential evolution and genetic algorithms [2]. Furthermore, a competition on niching methods for MMO has been regularly held at some reputable international conferences such as the *IEEE Congress on Evolutionary Computation* (CEC) and the *Genetic and Evolutionary Computation Conference* (GECCO) since 2013 [3].

Multimodal optimization also plays an important role in dynamic optimization, a field of optimization in which the problem landscape frequently changes over time [4], [5]. Quite often, these changes are not substantial [6], and the problem landscape after a change resembles the one before it [7]. An efficient dynamic optimization method should be able to exploit past information to re-optimize the changed problem as fast as possible. This can be achieved by analyzing the time-history of the global minimum. However, since the values of the optima change as well, a local optimum in the previous time step may become the global one in the new time step [8]. Therefore, an efficient dynamic optimization method should track the history of all good local optima even if only the best solution is desired.

Many MMO methods can have a good performance on simple problems. MMO imposes all the challenges associated with global optimization (e.g., presence of deceptive local optima, absence of a global structure, ill-conditioned landscapes, high dimensionality [9]). Besides, there are challenges peculiar to MMO, such as the existence of global optima with irregular shapes, sizes, and distribution. Many niching strategies and MMO methods fail in addressing these challenges effectively. More specifically, niching strategies that rely on a user-tuned distance metric (the *niche radius*) are vulnerable to problems in which global optima are irregularly distributed or have attraction regions of different size and shape. Therefore, many MMO methods show a significant performance drop when facing moderately hard problems. In the following, some of the most remarkable studies in the field of MMO are briefly reviewed.

A. Related Studies

Fitness sharing [10] and crowding [11] are the earliest niching strategies for MMO. Fitness sharing reduces the fitness of the solutions which are too close to each other. In crowding, offspring compete with their nearest parents for survival. Several varieties of these strategies have been later introduced for genetic algorithms [12], [13]; however, they can be integrated with other population-based methods after some customization.

Many MMO methods restrict the competition or information-sharing to close solutions, e.g., using a ring topology in particle swarm optimization (PSO) [14]. Some other methods, such as locally informed particle swarm (LIPS) [15] and close neighbor mobility optimization algorithm (CNMM) [16] determine neighbor particles based on their Euclidean distance. A similar idea has been followed with differential evolution by performing local mutation and recombination [17], [18], [19], [20]. A different mutation scheme was proposed in distributed individuals differential evolution (DIDE) [21] according to which random solutions are generated in the neighborhood of a solution for mutation. Voronoi neighborhood based crowding DE (VNCDE) [22] employs a more unconventional neighborhood criterion based on the Voronoi diagram.

Some studies have defined an auxiliary objective as a niching strategy. This strategy converts the MMO problem into a multiobjective optimization problem, a technique which is referred to as *multi-objectivization* [23]. For example, the bi-objective multi-population genetic algorithm (BMPGA) [24], defines minimization of the gradient vector norm as the second objective. This method requires derivatives of the objective function which are not available in black-box optimization problems. Deb and Saha [25] followed a similar methodology but provided a heuristic to avoid calculation of the gradient of the objective function. A few other methods [26], [27] defined an auxiliary objective to maximize the average distance of solutions to each other. Although the auxiliary objective may sufficiently encourage diversity, Yu et al. [28] defined two auxiliary objectives for this purpose in their DE-based method.

The success of the covariance matrix adaptation evolution strategy (CMA-ES) [29] in global optimization and its remarkable capability in learning the shape of the basin has resulted in several MMO methods using CMA-ES as the core search algorithm. For example, covariance matrix adaptation with adaptive niching (NCMA) [30] determines an elliptic shape for the limits of a niche by using the Mahalanobis distance metric given the covariance matrix of the population distribution. It adapts the niche radius such that each niche has ten members. The covariance matrix self-adaptation evolution strategy with repelling subpopulation (RS-CMSA-ES) [31] combines a strategy based on repelling subpopulations with CMSA-ES [32], a simple but effective evolution strategy. This method can learn both the shapes and the relative sizes of global optima.

Clustering has emerged as a popular strategy for MMO, considering the recent publications on this topic. Preuss [33] combined CMA-ES with the nearest-better clustering (NBC), a strategy to provide a rough estimate for the location of the niches based on initial random sampling and clustering. This strategy has also been used in combination with DE [34]. An improved variant of NBC-CMA-ES, named NEA2 [35], won the competition on niching methods for MMO held at the *2013 IEEE Congress on Evolutionary Computation* (CEC'2013). Automatic niching differential evolution (ANDE) [36] combines affinity propagation clustering [37] for niching with a contour prediction approach and a two-level local search to estimate the approximate location of the optimum. NCjDE-2LS$_{\text{ar}}$ [38] combines jDE with density-based spatial clustering of applications with noise (DBSCAN) [38], two local search mechanisms and an external archive. A few studies have employed other clustering methods, such as K-means clustering [39], and clustering solutions based on the similarity of their values [40].

The niching migratory multi-swarm optimizer (NMMSO) [41] divides the population into multiple swarms which search for niches in parallel. This method can migrate particles to create new swarms or to combine two swarms if they are exploring the same basin. NMMSO performs spectacularly well on low-dimensional separable problems (e.g., the Vincent function). It outperforms NEA2 on the CEC'2013 test suite for MMO; however, it fails when the problem landscape is more complicated, and the problem's dimensionality is not low.

Maree et al. [42] developed a two-phase MMO method. The first phase employs a sequential clustering heuristic based on the hill-valley method developed by Ursem [43] to cluster a population of randomly sampled solutions. In the second phase, a core search algorithm initiates from each cluster. They tested their hill-valley clustering method with different core search methods based on evolution strategies. Later on, the authors improved their method by introducing new stopping criteria to terminate subpopulations that are deemed unlikely to converge to a new global optimum [44]. This method, called HillVallEA, has won the aforementioned competition on niching methods for MMO in 2018 and 2019.

As discussed earlier, MMO plays an important role in dynamic optimization, and thus, there have been several studies on dynamic MMO (see [45] for a review of these methods). The moving peak benchmark (MPB) [46] has served as a widely-accepted tunable benchmark generator for dynamic MMO [45] owing to its flexibility and formulation simplicity. More recently, dynamic MMO has been formulated for tracking multiple global optima over time. Some practical applications of this formulation include dynamic multipath routing [47], dynamic tracking of multiple targets [47], and solving a time-dependent system of equations [48]. Some recent studies have concentrated on this type of dynamic MMO [49], [48].

B. Contribution and Outline

This study introduces an improved RS-CMSA-ES, denoted by RS-CMSA-ESII, which overcomes some major drawbacks of the existing variant. The introduced improvements include 1) a novel scheme for learning the basin sizes of the global minima, 2) a new adaptation scheme for the mutation profile

when elite solutions are involved, 3) two additional termination criteria that can predict whether a subpopulation can find a new global minimum, 4) a new bound-handling mechanism, 5) a time-efficient initialization method, and 6) a more accurate estimate for determining critical taboo regions. As a result of these improvements, RS-CMSA-ESII outperforms the best existing MMO methods by a clear margin, at least when evaluated on the widely accepted CEC'2013 test suite for MMO [3]. Finally, this study extends RS-CMSA-ESII to dynamic MMO problems by combining it with a prediction method to exploit past information.

The rest of this article is organized as follows: Section II provides a brief review of RS-CMSA-ES. The improved method, RS-CMSA-ESII, is explained in Section III. Section IV runs a controlled numerical simulation to clarify the importance of the improvements made to the components of RS-CMSA-ES. Section V compares RS-CMSA-ESII with the best and most well-known niching methods for MMO. Section VI extends RS-CMSA-ESII to dynamic MMO and compares its performance with a few relevant and recently developed methods for this purpose. Finally, conclusions are drawn in Section VII.

II. A Brief Review of RS-CMSA-ES

In RS-CMSA-ES, N_s subpopulations search the space of decision parameters in parallel. Ideally, each subpopulation would converge to a distinct global minimum. Each subpopulation ($\mathbb{P}_i,\ i = 1, 2, \ldots, N_s$) has its center ($\boldsymbol{x}_{\text{mean}_i}$), global step size ($\sigma_{\text{mean}i}$), covariance matrix ($\mathbf{C}_i$), and best solution ($\boldsymbol{x}_{\text{best}_i}$). Diversity is preserved by enforcing weaker subpopulations to stay away from taboo regions. These taboo regions are ellipsoids whose centers are taboo points. For $\mathbb{P}_i$, these taboo points are the union of:

- previously identified global minima stored in an archive (denoted by $\mathbb{A}$) unless $\boldsymbol{x}_{\text{best}_i}$ is better than the value of that archived solution.
- centers of superior subpopulations, which are the subpopulations whose best values are better than $\boldsymbol{x}_{\text{best}_i}$. In this case, $\mathbb{P}_i$ is considered as an inferior subpopulation to those subpopulations.

The m^{th} solution in $\mathbb{A}$ ($m = 1, 2, \ldots, |\mathbb{A}|$) has three properties: location ($\boldsymbol{x}_{A_m}$), value (f_{A_m}), and a scalar parameter called the normalized taboo distance ($\hat{d}_{A_m}$), which determines the relative size of the corresponding taboo region. This parameter is adapted during the optimization process. For subpopulation $\mathbb{P}_i$, the shapes of taboo regions are determined by the covariance matrix of $\mathbb{P}_i$. In contrast, the sizes of these taboo regions are determined by both the mutation profile of $\mathbb{P}_i$ and the normalized taboo distances of the taboo regions (for archived solutions) and default normalized taboo distance $\hat{d}_{\text{def}}$ (for superior subpopulations); for example, Fig. 1 illustrates the taboo regions for $\mathbb{P}_2$ in the presence of archived solutions $\boldsymbol{x}_{A_1}$, $\boldsymbol{x}_{A_2}$ with $\hat{d}_{A_1} = 0.5$ and $\hat{d}_{A_2} = 1.5$, the superior subpopulation $\mathbb{P}_1$, and the inferior subpopulation $\mathbb{P}_3$. For this example, $\hat{d}_{\text{def}} = 0.75$.

The evolution scheme for each subpopulation is based on CMSA-ES [32], except that elitism has been incorporated to

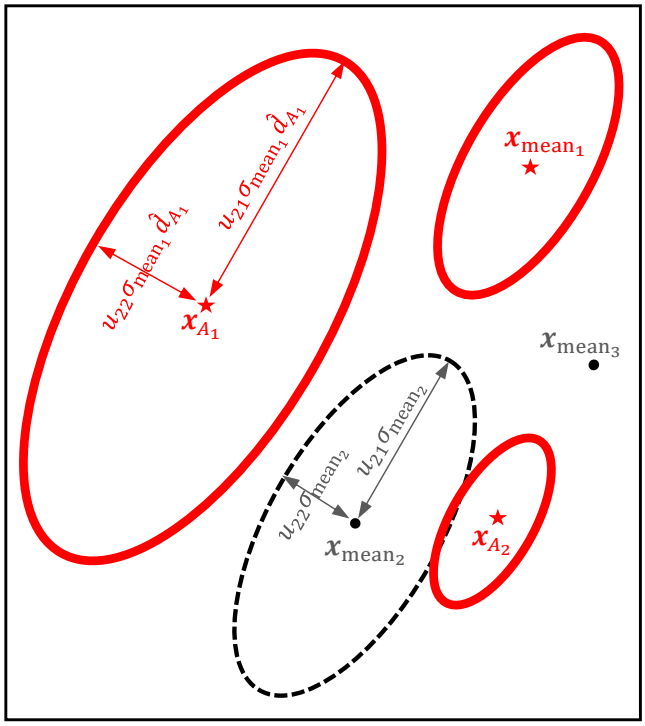


Fig. 1. Illustrations of taboo regions (ellipses with solid lines) for the subpopulation $\mathbb{P}_2$ in the presence of a superior subpopulation ($\mathbb{P}_1$), an inferior subpopulation ($\mathbb{P}_3$), and two archived solutions $\boldsymbol{x}_{A_1}$, $\boldsymbol{x}_{A_2}$ with $\hat{d}_{A_1} = 1.5$ and $\hat{d}_{A_2} = 0.5$. $\hat{d}_{\text{def}} = 0.75$ and u_{21}, u_{22} are the square root of the first and the second eigenvalue of $\mathbf{C}_2$.

the selection scheme. At each iteration, the j^{th} solution is sampled using the sampling strategy of CMSA-ES:

$$\begin{aligned} \sigma_j \leftarrow \sigma_{\text{mean}} \exp\left(\tau_\sigma \mathcal{N}(0,1)\right),\ \boldsymbol{s}_j \leftarrow \mathcal{N}_{\boldsymbol{D}}(\mathbf{0},\mathbf{1}), \\ \boldsymbol{x}_j \leftarrow \boldsymbol{x}_{\text{mean}} + \sigma_j\, \boldsymbol{s}_j, \end{aligned} \tag{1}$$

in which τ_σ is the learning rate for the global step size and $\mathcal{N}_{\boldsymbol{D}}(\mathbf{0},\mathbf{1})$ samples a vector of D random numbers from the standard normal distribution.

$\boldsymbol{x}_j$ should be outside all taboo regions defined for $\mathbb{P}_i$ to be a taboo-acceptable solution; otherwise, it is rejected, and a new $\boldsymbol{x}_j$ is sampled. If multiple successive samples are rejected, the taboo regions temporarily and slightly shrink. Since checking against all taboo regions may require a lot of computations, $\boldsymbol{x}_j$ is checked against critical taboo points only to improve the time-complexity of the sampling process. Critical taboo points are a subset of taboo points whose taboo regions may reject at least 1% of the sampled solution ($P_{\text{trej}} \geq 0.01$). Calculation of P_{trej} is not simple, and RS-CMSA-ES employs a simple model to provide an upper estimate for P_{trej} of each taboo region [31].

After sampling and evaluating λ taboo-acceptable solutions, the best N_{elt} solutions from the previous iteration are appended to the subpopulation. These solutions are then sorted according to their values, and the μ-best solutions are selected to update $\mathbb{P}_i$:

$$\mathbf{C} \leftarrow \left(1 - \frac{1}{\tau_{\text{c}}}\right)\mathbf{C} + \frac{1}{\tau_{\text{c}}} \sum_{j=1}^{\mu} w_j \left(\boldsymbol{s}_j \boldsymbol{s}_j^T\right), \tag{2}$$

in which τ_c is the decay time constant [32], and w_j's are logarithmically decreasing weights as used for CMA-ES [29]:

$$w_j = \frac{\ln(\mu+1) - \ln(j)}{\sum_{k=1}^{\mu} (\ln(\mu+1) - \ln(k))}, j = 1, 2, \ldots, \mu. \quad (3)$$

$\boldsymbol{x}_{\text{mean}}$ and σ_{mean} are updated as follows [39]:

$$\boldsymbol{x}_{\text{mean}} \leftarrow \sum_{j=1}^{\mu} w_j \boldsymbol{x}_j, \; \sigma_{\text{mean}} \leftarrow \sigma_{\text{mean}} \left(\frac{\prod_{j=1}^{\mu} \sigma_j^{w_j}}{\prod_{j=1}^{\lambda+N_{\text{elt}}} \sigma_j^{\left(\frac{1}{\lambda+N_{\text{elt}}}\right)}} \right) \quad (4)$$

When all subpopulations have been terminated because of convergence, stagnation, or divergence, their best solutions ($\boldsymbol{x}_{\text{best}_i}$'s) are analyzed. The best solutions that do not represent a global minimum, when compared with the best archived value ($f_A^{\min}$), are discarded. The remaining ones are checked against the archived solutions using the hill-valley strategy [43] to reveal which ones represent new global minima. The hill-valley heuristic is a simple yet efficient technique to detect whether two solutions (say $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$) share the same basin. It evaluates a third solution ($\boldsymbol{x}_3$) between these two solutions. If $\boldsymbol{x}_3$ is worse than both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, then it is concluded that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ belong to different basins. It is worth noting that the hill-valley heuristic is not always accurate and may result in a false negative or a true positive outcome.

The newly detected global minima are added to $\mathbb{A}$. Then, the normalized taboo distances of the archive solutions are updated based on the number of subpopulations that have converged to them. Given the updated $\mathbb{A}$, a new restart with an updated population size is performed, and this process continues until the evaluation budget is exhausted. The flowchart of RS-CMSA-ES is provided Fig. 2.

RS-CMSA-ES emerged as the most successful method on the CEC'2013 test suite for multimodal optimization in 2016. The method, however, suffers from some shortcomings:

- The adaptation of normalized taboo distances stalls if the subpopulations do not converge to global minima.
- The adaptation of the covariance matrix involving elite solutions favors search directions that have been successful in the past. However, these directions may not be the best ones after some iterations since the center of each subpopulation changes over time.
- The subpopulations that converge to undesirable local minima or already identified global minima waste a lot of the evaluation budget.
- The estimate for the criticality of a taboo region highly overestimates the actual value which may result in unnecessarily high time-complexity.
- The treatment for bound-violating solutions is primitive: it always prefers feasible solutions to infeasible ones, making it difficult to find near-the-bound global minima (if any).
- The subpopulation initialization process is unnecessarily time consuming.

RS-CMSA-ESII aims to address these drawbacks of RS-CMSA-ES.

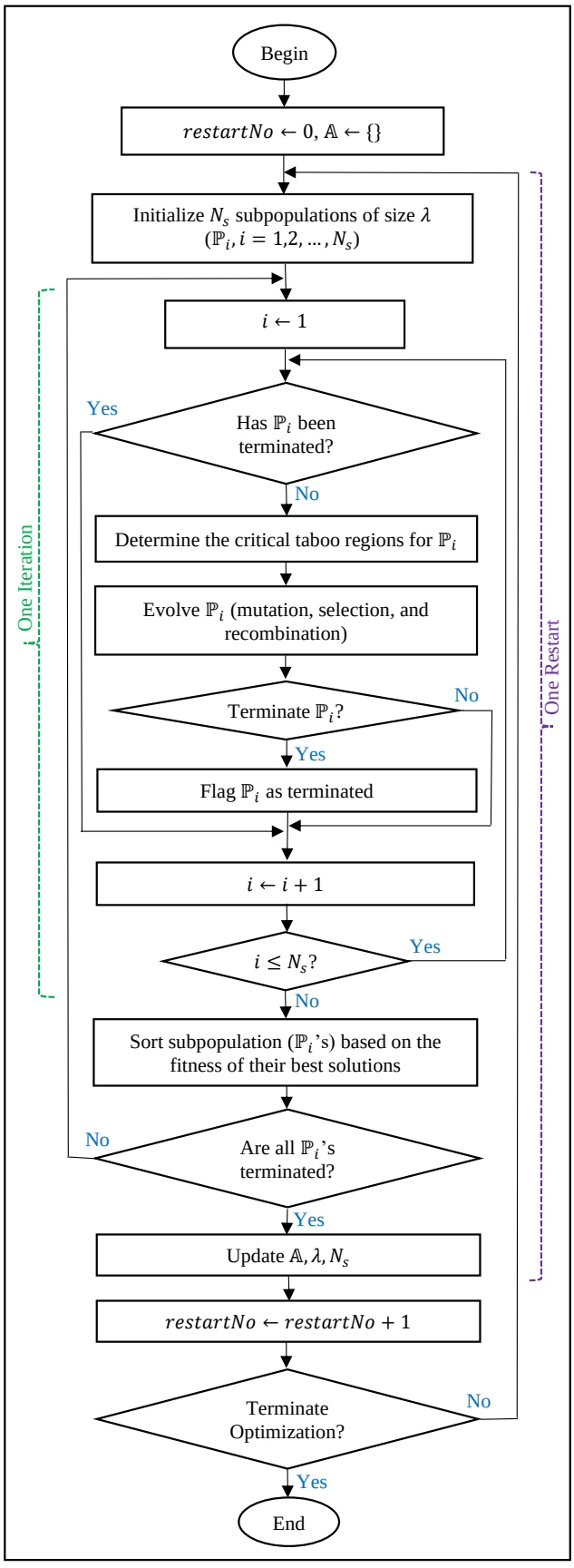


Fig. 2. Flowchart of RS-CMSA-ES

III. RS-CMSA-II

The general structure of RC-CMSA-ESII is similar to that of RS-CMSA-ES. However, the new variant enjoys several improvements in its components. These improvements are discussed in this section. Following the findings in [50], the number of subpopulations in RS-CMSA-ESII is one ($N_\mathrm{s} = 1$). This means that the whole population consists of one subpopulation, which is denoted by $\mathbb{P}$. This setting simplifies the search operators and improves the exploitation of information. Besides, it means the taboo regions are defined only by the archived solutions.

A. Update of the Normalized Taboo Distances

When updating $\hat{d}_{A_m}$'s, RS-CMSA ignores the subpopulations that failed in converging to a global minimum (new or already identified). If this is the case with most subpopulations, it slows down the adaptation of $\hat{d}_{A_m}$'s. The new updating scheme has addressed this issue. Furthermore, it follows a completely distinct idea. Based on the analysis of $\boldsymbol{x}_\mathrm{best}$ using the hill-valley strategy [43], three cases may happen:

- Case I: $\mathbb{P}$ has converged to a new global minimum.
- Case II: $\mathbb{P}$ has converged to the archived solution $\boldsymbol{x}_{A_{\overline{m}}}$.
- Case III: $\mathbb{P}$ has not converged or has converged to an undesirable minimum.

For each case, the proper action is performed. Case I is the most desirable case; therefore, RS-CMSA-ESII does not change the normalized taboo distances. It simply appends $\boldsymbol{x}_\mathrm{best}$ to $\mathbb{A}$ with the normalized taboo distance of $\hat{d}_\mathrm{def}$.

Case II indicates that $\hat{d}_{A_{\overline{m}}}$ was not large enough to keep $\mathbb{P}$ away from $\boldsymbol{x}_{A_{\overline{m}}}$. Thus, RS-CMSA-ESII increases $\hat{d}_{A_{\overline{m}}}$ and slightly reduces the normalized taboo distances of other archived solutions.

$$\hat{d}_{A_m} \leftarrow \begin{cases} \hat{d}_{A_m} \exp\left(\tau_{\hat{d}}\right) & \text{if } m = \overline{m} \\ \hat{d}_{A_m} \exp\left(-\dfrac{\tau_{\hat{d}}\,(1-\alpha_\mathrm{new})}{|\mathbb{A}|-1}\right) & \text{if } m \neq \overline{m} \text{ and } |\mathbb{A}|>1 \\ m = 1, 2, \ldots, |\mathbb{A}|\,. \end{cases} \tag{5}$$

In this equation, $0 \leq \tau_{\hat{d}}$ is the learning rate for the adaptation of the normalized taboo distances, and $0 \leq \alpha_\mathrm{new} \leq 1$ determines the expected likelihood that $\boldsymbol{x}_\mathrm{best}$ is a new global minimum if it is a global one. A smaller α_new means a greater reduction in $\hat{d}_{A_{m\neq\overline{m}}}$. This reduction is important since the normalized taboo distances of some archived solutions might be unnecessarily large. The default values of $\alpha_\mathrm{new} = 0.5$ and $\tau_{\hat{d}} = 1/\sqrt{D}$ are recommended following the settings for RS-CMSA-ES [31]. This setting means that if $\boldsymbol{x}_\mathrm{best}$ is a global minimum, it should be a new one 50% of the times. If this ratio is less than 50%, the geometric mean of $\hat{d}_{A_m}$'s will increase to further push future populations away from already detected global minima.

Case III implies that the normalized taboo distances of all archived solutions might have been unnecessarily large. Therefore, RS-CMSA-ESII slightly reduces $\hat{d}_{A_m}$'s:

$$\hat{d}_{A_m} \leftarrow \hat{d}_{A_m} \exp\left(-\frac{\tau_{\hat{d}}\,\alpha_\mathrm{global}}{|\mathbb{A}|}\right), m = 1, 2, \ldots, |\mathbb{A}|\,. \tag{6}$$

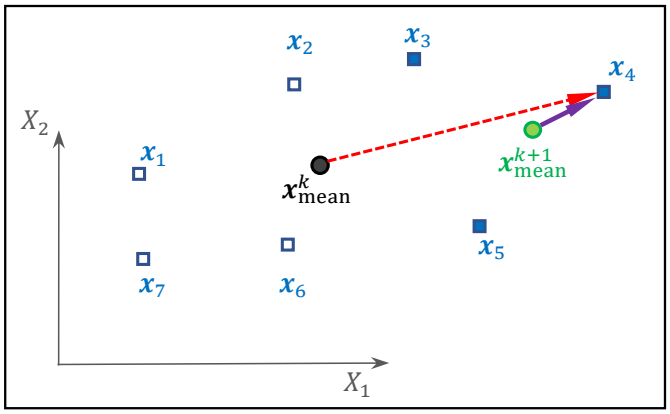


Fig. 3. Update of the suggested directions ($\boldsymbol{s}_j$'s in (7)) for adaptation of the covariance matrix.

In this equation, $0 < \alpha_\mathrm{global}$ controls the reduction amount. A greater α_global results in a higher chance of converging to a global minimum; however, it is more likely that this global minimum is an already identified one. Based on some preliminary numerical simulations, the default value of $\alpha_\mathrm{global} = 0.5$ is recommended.

B. Elitism and Adaptation of the Covariance Matrix

Elitism is not a part of CMSA-ES [32]; however, it is beneficial for MMO. The elite selection of RS-CMSA-ES is particularly useful in low dimensional problems. RS-CMSA-ESII follows the same formulation for updating $\boldsymbol{x}_\mathrm{mean}$ and σ_mean when performing elite selection; however, a new strategy to update $\mathbf{C}$ is proposed in this study because of the highlighted drawbacks of the corresponding scheme in RS-CMSA-ES (see Section II). The adaptation mechanism in RS-CMSA-ESII, unlike the corresponding one in RS-CMSA-ES, takes the iterative changes in the population center into account. After the calculation of the new $\boldsymbol{x}_\mathrm{mean}$, this strategy first updates the vector $\boldsymbol{s}_j$ of the survived elite solutions from previous iterations:

$$\boldsymbol{s}_j \leftarrow \frac{\boldsymbol{x}_j - \boldsymbol{x}_\mathrm{mean}}{\sigma_j},\ j = 1+\lambda, 2+\lambda, \ldots, N_\mathrm{elt}+\lambda. \tag{7}$$

This update encourages searching along the line connecting $\boldsymbol{x}_j$ to the new $\boldsymbol{x}_\mathrm{mean}$. Then, this vector is used to update the covariance matrix according to (2).

An exemplary case is depicted in Fig. 3. The subpopulation with its center at $\boldsymbol{x}^k_\mathrm{mean}$ has generated seven sample offspring (squares) at iteration $\#k$, three of which are selected for recombination (solid squares) and update of the subpopulation center ($\boldsymbol{x}^{k+1}_\mathrm{mean}$). Let us assume $\boldsymbol{x}_4$ is an elite solution. In RS-CMSA-ES, $\boldsymbol{s}_4$ is $(\boldsymbol{x}_4 - \boldsymbol{x}^k_\mathrm{mean})$ (red dashed vector), whereas in RS-CMSA-ESII, it is $\boldsymbol{x}_4 - \boldsymbol{x}^{k+1}_\mathrm{mean}$ (solid purple vector).

C. New Termination Criteria

In RS-CMSA-ES, a subpopulation is considered as converged if the change in the value of the best non-elite solution in the last $tolHistSize = 10 + \lfloor 30D/\lambda \rfloor$ iterations is less than $tolHistFun$, the value of which depends on the target tolerance

of the objective function (ϵ_f). RS-CMSA-ESII employs two additional termination criteria that identify and terminate a subpopulation which is unlikely to converge to a new global minimum in the future. If effectively formulated, they can substantially improve the efficiency of the search. These two termination criteria follow the same goals as the additional stopping criteria introduced by Maree et al. [44], but they use different formulations. The first stopping criterion, the merge operator, aims to detect convergence to a solution in $\mathbb{A}$. The second stopping criterion predicts if the population may converge to a global minimum at all.

1) Merge Operator: Although archived solutions repel $\mathbb{P}$, the normalized taboo distance of an archived solution might be small in comparison with the attraction region of that global minimum. In such a case, $\mathbb{P}$ may converge to that minimum; however, the convergence rate would be low because the taboo region that circumscribes the global minimum repels $\mathbb{P}$. Convergence to an already identified global minimum may be useful to update (in this case, to increase) the normalized taboo distance of that archived solution for the future restarts. However, since the convergence can be very slow, it will deplete the evaluation budget. To overcome this issue, RS-CMSA-ESII utilized a merge operator which aims to predict if $\mathbb{P}$ is converging to an archived solution, and if so, it identifies the corresponding archived solution.

The merge operator exploits the hill-valley strategy to check if $\boldsymbol{x}_{\mathrm{best}}$ shares the same basin with a solution in $\mathbb{A}$. There are two main challenges when employing this strategy:

- The hill-valley strategy requires additional function evaluations whenever it is executed. Considering that the population evolves, $\boldsymbol{x}_{\mathrm{best}}$ should be checked against the solutions in $\mathbb{A}$ frequently. This process itself will consume a lot of function evaluations, reducing the benefits that can be gained from the merge operator.
- It is possible that $\boldsymbol{x}_{\mathrm{best}}$ is in the basin of an archived solution; however, it will leave that basin in the future if it is allowed to evolve. It can happen especially at early iterations of a restart when the mutation strength is high and $\boldsymbol{x}_{\mathrm{best}}$ can significantly change.

To overcome these challenges, a mergeability indicator is defined which quantifies how likely $\mathbb{P}$ is converging to the archived solution $\boldsymbol{x}_{A_m}$:

$$\mathcal{M}(\boldsymbol{x}_{A_m}, \mathbb{P}) = \frac{1 + \hat{d}_{A_m}}{\hat{L}_m}, \quad (8)$$
$$\hat{L}_m = \sqrt{(\boldsymbol{x}_{A_m} - \boldsymbol{x}_{\mathrm{mean}})^{\mathrm{T}} (\sigma^2_{\mathrm{mean}} \mathbf{C})^{-1} (\boldsymbol{x}_{A_m} - \boldsymbol{x}_{\mathrm{mean}})}$$

in which $\mathcal{M}(\boldsymbol{x}_{A_m}, \mathbb{P})$ indicates the mergeability of $\mathbb{P}$ into $\boldsymbol{x}_{A_m}$, and $\hat{L}_m$ is the Mahalanobis distance between $\boldsymbol{x}_{A_m}$ and $\boldsymbol{x}_{\mathrm{mean}}$. Archived solutions with a mergeability greater than the mergeability threshold T_{merge} are flagged as potential candidates, in which T_{merge} is a control parameter set by the user. The hill-valley test is then executed for $\boldsymbol{x}_{A_m}$ if it is the only potential candidate for merge in the last $0.1 tolHistSize$ iterations. If the test outcome is positive, $\mathbb{P}$ is predicted to converge to $\boldsymbol{x}_{A_m}$. Therefore, $\mathbb{P}$ is terminated and $\boldsymbol{x}_{\mathrm{best}} \leftarrow \boldsymbol{x}_{A_m}$; otherwise this test will not be executed for the next $0.1 tolHistSize$ iterations.

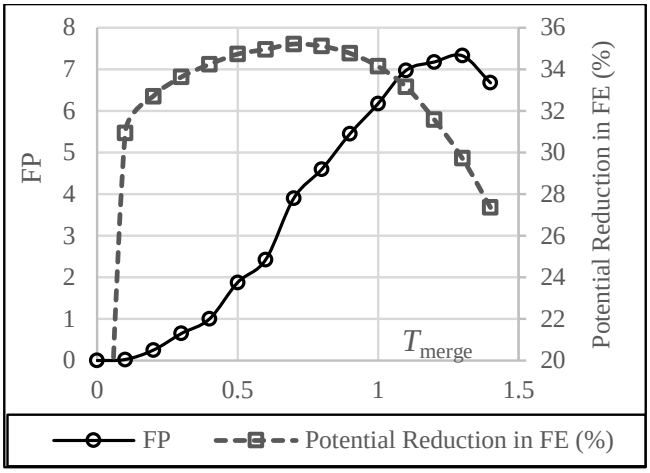


Fig. 4. Trade-off between the potential reduction in the used FEs and false positive (FP) predictions by the merge operator for different values of the mergeability threshold (T_{merge})

Like any other binary classification test, the merge operator may make false or true predictions. The value of T_{merge} should be selected such that the number of false positives (FP) is minimized while the number of function evaluations saved by the merge operator is maximized. This can be achieved by a proper selection of T_{merge}. A simple simulation is designed in this study to find such a value. In this simulation, the function evaluation at which the merge operator has flagged $\mathbb{P}$ for termination is recorded; however, $\mathbb{P}$ is not terminated because of the merge operator. At the end of the restart, the potential savings in the number of function evaluations (FE_{save}) by the merge operator is calculated, which is the function evaluation at which $\mathbb{P}$ was terminated minus the one at which $\mathbb{P}$ was flagged for merge. The overall potential reduction in the consumed function evaluations (FE_{SAVE}) is the sum of FE_{save} in all restarts.

For this simulation, 3D Shubert and 3D Vincent problems [3] are considered because of their unique features: The Shubert function has 81 irregularly distributed global minima of similar basin sizes and many local minima. In contrast, the Vincent function has no local minima; however, the basins vastly differ in the shape and size. The evaluation budget (FE_{max}) of each problem is 400,000.

Fig. 4 shows $FE_{\mathrm{SAVE}}/FE_{\mathrm{max}}$ and the number of false positives for different values of T_{merge}, averaged over 20 independent runs and two problems. As observed, the number of FP monotonously increases with T_{merge}. FE_{SAVE} increases with T_{merge} up to $T_{\mathrm{merge}} = 0.7$. Therefore, T_{merge} must be less than 0.7. The suggested value is $T_{\mathrm{merge}} = 0.5$.

2) Local Convergence Predictor: A subpopulation that seems to be converging to an undesirable local minimum can be identified before convergence when performing MMO. The reason for this is that there is the "best-value-so-far" from previous restarts (f_A^{min}). It is possible to predict if $\mathbb{P}$ is going to converge to an equally good or even a better solution in the future. For example, Maree et al. [44] assumed a "convergence rate" and compared the actual convergence rate of the subpopulation with the expected convergence rate. If the actual

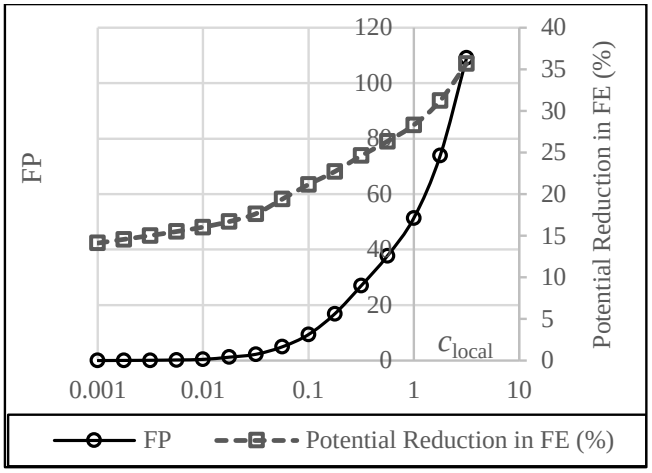


Fig. 5. Trade-off between the potential reduction in the adopted FEs and the number of FP by the local convergence predictor for different values of c_{local}.

convergence rate is not high enough, then the subpopulation is predicted to converge to a local minimum. RS-CMSA-ESII follows the same goal but uses a different method since the presence of taboo regions and elitism may hinder the progress rate. The employed formulation calculates the average fluctuations in the value of $\boldsymbol{x}_{\text{best}}$ in the last $0.5 tolHistSize$ iterations. If it is less than $c_{\text{local}}\left(f(\boldsymbol{x}_{\text{best}}) - tolHistFun - f_A^{\min}\right)$, it concludes that $\mathbb{P}$ might not converge to a global minimum. c_{local} is a control parameter defined by the user. A greater c_{local} terminates $\mathbb{P}$ earlier, which results in a greater FE_{SAVE}; however, it increases the number of FP.

A simulation similar to the one proposed earlier is performed to provide insights into the effect of c_{local}. The local convergence predictor may flag $\mathbb{P}$ for termination, but it does not terminate it. FE_{SAVE} from this termination criterion and the number of FP are then calculated for different values of c_{local}. For this simulation, the merge operator has been deactivated.

Fig. 5 shows the effect of c_{local} on FE_{SAVE} and the average number of FP for 3D Shubert and 3D Vincent functions, when $FE_{\max} = 400000$. Both the number of FP and FE_{SAVE} monotonously increase with c_{local}; however, the number of FP increases at a faster rate. The default value of $c_{\text{local}} = 0.04$ is suggested based on this graph and a few other numerical studies.

D. Bound Handling

At early iterations of a restart, it is common that some sampled solutions fall outside the search range. RS-CMSA-ES sorts infeasible solutions based on their constraint violations while feasible solutions are always preferred over infeasible ones. A repair operator is used in RS-CMSA-ESII to improve its near-bound search capability. This operator relocates the variable that has violated the bound constraint to a random point on a line segment whose middle point lies on $\boldsymbol{x}_{\text{mean}}$. This process is illustrated schematically in Fig. 6.

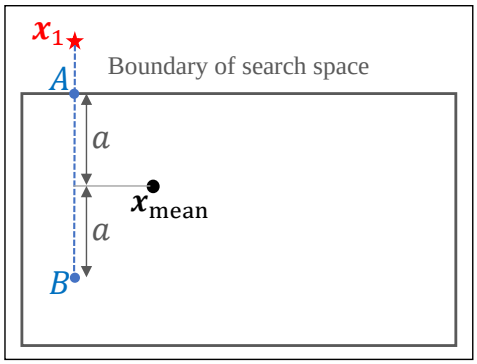


Fig. 6. Repair of a bound-violating sample ($\boldsymbol{x}_1$). The repaired solution is randomly generated on the line segment *AB*.

E. Initialization of the population

A shortcoming of the initialization process of RS-CMSA-ES is that it initiates with a conservatively large $\sigma_{\text{mean}}^{\text{ini}}$ at the beginning of each restart. This value gradually decreases when multiple successive attempts to generate $\boldsymbol{x}_{\text{mean}}$ fail; however, this process may take an unnecessarily long time even though the initialization process does not use any function evaluations. The improved initialization process addresses this issue. For the first restart, $\sigma_{\text{mean}}^{\text{ini}}$ is set to a conservatively large value ($\sigma_{\text{mean}}^{\text{ini}} = \sqrt{D}$). This value may require reduction so that $\boldsymbol{x}_{\text{mean}}$ can be generated. The value of $\sigma_{\text{mean}}^{\text{ini}}$ at which $\boldsymbol{x}_{\text{mean}}$ is successfully generated is preserved ($\sigma_{\text{mean}}^{\text{ini-s}}$). Since the update of taboo regions is gradual, it is unlikely that in the next restart, an acceptable $\boldsymbol{x}_{\text{mean}}$ may be generated if $\sigma_{\text{mean}}^{\text{ini}}$ is much greater than $\sigma_{\text{mean}}^{\text{ini-s}}$. Therefore, for the next restart, $\sigma_{\text{mean}}^{\text{ini}}$ initiates with a slightly greater value than $\sigma_{\text{mean}}^{\text{ini-s}}$ ($\sigma_{\text{mean}}^{\text{ini}} \leftarrow 1.04\,\sigma_{\text{mean}}^{\text{ini-s}}$). $\sigma_{\text{mean}}^{\text{ini}}$ is likely to gradually decrease until an acceptable point for $\boldsymbol{x}_{\text{mean}}$ is generated and selected. Then, the mutation profile of the population is defined as follows:

$$\begin{aligned} \sigma_{\text{mean}} &\leftarrow \min\left\{2\sigma_{\text{mean}}^{\text{ini}}, 0.3\right\} \\ \mathbf{C} &\leftarrow \left[\boldsymbol{x}^{\text{U}} - \boldsymbol{x}^{\text{L}}\right]\left[\boldsymbol{x}^{\text{U}} - \boldsymbol{x}^{\text{L}}\right]^T. \end{aligned} \tag{9}$$

This strategy can substantially reduce the time required for initialization. This is especially important for RS-CMSA-ESII since there is only one subpopulation at each restart, and thus, many restarts will be performed until the end of the optimization process.

F. Calculation of criticality of a taboo region

The time required for sampling new solutions strongly depends on the number of critical taboo points. One problem with RS-CMSA-ES is that the estimated value for P_{trej}, which determines which taboo points are critical, can be much greater than the actual one. RS-CMSA-ESII overcomes this shortcoming by providing a more accurate estimate of P_{trej} by utilizing the properties of the Mahalanobis distance metric to rescale the search space according to the mutation profile of the population, as shown in Fig. 7.

Fig. 7a illustrates the population $\mathbb{P}$ in the proximity of the taboo point $\boldsymbol{x}_A$ with the normalized taboo distance of $\hat{d}_A$. Let

X_1X_2 be the global coordinate system of the search space and x_1x_2 the auxiliary coordinate system along the main axes of the covariance matrix. The triangles represent some sampled solutions, one of which is not taboo acceptable (the solid triangle). Fig. 7b replots Fig. 7a when the space has been scaled along x_1 and x_2 with the scaled factor of $1/\left(\sigma_{\text{mean}}\, u_1\right)$ and $1/\left(\sigma_{\text{mean}}\, u_2\right)$, in which u_1 and u_2 are the square root of the eigenvalues of $\mathbf{C}$. This scaling does not change the acceptability of the sampled solutions. The distribution of samples in the rescaled space (x_1x_2) follows an isotropic normal distribution with standard deviation of one, and the taboo region is now a circle of radius $\hat{d}_A$.

From all the points on the line segment AC, point B has the highest probability density function. Using the value of the density function at this point for all the points on the line passing through AC, an upper limit for the probability that the sampled solution falls inside the taboo region can be easily calculated as follows:

$$P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A) = \int_{\hat{L}-\hat{d}_A}^{\hat{L}+\hat{d}_A} \varphi(x)dx,\ \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \tag{10}$$

in which $\varphi(x)$ is the probability density of the standard normal distribution function. Like before, taboo points with $P_{\text{trej}} > 0.01$ are considered critical.

In order to compare this new estimate for P_{trej} with the existing one, a simple numerical simulation is performed. In this simulation, the subpopulation $\mathbb{P}$ has its center on the origin ($\boldsymbol{x}_{\text{mean}} = \mathbf{0},\ \sigma_{\text{mean}} = 1$) and $\mathbf{C}$ is a diagonal matrix. The taboo region is $\boldsymbol{x}_A = r_0\mathbf{1}$ with $\hat{d}_A = 2$. $P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A)$ is then calculated using three methods:

- the existing estimation method adopted in RS-CMSA-ES [31]
- the new estimation method according to (10)
- a Monte Carlo simulation with 10000 sampled solutions according to the defined mutation profile.

Fig. 8 shows $P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A)$ calculated using these three methods for different values of r_0 when $\mathbf{C}$ has a condition number of 10 (Fig. 8a) and 10000 (Fig. 8b). The 95% confidence interval for the Monte Carlo simulation was too narrow to be shown on the graph; therefore, it can be perceived as a relatively accurate estimate of the true P_{trej}. This figure shows that:

- Both estimation methods overestimate the true $P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A)$; however, the new one is more accurate.
- The superiority of the new estimation method over the existing one intensifies when the condition number of $\mathbf{C}$ is high (Fig. 8b). In this case, the new estimate still provides a good approximation, whereas the existing one highly overestimates $P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A)$.

G. An Illustrative Example

The capabilities of RS-CMSA-ESII and its ability to learn the basin sizes of global minima are explored by analyzing its behavior when optimizing the 2D Vincent function (Fig. 9a). This function has 36 global minima with diverse sizes and relative distances, which challenge the niching capability of an MMO method. Fig. 9 illustrates snapshots of RS-CMSA-ESII immediately after reinitialization of the subpopulation at different restarts, including the taboo regions (thin red circles), the initialized subpopulation (thick black circle), and the global minima of the problem (blue stars). The number of detected global minima (NDGM) is also provided. It can be observed that:

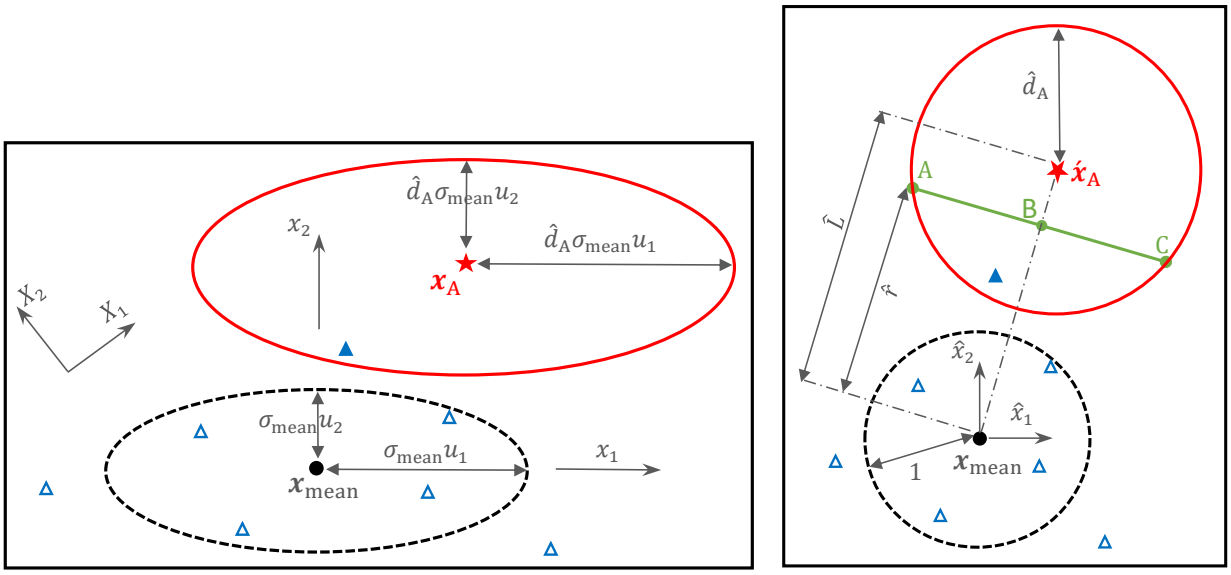
- By restart #7, seven global minima have been detected, indicating that the subpopulation in each previous restart has converged to a new global minimum. RS-CMSA-ESII has defined a taboo region around each detected global minimum. The normalized taboo distances of all these taboo regions are equal at this stage.
- At restart #22, some taboo points have larger taboo regions. More importantly, there is a noticeable correlation between the size of the global minimum and the radius of the corresponding taboo region. This means RS-CMSA-ESII is learning the relative sizes of the basins of the detected global minima.
- From restart #22 to restart #84, RS-CMSA-ESII detects new global minima in some restarts. In other restarts, it adapts the size of the existing taboo regions according to the adaptation mechanism explained in Subsection III-A. A noticeable (but not perfect) correlation between the size of the basins of the global minima and the corresponding taboo regions can always be detected.
- In this specific run, RS-CMSA-ESII detected all 36 global minima by the beginning of restart #96. At this point, it has used only 27,827 function evaluations, about 14% of the evaluation budget defined for this problem in the CEC'2013 test problems for MMO [3].

It is remarkable that the shape and sizes of the taboo regions change during a restart since they are coupled to the mutation profile of the subpopulation. The initial mutation profile is determined by the search space ranges. For this problem, the search space is $[0.25, 10]^2$; therefore, the initial mutation profile is Isotropic, resulting in circular taboo regions at the beginning of each restart.

IV. Modular Analysis

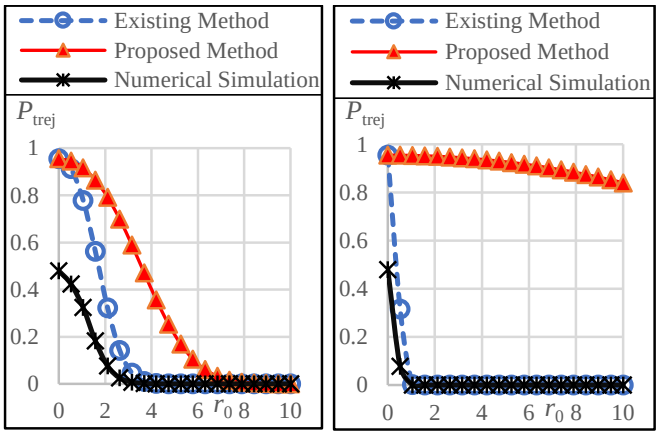
This section runs a controlled experiment to study the impact of the modifications made to RS-CMSA-ESII on its performance. In addition to RS-CMSA-ESII, five other variants of this method are considered. In each variant, one component of this method that was improved or developed in this study is suppressed or replaced by the corresponding component in RS-CMSA-ES. These variants are:

- RS-CMSA-ESII\{Archive}: The proposed method for the adaptation of the normalized taboo distances is replaced by the old adaptation method in RS-CMSA.
- RS-CMSA-ESII\{Elite}: The proposed method for the adaptation of $\mathbf{C}$ involving elite solutions is replaced with the corresponding method in RS-CMSA-ES.
- RS-CMSA-ESII\{Repair}: The proposed repair strategy is suppressed. Instead, the strategy of RS-CMSA-ES for treating bound-violating solutions is employed.



(a) (b)

Fig. 7. Providing an upper estimate for the criticality of a taboo region for population $\mathbb{P}$. A sample solution is taboo acceptable if it is outside the taboo region. a) The taboo region, population $\mathbb{P}$ and some sampled solution (triangles). The solid triangle is an unacceptable solution and will be discarded. The search space coordinate system is X_1X_2. x_1x_2 is an auxiliary coordinate system, which aligns with the eigenvectors of the covariance matrix. b) The space has been re-scaled along $\hat{x}_1$ and $\hat{x}_2$ with the scaled factor of $1/(\sigma_{\text{mean}}u_1)$ and $1/(\sigma_{\text{mean}}u_2)$.



(a) cond $(\mathbf{C}) = 10$ (b) cond $(\mathbf{C}) = 10000$

Fig. 8. Estimation of $P_{\text{trej}}(\boldsymbol{x}_A, \hat{d}_A)$ in 5-D space ($\boldsymbol{x}_A = r_0\mathbf{1}, \hat{d}_A = 2$) for the population $\mathbb{P}$ with $\boldsymbol{x}_{\text{mean}} = \mathbf{0}$ and $\sigma_{\text{mean}} = 1$ as a function of r_0 using three estimation methods. The condition number of $\mathbf{C}$ is a) 10 and b) 10000.

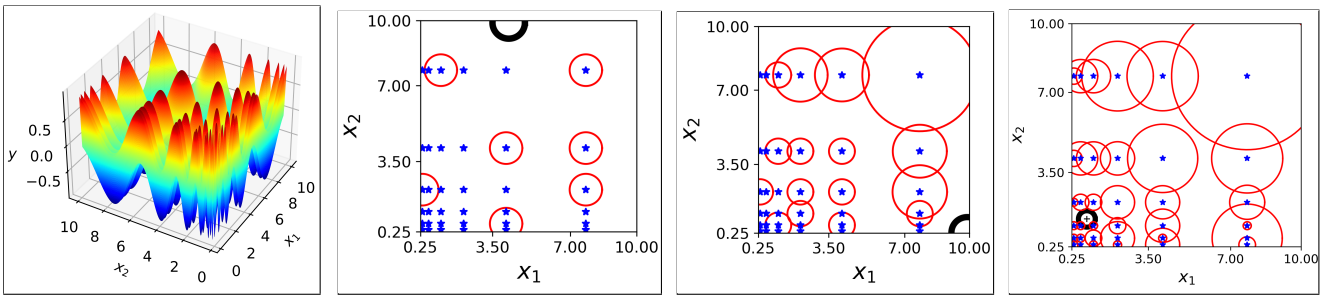
- RS-CMSA-ESII\{Merge}: The formulated merge operator is suppressed.
- RS-CMSA-ESII\{Local}: The formulated local convergence predictor is suppressed.

Except for the modified component, all other components of these variants are identical to those of RS-CMSA-ESII. The control parameters are set to their default values and those recommended in RS-CMSA-ES [31], except for the following control parameters:

- Number of subpopulations (N_s) = 1
- Population Size (λ) = $6\sqrt{D}$ (fixed)
- Learning rate for the step size ($\tau_\sigma = \frac{1}{2\sqrt{D}}$)
- Parents size (μ) = $\max\{1, \lfloor 0.2\lambda + .5 \rfloor\}$
- Number of elite solutions $\lceil 0.1\lambda \rceil$
- $tolHistFun = 10^{-6}$

The well-known peak ratio (PR) performance indicator [3] is employed to compare the performance of these variants with RS-CMSA-ESII. Twelve benchmark problems from the CEC'2013 test suite for MMO [3] are considered for this purpose. The other eight problems were too easy to distinguish the pros and cons of each variant. Each experiment was repeated 50 times independently using a single core and 1GB of allocated memory. Table I presents the mean and standard error of the calculated PR for each problem and each variant of RS-CMSA-ESII. The problems in which the studied component has provided a practically meaningful difference are highlighted in boldface. The obtained results reveal that:

- Each new or modified component of RS-CMSA-ESII provides a detectable benefit to RS-CMSA-ESII when the overall performance is considered.
- When each problem is considered, the updated component is either beneficial or at least does not deteriorate the performance. The only exception is the introduced repair operator which turned out to be detrimental for PID=14.
- The studied components improve or do not change the time complexity of the method, except for the new elite strategy. However, considering the number of function evaluations for all 12 problems ($9 \times 400,000 + 3 \times 200,000 = 4,200,000$), the time per evaluation is in the range of 0.5-1.5 ms (including the time required for a function evaluation). Therefore, the required time by the optimization operators is negligible considering that in most practical problems, each function evaluation may take from a few seconds to a few hours.



(a) 2D Landscape (b) Restart #7, NDGM=7 (c) Restart #22, NDGM=16 (d) Restart #84, NDGM=35

Fig. 9. Response of RS-CMSA-ESII when optimizing the 2D Vincent function. a) Fitness landscape of the problem. b-d) Taboo regions (red thin circles) and the initialized subpopulation (black thick circle) at different restarts, immediately after the initialization of the subpopulation. The number of detected global minima (NDGM) is reported for each case. Global minima are shown by blue stars. Taboo regions are shown with a scale factor of 0.25 for better visualization. The center of the subpopulation is the center of the black circle, the radius of which is one-fourth of the mutation strength of the initialized subpopulation.

TABLE I

PEAK RATIO (MEAN ± SE) OF DIFFERENT VARIANTS OF RS-CMSA-ESII ON 12 TEST PROBLEMS OF THE CEC'2013 COMPETITION ON NICHING METHODS FOR MMO [3]. THE NUMBERS IN BOLDFACE INDICATE A MEANINGFUL PERFORMANCE DIFFERENCE. THE REPORTED TIME IS THE AVERAGE TIME TO RUN ALL 12 PROBLEMS

PID	RS-CMSA-ESII	RS-CMSA-ESII \{Archive}	RS-CMSA-ESII \{Elite}	RS-CMSA-ESII \{Repair}	RS-CMSA-ESII \{Merge}	RS-CMSA-ESII \{Local}
8	0.997 ± 0.001	0.996 ± 0.001	0.984 ± 0.002	0.999 ± 0.000	**0.982 ± 0.002**	**0.955 ± 0.004**
9	0.990 ± 0.001	**0.965 ± 0.002**	**0.985 ± 0.001**	0.987 ± 0.001	**0.899 ± 0.002**	0.990 ± 0.001
11	1.000 ± 0.000	**0.977 ± 0.008**	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
12	1.000 ± 0.000	0.990 ± 0.005	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
13	0.993 ± 0.005	**0.870 ± 0.010**	**0.903 ± 0.012**	1.000 ± 0.000	**0.963 ± 0.010**	**0.930 ± 0.012**
14	0.850 ± 0.007	0.837 ± 0.003	0.833 ± 0.000	**0.883 ± 0.011**	0.837 ± 0.003	0.840 ± 0.005
15	0.750 ± 0.000	0.750 ± 0.000	0.750 ± 0.000	0.743 ± 0.005	0.750 ± 0.000	0.750 ± 0.000
16	0.833 ± 0.000	**0.773 ± 0.011**	0.820 ± 0.006	**0.787 ± 0.011**	**0.793 ± 0.010**	**0.790 ± 0.010**
17	0.750 ± 0.000	0.750 ± 0.000	0.750 ± 0.000	0.738 ± 0.009	0.738 ± 0.005	**0.735 ± 0.006**
18	0.667 ± 0.000	0.667 ± 0.000	0.667 ± 0.000	0.637 ± 0.009	0.667 ± 0.000	0.667 ± 0.000
19	0.703 ± 0.009	**0.667 ± 0.008**	0.705 ± 0.008	**0.578 ± 0.018**	**0.505 ± 0.005**	0.667 ± 0.008
20	0.618 ± 0.004	**0.593 ± 0.008**	0.615 ± 0.005	**0.410 ± 0.013**	**0.365 ± 0.015**	0.603 ± 0.008
Avg.	0.846 ± 0.001	0.820 ± 0.002	0.834 ± 0.001	0.813 ± 0.003	0.792 ± 0.002	0.827 ± 0.002
Time (min)	51.9	80.4	37.9	59.1	51.5	102.0

V. COMPARISON WITH OTHER MMO METHODS

This section compares the performance of RS-CMSA-ESII with the most successful niching methods for MMO. The test suite and experimental setup of the CEC'2013 competition on niching methods for MMO [3] are followed for this purpose. This test suite has been employed in similar events in the subsequent years. Besides, many researchers have adapted this test suite and reported their results on its problems, which facilitates a fair performance evaluation and comparison. The selected methods for comparison are:

- The winners of this competition in previous years, which are NEA2 [35], NMMSO [41], RS-CMSA-EA [31], HillVallEA18 [44], and HillVallEA19 [51]. These methods were the winners of this competition in 2013, 2015, 2016, 2018, and 2019, respectively.
- MMO methods that have been recently published in top journals and reported their results on this test suite, including ANDE [36], LBPADE [20], DIDE [21], FBK-DE [34], MMDE [40], TS-ABC [52], and NCjDE- 2LS$_{ar}$ [38]. The results of these methods have been directly excerpted from the corresponding publications.

RS-CMSA-ESII is employed for MMO of all 20 problems of the CEC'2013 test suite for MMO [3]. Each problem was optimized 50 times independently with random seeds $0, 1, \ldots, 49$, and PR for each problem was calculated and averaged for three function tolerances ($\varepsilon_f = 10^{-5}, 10^{-4}, 10^{-3}$). These function tolerances were selected since many of these methods have reported results for them. Besides, a looser function tolerance may calculate a PR greater than the actual one for the problems with basins of different sizes (e.g., the 3D Vincent function). For TS-ABC [52], only the results for $\epsilon_f = 10^{-4}$ have been reported in the corresponding publication; therefore, for this method, the reported PR represents the averaged PR over three function tolerances. For NCjDE-2LS$_{ar}$ [38], PR has been reported for $\epsilon_f = 10^{-3}$ and $\epsilon_f = 10^{-5}$ only; therefore, for this method, the calculated PR is the average of the reported PR for these two tolerances.

Table II presents the calculated PR for each problem and each method. For RS-CMSA-ESII, the standard error of the calculated PR is also provided. Furthermore, the last column shows the standard deviation of PR of the tested methods for each problem. For RS-CMSA-ESII, the calculated PR for each independent run has been provided in the Supplementary Material **S1**. For this method, the individual PR values were identical for $\epsilon_f = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$. The obtained results reveal that:

- When MPR is considered, RS-CMSA-ESII outperforms all the tested methods, including HillVallEA19, which

was the winner of the competition on MMO in 2019. The standard error for the MPR of RS-CMSA-ESII is small. Assuming that the standard error of other methods is similar to that of RS-CMSA-ESII, any difference in the mean peak ratio (MPR) greater than 0.002 is statistically significant. When comparing RS-CMSA-ESII with HillVallEA19 (given the reported results in the GECCO'2019 competition), the Friedman test returns a p-value of 3.26×10^{-40}.

- The first five and the tenth problems are easy, and almost all methods can reach a peak ratio close to one. These problems have at most two variables. Such easy problems can hardly be used to differentiate between superior and inferior methods.
- Problems 6 and 8 are the 2-D and 3-D Shubert function with many local minima, although the global minima have similar sizes. For this problem, NCjDE-2LS$_{\text{ar}}$ and RS-CMSA-ESII obtained the best results, followed by HillVallEA19.
- When the basins significantly vary in their sizes and relative distribution (PID = 7, 9), RS-CMSA-ESII outperforms all other methods, thanks to the additional termination criteria and the ability to learn the relative sizes of global minima. The closest competitors of RS-CMSA-ESII in these two problems are HillVallEA19 and NMMSO.
- For problems 11 and 12, some of these methods can reach a peak ratio close to one. These 2-D composite functions have complex definitions and the resultant landscape does not have the simplicity of the previous problems.
- Problems 13, 14, 16, and 18 are based on the same composite function (CF3) [3] with six global minima when $D = 2, 3, 5, 10$, respectively. In low dimensions (PID = 13, 14), there are a few methods that outperform RS-CMSA-ESII (e.g., FBK-DE, HillVallEA18 and HillVallEA19). However, in higher dimensions, RS-CMSA-ESII outperforms these methods. This indicates that the performance of RS-CMSA-ESII scales better with the problem's dimensionality.
- The behavior of LBPADE on CF3 is odd since it provides a higher PR for the 10-D problem (PID = 18) than the 5-D problem (PID = 16). Furthermore, no other method could reach a PR greater than 0.667 for PID = 18. More importantly, analyzing the reported global minima obtained by HillVallEA19, HillVallEA2018, and RS-CMSA-ESII reveals that all these methods could find only four specific global minima for PID = 18 when all their runs are considered, resulting in $PR = 4/6 = 0.667$. Besides, the higher PR of LBPADE is only observed when $\varepsilon_f = 10^{-3}$, and for $\varepsilon_f = 10^{-5}$ and $\varepsilon_f = 10^{-4}$, PR is 0.667. A simple explanation for this odd observation is that LBPADE could find a fifth global minimum in some runs with low precision. Another possibility, which can explain a higher PR in a higher dimensions, is that two or more solutions sharing the same basin have been counted as approximations of different global minima by the PR calculator employed in the CEC'2013 competition package. This redundancy may occur only if ε_f is not small enough and/or the defined niche radius for post-processing [3] is not large enough in comparison with the basin sizes of the global minima. Data on the reported global minima are required to discover the actual cause of this odd observation.
- Problems 15, 17, 19, and 20 have the same fitness function (CF4) with 3, 5, 10, and 20 variables, respectively [3]. In the lowest dimension, there are several methods that tie with RS-CMSA-ES II; however, in higher dimensions, RS-CMSA-ESII outperforms the other methods by a practically significant margin.
- One remarkable observation is that for PID = 20, RS-CMSA-ESII is the only method that could reach a PR greater than 0.5. Since this function has eight global minima, we speculate that other methods could only find four specific global minima. To check this, we analyzed the reported results of RS-CMSA-ES, HillVallEA18, and HillVallEA19, for which detailed data of each run is publicly available. Our analysis has supported this speculation: none of these methods could find any of the four harder-to-find global minima of this problem in any run.
- Overall, the improvements made to RS-CMSA-ESII have resulted in an increase of 0.052 in MPR, which is considerable since for some of these problems both variants could reach PR ≈ 1.
- The standard deviation calculated over the PR of the tested methods (SDPR) may reveal the usefulness of a test problem for benchmarking purposes; for example, for PID = 1, 2, 3, 4, 5, 10, SDPR is almost zero, which means that all the tested methods have similar performance on these problems. These problems are too easy; however, it is not the only case that can result in a small SDPR. It is also possible that some of the global minima in a problem are so easy that most methods can find them and the rest of global minima are so hard to find that no MMO method can detect them. Such a problem is not a discriminative test problem either since it cannot highlight the differences between different methods.

VI. Extension of RS-CMSA-ESII for Dynamic Problems

RS-CMSA-ES and RS-CMSA-ESII can be extended to dynamic MMO by employing a prediction method which provides the initial population after the change. The dynamic variants of these methods are denoted by DRS-CMSA-ES(II), which can be DRS-CMSA-ES or DRS-CMSA-ESII. DRS-CMSA-ES is formed by suppressing five main modifications proposed in this study (see Table I). The other modifications that aimed to improve the time efficiency of the method are kept active to save computation resources.

At the onset of time step $t+1$, a set of candidate populations are suggested by the prediction method. These suggested populations will be used in the subsequent restarts as the initial population. Let $\boldsymbol{x}_{A_m}^t$ be the m^{th} archived solution at the end of time step t. The center and covariance matrix of the initial population for the $(m-1)^{\text{th}}$ restart at time step $t+1$ are calculated as follows:

TABLE II
PEAK RATIO FOR EACH PROBLEM OF THE CEC'2013 TEST SUITE FOR MMO [3] OBTAINED USING DIFFERENT METHODS. THE REPORTED PR IS THE AVERAGE PR FOR THREE FUNCTION TOLERANCES $\varepsilon_f = 10^{-3}, 10^{-4}, 10^{-5}$. THE PR OF THE DOMINANT METHODS FOR EACH PROBLEM ARE PROVIDED IN BOLD FONT UNLESS THERE ARE MORE THAN THREE DOMINANT METHODS FOR THAT PROBLEM. FOR RS-CMSA-ESII, THE STANDARD ERROR IS ALSO PROVIDED. THE LAST COLUMN SHOWS THE STANDARD DEVIATION OF THE PR OF THE COMPARED METHODS FOR EACH PROBLEM.

PID	Method													SDPR
	ANDE	LBPADE	DIDE	FBK-DE	MM-DE	TS-ABC	NCjDE-2LS$_{ar}$	NEA2	NMMSO	RS-CMSA-ES	HillValleA18	HillValleA19	RS-CMSA-ESII	
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.000
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.001
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.000
6	1.000	0.985	1.000	0.660	1.000	1.000	1.000	0.636	0.661	0.999	1.000	1.000	1.000 ± 0.0000	0.152
7	0.937	0.886	0.921	0.813	0.916	0.799	0.976	0.914	1.000	0.998	1.000	1.000	1.000 ± 0.0000	0.070
8	0.946	0.624	0.691	0.824	0.971	0.910	**1.000**	0.240	0.897	0.875	0.920	0.975	**0.997** ± 0.0008	0.212
9	0.511	0.510	0.571	0.425	0.463	0.457	0.833	0.581	0.978	0.734	0.945	0.972	**0.990** ± 0.0010	0.225
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.989	1.000	1.000	1.000	1.000	1.000 ± 0.0000	0.003
11	1.000	0.707	1.000	1.000	1.000	0.980	1.000	0.962	0.990	0.997	1.000	1.000	1.000 ± 0.0000	0.080
12	1.000	0.752	1.000	0.935	1.000	0.998	1.000	0.838	0.993	0.948	1.000	1.000	1.000 ± 0.0000	0.078
13	0.714	0.699	0.977	1.000	0.667	0.693	0.978	0.954	0.983	0.997	1.000	1.000	0.993 ± 0.0046	0.142
14	0.667	0.667	0.761	**0.906**	0.667	0.667	0.714	0.806	0.721	0.803	**0.917**	**0.923**	0.850 ± 0.0071	0.100
15	0.636	0.678	0.748	0.729	0.750	0.518	0.746	0.717	0.635	0.745	0.750	0.750	0.750 ± 0.0000	0.070
16	0.667	0.637	0.667	0.709	0.667	0.667	0.667	0.673	0.660	0.667	0.687	0.723	**0.833** ± 0.0000	0.049
17	0.397	0.527	0.591	0.633	0.636	0.453	0.694	0.695	0.466	0.695	**0.750**	**0.750**	**0.750** ± 0.0000	0.122
18	0.653	**0.695**	0.667	0.667	0.658	0.667	0.667	0.666	0.650	0.667	0.667	0.667	0.667 ± 0.0000	0.011
19	0.363	0.527	0.538	0.522	0.500	0.505	0.575	0.667	0.448	0.508	0.585	0.593	**0.703** ± 0.0086	0.088
20	0.249	0.300	0.352	0.449	0.088	0.293	0.492	0.357	0.172	0.468	0.483	0.480	**0.618** ± 0.0042	0.148
MPR	0.787	0.760	0.824	0.814	0.799	0.780	0.867	0.785	0.813	0.855	0.885	0.892	0.907 ± 0.0006	

$$\boldsymbol{x}^{t+1}_{\text{mean}_{m-1}} \leftarrow \boldsymbol{x}^{t}_{A_m},\ \mathbf{C}^{t+1}_{m-1} \leftarrow \left(\boldsymbol{x}^{\text{U}} - \boldsymbol{x}^{\text{L}}\right)\left(\boldsymbol{x}^{\text{U}} - \boldsymbol{x}^{\text{L}}\right)^{T}, \tag{11}$$

in which $\boldsymbol{x}^{\text{U}}$ and $\boldsymbol{x}^{\text{L}}$ define the search space. The global step size is calculated based on the change in the m^{th} global minimum from time step $t-1$ to time step t. To calculate this change, the archived solution at time step $t-1$ that corresponds to $\boldsymbol{x}^{t}_{A_m}$ should be determined first. This solution is denoted by $\boldsymbol{x}^{t-1}_{A_m}$ and is the one that has the smallest normalized Euclidean distance to $\boldsymbol{x}^{t}_{A_m}$:

$$\boldsymbol{x}^{t-1}_{A_m} = \underset{\boldsymbol{x}^{t-1}_{A_k}}{\text{argmin}} \left\{ \tilde{d}\left(\boldsymbol{x}^{t}_{A_m}, \boldsymbol{x}^{t-1}_{A_k}\right) \right\}, \tag{12}$$

in which $\tilde{d}\left(\boldsymbol{x}^{t}_{A_m}, \boldsymbol{x}^{t-1}_{A_k}\right)$ calculates the normalized Euclidean distance between $\boldsymbol{x}^{t}_{A_m}$ and $x^{t-1}_{A_k}$:

$$\tilde{d}\left(\boldsymbol{x}^{t}_{A_m}, \boldsymbol{x}^{t-1}_{A_k}\right) = \sqrt{\sum_{j=1}^{D} \left(\frac{x^{t}_{A_{mj}} - x^{t-1}_{A_{kj}}}{x^{\text{U}}_j - x^{\text{L}}_j} \right)^2}. \tag{13}$$

This equation shows that the distances have been normalized with respect to the search range. Now $\sigma^{t+1}_{\text{mean}_{m-1}}$ is set as follows:

$$\sigma^{t+1}_{\text{mean}_{m-1}} = \min \left\{ 0.3, 0.3 \frac{\tilde{d}(\boldsymbol{x}^{t}_{A_m}, \boldsymbol{x}^{t-1}_{A_m})}{\sqrt{D}} \right\},\ m = 1, 2, \ldots, \left|\mathbb{A}^t\right| \tag{14}$$

For the first $|\mathbb{A}^t|$ restarts at time step $t+1$, $\mathbb{P}$ is formed using (11) and (14) to exploit the history of past information. The benefits of this exploitation depend on how many global minima have been successfully detected in the previous time step. Therefore, for the subsequent restarts (restarts $|\mathbb{A}^t|$ onward), this initial solution is determined using the initialization method of RS-CMSA-ESII to improve exploration for global minima that were not found in the previous time step.

Dynamic RS-CMSA-ES(II) employs a simple reevaluation-based change detection mechanism. It generates and evaluates a set of λ randomly selected solutions in the search space at the beginning of each time step. Afterwards, $\max\{1, \lfloor 0.1\lambda \rfloor\}$ of these solutions are selected and reevaluated at each iteration to detect a potential change.

A. Comparison with Existing Dynamic MMO Methods

The modified moving peak benchmark (MMPB) developed in [49] is employed to evaluate the performance of DRS-CMSA-ES(II). *MMPB*(D, G, L) generates a dynamic multi-modal function in D-dimensional space which has G global minima and L local ones. These minima move along a random direction with a step defined by parameter S. Luo et al. [49] employed this test suite with different values for the parameters of the MMPB generator; however, only their most complex case is considered here ($D = L = G = 5$) for brevity. Following their experimental settings, DRS-CMSA-ES and DRS-CMSA-ESII perform 30 independent runs to optimize 30 instances of *MMPB*$(5, 5, 5)$. The target function tolerance is 10^{-5} and a total of 60 time steps are considered for each problem. The mean peak ratio (MPR) is then calculated for each problem, which is the average fraction of global minima detected at the end of each time step. Table III compares the

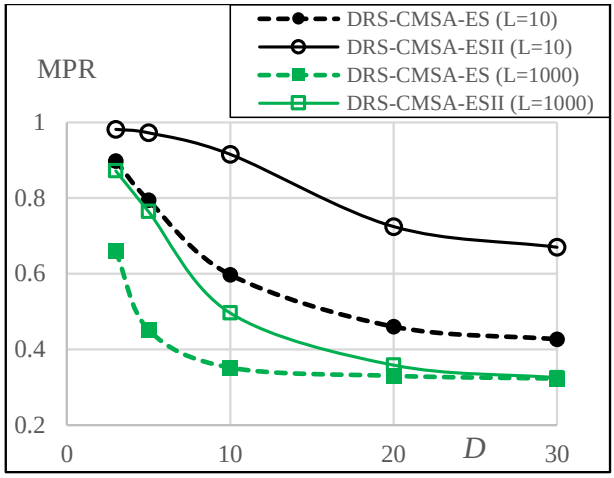
TABLE III
COMPARISON BETWEEN THE PERFORMANCE OF DRS-CMSA AND DRS-CMSA-ESII ON THE *MMPB*$(5, 5, 5)$ PROBLEM (MEAN $\pm$ SE) WITH THAT OF THREE OTHER METHODS REPORTED IN [49]

S	DMMCSA	CSA	PSO	DRS-CMSA-ES	DRS-CMSA-ESII
0.1	0.341	0.087	0.117	0.981 $\pm$ 0.002	**0.993** $\pm$ **0.001**
0.2	0.329	0.083	0.103	0.976 $\pm$ 0.004	**0.993** $\pm$ **0.001**
0.3	0.463	0.085	0.092	0.981 $\pm$ 0.002	**0.992** $\pm$ **0.001**
0.4	0.587	0.09	0.093	0.981 $\pm$ 0.002	**0.994** $\pm$ **0.001**
0.5	0.622	0.096	0.089	0.970 $\pm$ 0.003	**0.993** $\pm$ **0.001**

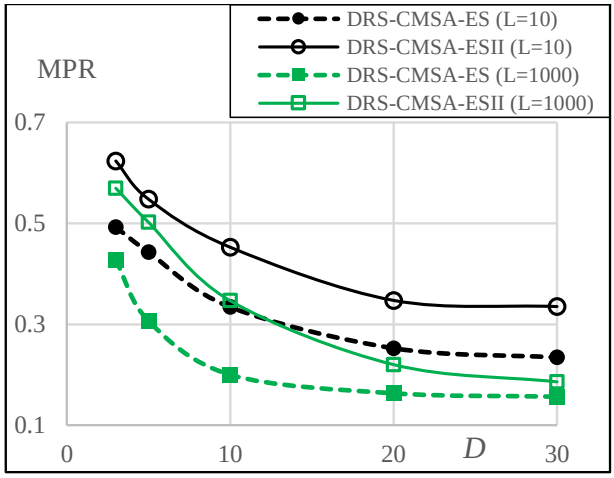
MPR of DRS-CMSA-ES and DRS-CMSA-ESII with that of three other methods reported in [49]. A comparison of results reveals that both DRS-CMSA-ES variants outperform the other methods by a considerable margin. The second variant shows slightly better results which is statistically significant, assuming that MPR follows a normal distribution. Both methods could detect and track almost all global minima in all runs, indicating that these test problems were too easy for them.

B. Results for more Challenging Dynamic MMO Problems

A more challenging setting for the MMPB generator is suggested in this study. First, we noticed that the search range of $[-1, 1]^D$ is too small considering the widths of minima, and consequently, most local minima would be masked by the global minima. Alternatively, the search range is set to $[-500, 500]^D$ in this subsection. Besides the problem dimensionality, the effects of the number of global minima ($G = 5, 10$) and the number of global minima ($L = 10, 1000$) are studied. Each experiment is repeated 50 times independently when $S = 10$. Other settings are similar to those of the previous experiment. Fig. 10 shows the MPR calculated for RS-CMSA-ES and RS-CMSA-ESII. Details of the obtained results are provided in the supplementary material **S2**. As observed, RS-CMSA-ESII outperforms the older variant in most, if not all, of the tested problems. More importantly, the results of DRS-CMSA-ES(II) is expected to provide a benchmark for challenging dynamic MMO test problems for future studies. It should be highlighted that any method would show a dramatic performance decline when G is high because the allocated budget per time step does not increase with G; however, some methods may show an earlier or a more severe performance decline than others.



(a) $G = 5$



(b) $G = 10$

Fig. 10. MPR calculated for DRS-CMSA-ES and DRS-CMSA-ESII for the MMPB problems with different values for the number of global minima (G) and local minima (L), and problem dimension (D)

VII. SUMMARY AND CONCLUSIONS

This study has introduced a new variant of the covariance matrix self-adaptation evolution strategy with repelling subpopulations (RS-CMSA-ES), a recent and successful for multimodal optimization (MMO). The resulting approach, called RS-CMSA-ESII, has undergone major improvements in several of its components including: i) the adaptation scheme of the normalized taboo distances of the archived solutions, ii) the elite selection scheme for adaptation of the covariance matrix, iii) the stopping criteria for predicting convergence to local or already detected minima, iv) handling bounds of the search space when sampling solutions, v) the initialization of a subpopulation, and vi) the determination of critical taboo points. Controlled numerical simulations have shown the importance of these improvements.

RS-CMSA-ESII was compared with the most successful MMO methods on the widely accepted CEC'2013 test suite for MMO. The comparison of results has shown the superiority of RS-CMSA-II over the existing methods, including its older variant (RS-CMSA-ES). RS-CMSA-ESII was also extended to dynamic MMO by reinforcing it with a prediction method and a change detection mechanism. Dynamic RS-CMSDA-ESII outperformed a few recent methods for dynamic MMO when compared on the modified dynamic moving benchmark problems.

The CEC'2013 test suite for MMO has made a substantial contribution to advancing knowledge in this field by providing a widely accepted experimental setup to facilitate a fair comparison of MMO methods. However, this test suite has remained unchanged since 2013. The caveat of specialization of niching methods for this specific test suite may produce a negative bias for the research conducted in this field. Strengthening this test suite with a more diverse set of benchmark

functions with known properties can overcome this problem.

The proposed elite selection mechanism provides a new research direction for elite selection in evolution strategies, which has been overlooked in favor of the non-elite comma selection scheme for single-objective optimization. For MMO and multi-objective optimization, elite selection has been the most preferred choice. More analyses on the proposed elite selection scheme may result in a robust selection approach that can compete with the preferred comma scheme in unconstrained single objective optimization as well.

ACKNOWLEDGMENT

This research has been funded by the Australian Research Council Discovery Project DP190102637. Computational resources for this study were provided by the National Computational Infrastructure (NCI), which is supported by the Australian Government. The source code of RS-CMSA-ESII in Python is available upon request. RS-CMSA-ESII with slightly different settings has participated and won the competitions on niching methods for multimodal optimization, which were held at GECCO and CEC in 2020. The last author gratefully acknowledges support from CONACyT grant no. 2016-01-1920 (*Investigación en Fronteras de la Ciencia 2016*) and from a SEP-Cinvestav grant (application no. 4). He was also partially supported by the Basque Government through the BERC 2018-2021 program by the Spanish Ministry of Science.

REFERENCES

[1] S. Das, S. Maity, B.-Y. Qu, and P. N. Suganthan, "Real-parameter evolutionary multimodal optimization-a survey of the state-of-the-art," *Swarm and Evolutionary Computation*, vol. 1, no. 2, pp. 71–88, 2011.

[2] X. Li, M. G. Epitropakis, K. Deb, and A. Engelbrecht, "Seeking multiple solutions: an updated survey on niching methods and their applications," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 518–538, 2016.

[3] X. Li, A. Engelbrecht, and M. G. Epitropakis, "Benchmark functions for cec'2013 special session and competition on niching methods for multimodal function optimization," RMIT University, Tech. Rep., 2013.

[4] R. Azzouz, S. Bechikh, and L. B. Said, "Dynamic multi-objective optimization using evolutionary algorithms: A survey," in *Recent advances in evolutionary multi-objective optimization*. Springer, 2017, pp. 31–70.

[5] T. T. Nguyen, S. Yang, and J. Branke, "Evolutionary dynamic optimization: A survey of the state of the art," *Swarm and Evolutionary Computation*, vol. 6, pp. 1–24, 2012.

[6] L. Cao, L. Xu, E. D. Goodman, and H. Li, "Decomposition-based evolutionary dynamic multiobjective optimization using a difference model," *Applied Soft Computing*, vol. 76, pp. 473–490, 2019.

[7] K. Deb, S. Karthik *et al.*, "Dynamic multi-objective optimization and decision-making using modified nsga-ii: a case study on hydro-thermal power scheduling," in *International conference on evolutionary multi-criterion optimization*. Springer, 2007, pp. 803–817.

[8] T. Blackwell and J. Branke, "Multi-swarm optimization in dynamic environments," in *Workshops on Applications of Evolutionary Computation*. Springer, 2004, pp. 489–500.

[9] N. Hansen, S. Finck, R. Ros, A. Auger *et al.*, "Real-parameter black-box optimization benchmarking 2009: noiseless functions definitions," INRIA, Tech. Rep. RR-6829, 2009.

[10] D. E. Goldberg, J. Richardson *et al.*, "Genetic algorithms with sharing for multimodal function optimization," in *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*. Hillsdale, NJ: Lawrence Erlbaum, 1987, pp. 41–49.

[11] K. Dejong, "An analysis of the behaviour of a class of genetic adaptive systems. doctorat dissertation, dept. of computer and communication sciences, university of michigan," *Ann Arbor*, 1975.

[12] S. W. Mahfoud, "Niching methods for genetic algorithms," Doctoral dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 1995, uMI Order No. GAX95-43663.

[13] O. J. Mengshoel and D. E. Goldberg, "The crowding approach to niching in genetic algorithms," *Evolutionary computation*, vol. 16, no. 3, pp. 315–354, 2008.

[14] X. Li, "Niching without niching parameters: particle swarm optimization using a ring topology," *Evolutionary Computation, IEEE Transactions on*, vol. 14, no. 1, pp. 150–169, 2010.

[15] B.-Y. Qu, P. N. Suganthan, and S. Das, "A distance-based locally informed particle swarm model for multimodal optimization," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 3, pp. 387–402, 2012.

[16] J. Zou, Q. Deng, J. Zheng, and S. Yang, "A close neighbor mobility method using particle swarm optimizer for solving multimodal optimization problems," *Information Sciences*, vol. 519, pp. 332–347, 2020.

[17] S. Biswas, S. Kundu, and S. Das, "Inducing niching behavior in differential evolution through local information sharing," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 246–263, 2014.

[18] B.-Y. Qu, P. N. Suganthan, and J.-J. Liang, "Differential evolution with neighborhood mutation for multimodal optimization," *IEEE transactions on evolutionary computation*, vol. 16, no. 5, pp. 601–614, 2012.

[19] S. Biswas, S. Kundu, and S. Das, "An improved parent-centric mutation with normalized neighborhoods for inducing niching behavior in differential evolution," *IEEE transactions on cybernetics*, vol. 44, no. 10, pp. 1726–1737, 2014.

[20] H. Zhao, Z.-H. Zhan, Y. Lin, X. Chen, X.-N. Luo, J. Zhang, S. Kwong, and J. Zhang, "Local binary pattern-based adaptive differential evolution for multimodal optimization problems," *IEEE Transactions on Cybernetics*, 2019.

[21] Z.-G. Chen, Z.-H. Zhan, H. Wang, and J. Zhang, "Distributed individuals for multiple peaks: A novel differential evolution for multimodal optimization problems," *IEEE Transactions on Evolutionary Computation*, 2019.

[22] Y.-H. Zhang, Y.-J. Gong, Y. Gao, H. Wang, and J. Zhang, "Parameter-free voronoi neighborhood for evolutionary multimodal optimization," *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 2, pp. 335–349, 2019.

[23] J. D. Knowles, R. A. Watson, and D. W. Corne, "Reducing local optima in single-objective problems by multi-objectivization," in *International conference on evolutionary multi-criterion optimization*. Springer, 2001, pp. 269–283.

[24] J. Yao, N. Kharma, and P. Grogono, "Bi-objective multipopulation genetic algorithm for multimodal function optimization," *Evolutionary Computation, IEEE Transactions on*, vol. 14, no. 1, pp. 80–102, 2010.

[25] K. Deb and A. Saha, "Multimodal optimization using a bi-objective evolutionary algorithm," *Evolutionary computation*, vol. 20, no. 1, pp. 27–62, 2012.

[26] A. Basak, S. Das, and K. C. Tan, "Multimodal optimization using a biobjective differential evolution algorithm enhanced with mean distance-based selection," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 5, pp. 666–685, 2012.

[27] S. Bandaru and K. Deb, "A parameterless-niching-assisted bi-objective approach to multimodal optimization," in *2013 IEEE Congress on Evolutionary Computation*. IEEE, 2013, pp. 95–102.

[28] W.-J. Yu, J.-Y. Ji, Y.-J. Gong, Q. Yang, and J. Zhang, "A tri-objective differential evolution approach for multimodal optimization," *Information Sciences*, vol. 423, pp. 1–23, 2018.

[29] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[30] O. M. Shir, M. Emmerich, and T. Bäck, "Adaptive niche radii and niche shapes approaches for niching with the CMA-ES," *Evolutionary Computation*, vol. 18, no. 1, pp. 97–126, 2010.

[31] A. Ahrari, K. Deb, and M. Preuss, "Multimodal optimization by covariance matrix self-adaptation evolution strategy with repelling subpopulations," *Evolutionary Computation*, vol. 25, no. 3, pp. 439–471, 2017.

[32] H.-G. Beyer and B. Sendhoff, "Covariance matrix adaptation revisited–the cmsa evolution strategy–," in *International Conference on Parallel Problem Solving from Nature*. Springer, 2008, pp. 123–132.

[33] M. Preuss, "Niching the CMA-ES via nearest-better clustering," in *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*. ACM, 2010, pp. 1711–1718.

[34] X. Lin, W. Luo, and P. Xu, "Differential evolution for multimodal optimization with species by nearest-better clustering," *IEEE Transactions on Cybernetics*, 2019.

[35] M. Preuss, "Improved topological niching for real-valued global optimization," in *Applications of Evolutionary Computation*. Springer, 2012, pp. 386–395.

[36] Z.-J. Wang, Z.-H. Zhan, Y. Lin, W.-J. Yu, H. Wang, S. Kwong, and J. Zhang, "Automatic niching differential evolution with contour prediction approach for multimodal optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 1, pp. 114–128, 2019.

[37] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *science*, vol. 315, no. 5814, pp. 972–976, 2007.

[38] G. Dominico and R. S. Parpinelli, "Multiple global optima location using differential evolution, clustering, and local search," *Applied Soft Computing*, p. 107448, 2021.

[39] H. Li, P. Zou, Z. Huang, C. Zeng, and X. Liu, "Multimodal optimization using whale optimization algorithm enhanced with local search and niching technique," *Math. Biosci. Eng.*, vol. 17, no. 1, pp. 1–27, 2020.

[40] X. Wang, M. Sheng, K. Ye, J. Lin, J. Mao, S. Chen, and W. Sheng, "A multilevel sampling strategy based memetic differential evolution for multimodal optimization," *Neurocomputing*, vol. 334, pp. 79–88, 2019.

[41] J. E. Fieldsend, "Running up those hills: Multi-modal search with the niching migratory multi-swarm optimiser," in *2014 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2014, pp. 2593–2600.

[42] S. Maree, T. Alderliesten, D. Thierens, and P. A. Bosman, "Real-valued evolutionary multi-modal optimization driven by hill-valley clustering," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2018, pp. 857–864.

[43] R. K. Ursem, "Multinational evolutionary algorithms," in *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, vol. 3. IEEE, 1999, pp. 1633–1640.

[44] S. Maree, T. Alderliesten, D. Thierens, and P. A. Bosman, "Benchmarking the hill-valley evolutionary algorithm for the gecco 2018 competition on niching methods multimodal optimization," *arXiv preprint arXiv:1807.00188*, 2018.

[45] I. Moser and R. Chiong, "Dynamic function optimization: the moving peaks benchmark," in *Metaheuristics for Dynamic Optimization*. Springer, 2013, pp. 35–59.

[46] J. Branke, "Memory enhanced evolutionary algorithms for changing optimization problems," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, vol. 3. IEEE, 1999, pp. 1875–1882.

[47] S. Cheng, H. Lu, Y.-n. Guo, X. Lei, J. Liang, J. Chen, and Y. Shi, "Dynamic multimodal optimization: A preliminary study," in *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2019, pp. 279–285.

[48] S. Cheng, H. Lu, W. Song, J. Chen, and Y. Shi, "Dynamic multimodal optimization using brain storm optimization algorithms," in *International Conference on Bio-Inspired Computing: Theories and Applications*. Springer, 2018, pp. 236–245.

[49] W. Luo, X. Lin, T. Zhu, and P. Xu, "A clonal selection algorithm for dynamic multimodal function optimization," *Swarm and Evolutionary Computation*, vol. 50, p. 100459, 2019.

[50] A. Ahrari and K. Deb, "A novel class of test problems for performance evaluation of niching methods," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 6, pp. 909–919, 2017.

[51] S. Maree, T. Alderliesten, and P. A. Bosman, "Benchmarking hillvallea for the gecco 2019 competition on multimodal optimization," *arXiv preprint arXiv:1907.10988*, 2019.

[52] Y.-H. Zhang, Y.-J. Gong, H.-Q. Yuan, and J. Zhang, "A tree-structured random walking swarm optimizer for multimodal optimization," *Applied Soft Computing*, vol. 78, pp. 94–108, 2019.

Ali Ahrari received his Ph.D. degree in mechanical engineering from Michigan State University, Michigan, USA. He is now a postdoc research associate at the University of New South Wales, ACT, Australia. His research concentrates on evolutionary algorithms and engineering optimization. He has won GECCO and CEC competitions on multimodal optimization in 2016 and 2020, and ISCSO student competition in structural optimization in 2017 and 2018. He is now a member of the IEEE CIS Task Force on Multimodal Optimization.

Saber Elsayed received the Ph.D.degree in Computer Science from the University of New South Wales Canberra, Australia, in 2012. Currently, Saber is a Senior Lecturer with the School of Engineering and Information Technology, University of New South Wales Canberra. His research interests include the areas of evolutionary algorithms, constraint-handling techniques for evolutionary algorithms, scheduling, big data and cybersecurity using computational intelligence. Saber won several IEEE-CEC competitions. Dr. Elsayed is serving as the chair of the IEEE Computational Intelligence Society (ACT Chapter), and has held organizational roles in several conferences.

Ruhul Sarker received his Ph.D. in 1992 from Dalhousie University, Halifax, Canada. He is currently a Professor in the School of Engineering and IT, Canberra, Australia. His main research interests are Evolutionary Optimization, and Applied Operations Research. He is the lead author of the book *Optimization Modelling: A Practical Approach*. Prof. Sarker is a member of IEEE and INFORMS.

Daryl Essam received his B.Sc. degree in computer science from University of New England, Australia in 1990 and his Ph.D. degree from University of New South Wales, Australia, in 2000. Since 1994, he has been with the Canberra campus, UNSW, where he is currently a Senior Lecturer. His research interests include genetic algorithms, with a focus on both evolutionary optimization and large scale problems.

Carlos A. Coello Coello (M'98-SM'04-F'11) received a PhD in computer science from Tulane University, USA, in 1996. He is currently Professor with Distinction (CINVESTAV-3F Researcher) at the Computer Science Department of CINVESTAV-IPN, in Mexico City, Mexico. Dr. Coello has authored and co-authored over 500 technical papers and book chapters. He has also co-authored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Second Edition, Springer, 2007). His publications currently report over 58,300 citations in Google Scholar (his h-index is 96). Currently, he is the Editor-in-Chief of the IEEE Transactions on Evolutionary Computation and serves in the editorial board of several other international journals. He received the 2007 National Research Award from the Mexican Academy of Sciences in the area of Exact Sciences, the Medal to the Scientific Merit 2009, granted by Mexico City's congress and the National Medal of Science and Arts in the area of Physical, Mathematical and Natural Sciences. He is also a recipient of the 2013 IEEE Kiyo Tomiyasu Award, of the 2016 The World Academy of Sciences (TWAS) Award in "Engineering Sciences," and of the 2021 IEEE Computational Intelligence Society Evolutionary Computation Pioneer Award. He is a Fellow of the IEEE, and a member of the ACM and the Mexican Academy of Science. His major research interests are: evolutionary multi-objective optimization and constraint-handling techniques for evolutionary algorithms.