Large-Scale Optimization via Evolutionary Multitasking assisted Random Embedding

Yinglan Feng[1], Liang Feng[2], Yaqing Hou[3] and Kay Chen Tan[1]
Department of Computer Science, City University of Hong Kong, Hong Kong[1]
College of Computer Science, Chongqing University, China[2]
College of Computer Science and Technology, Dalian University of Technology, China[3]
Email: kaytan@cityu.edu.hk

Abstract—**Evolutionary algorithms (EAs) often lose their superiority and effectiveness when applied to large-scale optimization problems. In the literature, many research studies have been proposed to improve the search performance of EAs, such as cooperative co-evolution, embedding, and new search operator design. Among those, memetic multi-agent optimization (MeMAO) is a recently proposed paradigm for high-dimensional problems by using random embeddings. It demonstrated high efficacy with the assumption of "effective dimension". However, as prior knowledge is always unknown for a given problem, this method may fail on the large-scale problems that do not have low effective dimensions. Taking this cue, we propose an evolutionary multitasking (EMT) assisted random embedding method (EMT-RE) for solving large-scale optimization problems. Instead of conducting a search on the randomly embedded space directly, we treat the embedded task as the auxiliary task for the given problem. By performing EMT with both the given problem and the randomly embedded task, not only the useful solutions found along the search can be transferred across tasks toward efficient problem solving, but the effectiveness of the search on problems without a low effective dimensionality is also guaranteed. To evaluate the performance of newly proposed EMT-RE, comprehensive empirical studies are carried out on 8 synthetic continuous optimization functions with up to 2,000 dimensions.**

Index Terms—**Large-Scale Optimization, Evolutionary Multitasking, Random Embedding, Knowledge Transfer**

I. Introduction

In the last decades, evolutionary algorithms (EAs) such as differential evolution (DE) [1], [2], genetic algorithm (GA) [3], and evolutionary strategy (ES) [4], have been successfully applied to solve the numerical and combinatorial optimization problems with low or medium dimensionality [5]. However, many optimization problems in real life may involve a large number of decision variables, which are known as large scale optimization problems. Most EAs lose their superiority and effectiveness when applied to those large-scale optimization problems, suffering from the "curse of dimensionality" [6].

In the literature, there are some efforts to improve conventional EAs for large-scale optimization problems, which can be divided into two categories, i.e., decomposition and non-decomposition [7]. Decomposition algorithms based on the cooperative co-evolution (CC) method mainly decompose the original problem into several subproblems that can be handled easier by employing the idea of divide-and-conquer (DC) [8]. According to different grouping strategies, the two main directions are random grouping [6], [9] and grouping strategy based on variable correlation [10]–[14]. However, such methods of decomposition rely on the decomposability of decision variables, so they may fail to solve non-separable problems. Besides, the dimensionality mismatch problem (i.e., the partial solutions to a subproblem can hardly be precisely evaluated) is another major hurdle when applying DC to non-separable problems, even though a self-evaluation evolution approach is proposed to relieve the stress of evaluation [15]. The non-decomposition approaches have emerged next, such as self-adapting the control parameters [16], designing new search operators [17], introducing structured population and migration strategy [18], [19], and embedding local search strategy [20], [21].

Among the non-decomposition approaches, random embedding techniques have been shown effective for tackling large-scale optimization problems. With the assumption that most dimensions in the high-dimensional problem do not change the objective significantly and only a small subspace affects the function value [22]–[24], different independently drawn random embeddings are employed to identify the underlying linear effective subspace [25]–[28]. However, real-world problems may not have a certain effective subspace, and it is hard to verify the existence of such effective dimensions [29]. More recently, a memetic multi-agent optimization (MeMAO) paradigm has been proposed in the literature to improve the searching efficiency for high-dimensional optimization problems with a "low effective dimensionality" [30]. By leveraging a memetic multi-agent learning system [31], MeMAO first reformulates the target optimization problem into multiple low-dimensional tasks (i.e., multi-agent environment) via random embedding methods and then constructs the interaction mechanisms among agents for information sharing while the individual search progresses online. Nevertheless, since MeMAO conducts a search on the randomly embedded space directly, the limitation of the assumption of "effective dimension" has remained unresolved. Moreover, it is worth noting that there is no guarantee on the preservation of the original global optimum in the embedded low-dimensional subspaces.

Evolutionary multitasking (EMT) as a recently emerged research topic that contends to conduct an evolutionary search on multiple search spaces corresponding to different tasks concurrently, which utilizes the seamless transfer of knowledge among tasks to speed up the evolutionary process [32]. Taking this cue, if we treat the randomly embedded space as the auxiliary task for the target problem, by performing EMT with both the original problem and the embedded task in the multi-task scenario, the large-scale optimization problems can be solved with the following benefits: 1) The useful solutions found along the search can be transferred across tasks toward efficient problem solving; 2) The effectiveness of the search on problems without a low effective dimensionality is also guaranteed.

Inspired by this, here we propose a large-scale optimization method via evolutionary multitasking assisted random embedding, namely EMT-RE. Specifically, we construct several simple forms of the problem (i.e., reformulate it into different low-dimensional problems) by random embeddings and take them into account as auxiliary tasks of the original problem. These tasks work together to build an evolutionary multitasking environment. Next, domain-specific solvers are used for evolving selfish genetic materials of each task, while a unified solution representation is designed as a general solver to represent multiple task domains simultaneously. Further, the knowledge transfer across tasks is conducted implicitly through the chromosomal crossover with two solutions possessing different skill factors along the evolutionary search process. To evaluate the efficacy of the proposed method, comprehensive empirical studies on 8 synthetic continuous optimization functions are conducted. Experiment results highlight the significant improvement of the proposed EMT-RE method compared to MeMAO, thus verifying the superior search performances of the proposed evolutionary multitasking assisted random embedding for large scale optimization, in terms of both solution quality and convergence speed.

The rest of this paper is organized as follows. Section II gives the background of random embedding and the existing implicit evolutionary multitasking paradigm. Next, the detailed design of the proposed EMT-RE framework is illustrated in Section III. Section IV provides comprehensive empirical studies on the commonly used synthetic optimization functions. Finally, we discuss the concluding remarks of the paper in Section V.

II. Preliminaries

This section first presents a brief introduction to the concept of random embedding, especially the practice in [30]. Next, an introduction of the evolutionary multitasking optimization and the existing implicit EMT paradigm is also presented.

A. Random Embedding

For high-dimensional optimization problems, the search space will grow exponentially as the dimension increases. Random embedding methods are commonly used for dimensionality reduction. Previous studies [23], [27] have shown that it is feasible to optimize a high-dimensional function with low effective dimensionality by using random embedding which requires a simple modification (i.e., multiplication by a random embedding matrix). Given a high-dimensional function $f : \mathbb{R}^D \to \mathbb{R}$ with low effective dimensionality d_e and a random matrix $\mathbf{A} \in \mathbb{R}^{D \times d}$ of normally distributed random numbers and $d \leq d_e$. For any $\mathbf{x} \in \mathbb{R}^D$, there always exists a $\mathbf{y} \in \mathbb{R}^d$ such that $f(\mathbf{x}) = f(\mathbf{A}\mathbf{y})$.

Based on the existing study, MeMAO [30] employs the random embedding matrix $\mathbf{A} \in \mathbb{R}^{D \times d}$ to reconstruct the original high-dimensional optimization problem $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^D$, into several low-dimensional tasks $g(\mathbf{y}), \mathbf{y} \in \mathbb{R}^d$, in multi-agent environment. D and d represent high dimension and low dimension respectively, and there exists a $\mathbf{y} \in \mathbb{R}^d$ satisfying $f(\mathbf{x}) = f(\mathbf{A}\mathbf{y})$ for any $\mathbf{x} \in \mathbb{R}^D$. Notably, all low-dimensional optimization tasks on different agents, embedded from the target high-dimensional problem by different random embedding matrices, are faced with entirely different searching landscapes. Therefore, when conducting the meme transmission process, the genetic materials cannot be simply transferred from one agent to another (denoted as $agt(1)$ and $agt(2)$). Instead, the information is firstly embedded into the searching space of the target high-dimensional optimization problem by $agt(1)$'s random embedding matrix $\mathbf{A}_1$ and then mapped back to the search space of target agent $agt(2)$ by the pseudo inverse matrix $pinv(\mathbf{A}_2)$, where $pinv(\mathbf{A})$ equals to $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

B. Evolutionary Multitasking Optimization

Inspired by biocultural models of multifactorial inheritance, evolutionary multitasking optimization, which conducts evolutionary search concurrently on multiple search landscapes corresponding to different optimization problems or tasks, has been proposed to improve problem-solving performance across multiple optimization problems by seamlessly transferring knowledge among them. Currently, knowledge sharing across tasks in EMT algorithms is commonly realized by the implicit genetic transfer through chromosomal crossover [32], [33]. The implicit EMT usually employs a population of individuals with a unified solution representation for solving multiple tasks. To compare population members in the multi-task sense, a set of properties for every individual are defined as follow:

- *Factorial Cost*: For a given task T_i, the factorial cost f_p^i of an individual p denotes its fitness or objective value on this task.
- *Factorial Rank*: The factorial rank r_p^i of an individual p on task T_i denotes the index of p in the list of population members sorted in ascending order with respect to factorial cost on this specific task.
- *Scalar Fitness*: The scalar fitness ϕ_p of an individual p is defined as $\phi_p = 1/\min_{j\in\{1,...,K\}} r_p^j$ based on the best rank over all tasks, where K is the number of tasks.
- *Skill Factor*: The skill factor τ_p denotes the task on which p is most effective, where $\tau_p = \arg\min_j r_p^j, j \in \{1, ..., K\}$.

According to the above definitions, performance comparison can be carried out straightforwardly based on the scalarized fitness (i.e., *Scalar Fitness*) [32]. For example, individual p_1

is considered to dominate p_2 in a multitasking environment if $\phi_1 > \phi_2$. Further, the general structure of implicit EMT algorithms is provided next.

Algorithm 1: General Structure of Implicit EMT

Generate an initial population of N_p individuals as P in a unified search space Y;
for ***every*** p_i ***in*** P **do**
  Assign the skill factor τ_i of p_i ;
  Evaluate p_i for task τ_i ;
while *stopping conditions are not satisfied* **do**
  Apply genetic operators, i.e., **assortative mating**, on P to generate an off-spring population C (Refer to Alg. 2);
  for ***every*** c_j ***in*** C **do**
    Determine the skill factor τ_j of c_j based on **vertical cultural transmission** (See Alg. 3);
    Evaluate c_j on task τ_j only;
  Intermediate population $T = P \cup C$;
  Update the scalar fitness and skill factor of individuals in T;
  Select the fittest N_p individuals from T to survive into the next generation P;

As shown in Algorithm 1, the implicit genetic transfer is implemented via two features of multifactorial inheritance acting. In particular, assortative mating allows individuals with distinct skill factors to mate with some probability, and vertical cultural transmission denotes that offspring can then randomly select a parental skill factor for imitation. Details of these features are referred to Algorithm 2 and Algorithm 3, respectively. Taking the advantages of EMT, this paper proposes an evolutionary multitasking assisted random embedding method for the large-scale optimization problem. Besides the original problem, several low-dimensional subproblems are formed by random embeddings to assist target optimization in a multi-task scenario. Notably, for more details of the existing EMT algorithm, interested readers can refer to the study in [32], [34]–[36].

III. Proposed Method

In this section, the design of the proposed EMT-RE model is presented. In particular, we first give an overview of the EMT-RE framework, which reformulates the original problem into multiple low-dimensional auxiliary tasks by random embeddings and performs EMT on both the original and the formulated tasks. Next, two features of multifactorial inheritance, namely assortative mating and vertical cultural transmission, are detailed respectively.

A. Overview of the proposed EMT-RE

Fig. 1 illustrates the framework of the proposed EMT-RE paradigm. As can be observed, for a given large-scale optimization problem with high dimension (i.e., $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^D, D = 3$ and represented by a cube), a number of random embedding matrics $\mathbf{A}_i \in \mathbb{R}^{D \times d}$ are employed to reformulate it into N low-dimensional subtasks (i.e., $g(\mathbf{y})$, $h(\mathbf{y})$, and $z(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^d, d = 2$ and represented by squares) as auxiliary tasks for the original problem. Note that every solution of the subtask is evaluated by mapping it back to the original high-dimensional space. In the multi-task scenario, EMT-RE performs both evolutionary search on a single task for self-evolution and knowledge transfer across tasks for sharing beneficial solutions, which is drove by assortative mating with a predefined probability rmp, along the evolutionary search process. As for the former, shown at the bottom of Fig. 1, tasks hold their domain-specific representations. By employing genetic operators, including mutation, crossover, and selection, multiple domain-specific solvers are used for evolving selfish genetic materials of each task. The particular evolutionary

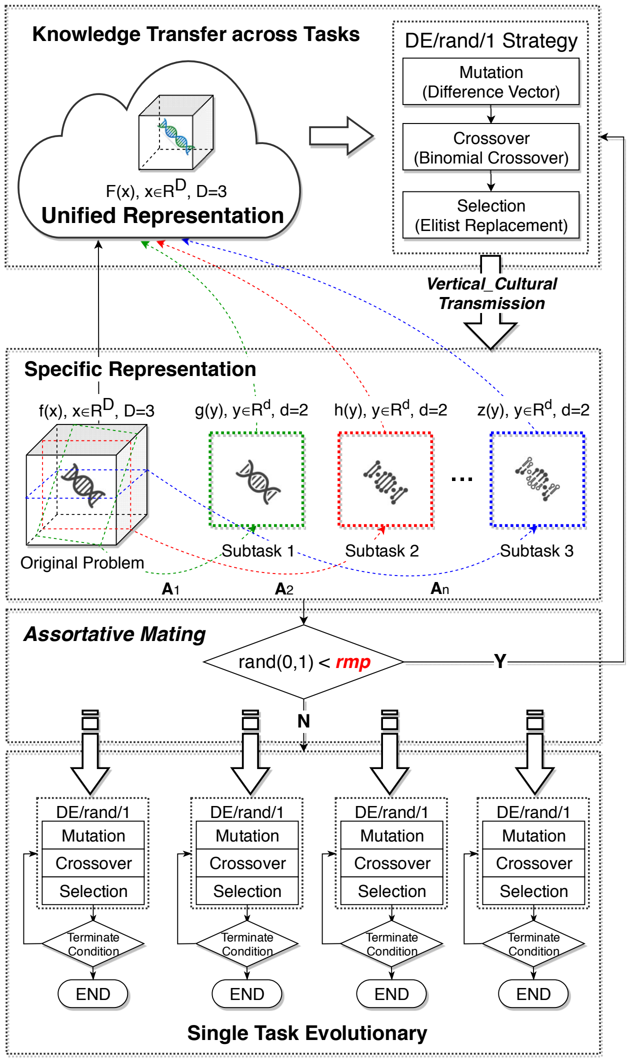


Fig. 1. The illustration of EMT-RE framework. N low-dimensional subspaces are randomly embedded as helper tasks for the given high-dimensional problem. EMT-RE is performed on both single task evolutionary and knowledge transfer across tasks.

solvers used herein are all DE with a strategy of DE/rand/1 [1]. For the latter, depicted on the top of Fig. 1, a general solver is operated directly on the unified representation by using the same DE as the basic solver. The knowledge transfer across tasks is triggered implicitly through the generation of the difference vector by two solutions possessing different skill factors and the binomial crossover. Domain-specific representations must be encoded within a unified representation scheme (i.e., mapping solutions of subtasks back to the original high-dimensional space by the corresponding random embedding matrix $\mathbf{A}_i$). After producing an offspring solution, vertical cultural transmission is conducted by selective imitation to evaluate the offspring for the selected task only. Further, the elitist strategy makes sure that only the best individuals survive through the generations. Details of genetic mechanisms (assortative mating) and selective evaluation (vertical cultural transmission) of EMT-RE shall be presented in the next two subsections.

B. Genetic Mechanisms

Algorithm 2: Assortative Mating for EMT-RE

```
1  for every individual p_i in current population P do
2      Get the skill factor τ_p of p_i to identify the current task;
3      Get the skill factor τ_b of the expert task;
4      if (τ_p == τ_b) or (rand(0, 1) > rmp) then
5          v_i^p = x_{r1}^p + F × (x_{r2}^p − x_{r3}^p) ;
6      else
7          u_i = A_b(x_{r2}^b − x_{r3}^b) ;
8          if p_i belongs to the original task (i.e., with high dimension) then
9              v_i^p = x_{r1}^p + F × u_i ;
10         else
11             v_i^p = x_{r1}^p + F × (pinv(A_p)u_i) ;
12     Perform binomial crossover on v_i^p to produce an offspring c.
```

After the population initialization step, every individual is endowed with a vector of high dimension D random variables. By employing different random embedding matrices $\mathbf{A}_i$, several low-dimensional optimization tasks with the specific representation are embedded from the original problem and differentiated by distinct skill factors.

A key to implicit EMT is that two parents must meet certain conditions to undergo crossover [32], which follows the principle of assortative mating [37]. In EMT-RE, the assortative mating is realized based on the DE/rand/1 mutation strategy and binomial crossover. Different from classical assortative mating in general implicit EMT, which completely select random parent candidates to undergo crossover [38], EMT-RE identifies the expert task holding the best fitness performance to interact with the current task. The learn-from-the-elitist principle [39] guarantees that useful solutions found along the search can be transferred across tasks toward efficient problem-solving. After that, if the current task and the expert task share a common skill factor, solutions are all randomly selected in this task to generate a mutant vector. In contrast, if their skill factor differs, the difference vector is formed by two random individuals from the expert task, which only occurs with a specified random mating probability (rmp). After the DE/rand/1 mutation, the binomial crossover is performed to generate an offspring. The detailed rules of assortative mating for proposed EMT-RE are provided in Algorithm 2.

In this case, the parameter rmp is predefined to balance depth and breadth of the evolutionary searching. A smaller value of rmp implies more internal evolution that only individuals sharing a common skill factor (i.e., $\mathbf{x}_{r2}^p$ and $\mathbf{x}_{r3}^p$) are allowed to conduct differential evolution for the single task evolutionary (see line 5 in Alg. 2). Conversely, a greater value of rmp leads to a higher frequency of knowledge transfer between the current task and the expert task. We generate a random number of $rand$ between 0 and 1. If $rand$ is less than predefined rmp, seamless information transmission from one task to another occurs.

Since domain knowledge of tasks is typically represented as the population-based genetic materials, the process of knowledge transfer across tasks is implemented through chromosomal information transmission. Firstly, two random and mutually exclusive individuals (i.e., $\mathbf{x}_{r2}^b$ and $\mathbf{x}_{r3}^b$) are selected from the expert task (i.e., under the skill factor τ_b) to create the difference vector. Since multiple low-dimensional optimization subtasks are embedded from the target problem via different random embedding matrices $\mathbf{A}_i$, they are confronted with completely different landscapes. We cannot directly transfer the genetic materials across tasks, but only encode the transferable information from the first step into the unified searching space by embedding matrix of the expert task $\mathbf{A}_b$ (see line 7 in Alg. 2). If the current task is exactly the original task (with high dimension), the mutant chromosome $\mathbf{v}_i^p$ for i^{th} individual p_i is simply generated from intermediate vector $\mathbf{u}_i$ by:

$$\mathbf{v}_i^p = \mathbf{x}_{r1}^p + F \times \mathbf{u}_i, \tag{1}$$

where $\mathbf{x}_{r1}^p$ is a randomly chosen individual from the current task, and F is the differential weight for controlling the amplitude of difference. If not, on the other hand, $\mathbf{u}_i$ will

Algorithm 3: Vertical Cultural Transmission via Selective Imitation

```
Given an offspring c mutated either from the current task τ_p or the expert task τ_b (see Alg. 2) ;
if c is mutated from τ_p then
    c imitate skill factor τ_p ;
else
    c imitate skill factor τ_b ;
    u_c = A_p v_c^p ;
    v_c^b = pinv(A_b) u_c ;
```

be mapped into the embedded space of the current task by the pseudo inverse of random embedding matrix $pinv(\mathbf{A}_p)$:

$$\mathbf{v}_i^p = \mathbf{x}_{r1}^p + F \times (pinv(\mathbf{A}_p)\mathbf{u}_i), \quad (2)$$

where $pinv(\mathbf{A})$ is approximated by $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

Once the trial vector $\mathbf{v}_i^p$ is created, the binomial crossover is performed on it to produce an offspring c with the same dimension of the current task.

C. Selective Evaluation

Evaluating every individual for every optimization task in EMT is computationally expensive. Hence, most EMT algorithms are designed to be efficient by reducing the total number of function evaluations. Specifically, the evaluation is performed only on selected tasks for which the individual is most likely to get high performance, as it is unlikely to perform well on all tasks. Taking the inspiration from vertical cultural transmission [40] in which the phenotype (e.g., skill factor) of an offspring is directly influenced by the phenotype of its parents, we propose a selective imitation strategy for EMT-RE. It allows offspring to imitate the skill factor of either the current task or the expert task and be evaluated for the selected task. If an offspring is mutated from only the current task, it will be evaluated only for task τ_p. While for the expert task, in addition to inheriting the skill factor of τ_b, the chromosome of offspring is firstly mapped back to the unified searching space and then decoded to a task-specific phenotype space (i.e., the expert task). The steps involved are summarized in Algorithm 3.

IV. Empirical Study

In this section, we conduct comprehensive empirical studies to evaluate the performance of the proposed EMT-RE paradigm. Details of experimental configuration and results are discussed herein.

A. Experiment Setup

In this paper, the MeMAO paradigm [30], which addresses high-dimensional optimization using random embeddings, is considered as the baseline for comparison. Accordingly, most of the experimental settings in our empirical study, including testing optimization problems and basic evolutionary solver, are the same as MeMAO to ensure a fair comparison.

Firstly, 8 single-objective synthetic optimization functions, including Sphere, Ackley, Rastrigin, Weierstrass, Rosenbrock, Griewank, Schwefel, and Levy functions, are employed to testify the efficacy of the EMT-RE framework for the large-scale optimization problem. All of the functions are randomly rotated at first. With a high dimensionality of $D = 2,000$, the global optimal value is randomly generated within the range of $[-0.2, 0.2]^D$ to avoid being located at 0. Further, by leveraging random embedding methods, multiple low-dimensional tasks (i.e., $d = 30$) are embedded randomly from the given optimization problems in the search scope of $[-0.3, 0.3]^d$. Detailed configurations of random embedding are referred to [23], [29].

Next, the traditional DE is employed as the fundamental population-based evolutionary solver, using DE/rand/1 operator. Further, to ensure a fair comparison, the number of tasks in EMT-RE, and the number of agents in MeMAO are both configured to 5 in the following experiment. Besides, other evolutionary operators and parameters in this study are kept consistent with MeMAO and summarized as follows.

1) Population size: $N_p = 100$.
2) Maximum generations: $MaxGen = 100$.
3) Independent number of runs: $runs = 20$.
4) F and CR in DE: $F = 0.5, CR = 0.9$.
5) Random mating probability in EMT-RE: $rmp = 0.2$.

B. Results and Discussions

Table I summarizes the performance of the proposed EMT-RE, MeMAO, and DE in terms of averaged objective value and standard deviation on the single-objective optimization benchmarks over 20 independent runs. Since DE has been employed as the basic solvers in the proposed algorithm for tackling multiple tasks in each benchmark, the results obtained by single-task DE on the original high-dimensional problems are also presented in the table for comparison. Further, the Wilcoxon Rank Sum Test [41] with a 95% confidence interval is conducted on experimental results to verify the statistical significance of strategy variance. Symbols "$\approx$", "+", and "-" denote the compared method statistically similar, better, and worse than the proposed EMT-RE, respectively.

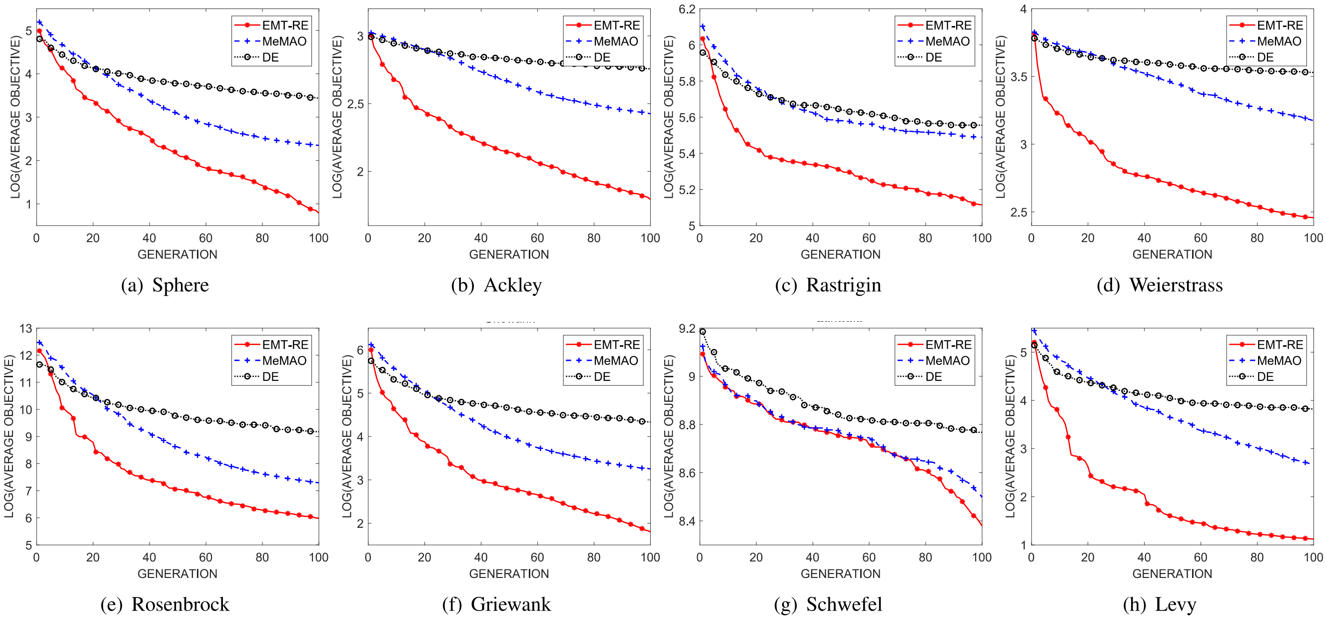
It can be observed in Table I, compared to MeMAO and single-task DE, the proposed EMT-RE achieved superior performance on most optimization problems in terms of averaged objective value. It is mainly because that EMT-RE treats the randomly embedded subproblem as a helper task for the original problem, and the useful solutions found along the search can be transferred across tasks. The improved solution quality confirmed the effectiveness of conducting implicit EMT for both the given problem and the randomly embedded tasks.

To assess the efficiency of the proposed EMT-RE, the average convergence traces of EMT-RE, MeMAO, and DE on all high-dimensional optimization functions are presented in Fig. 2. In the figure, the Y-axis denotes the averaged objective values obtained in log scale, while the X-axis gives the respective computational effort incurred in terms of the generation made so far. As can be observed, the traditional DE obtains better performance at the very beginning of the evolutionary process. However, with the useful solution transferring across tasks, the random embedding based methods including MeMAO and the proposed EMT-RE catch up from behind and exhibit superior performance over DE on most of the problems. Moreover, as MeMAO conducts searching on the embedded space directly and cannot guarantee the preservation of the original global optimum in the subspaces, the improvements in terms of convergence speed achieved by it are limited. Therefore, it is clear that EMT-RE converges much fast than MeMAO on almost all the problems except for the Schwefel function in Fig. 2(g), which is complicated

TABLE I

THE AVERAGED OBJECTIVE VALUE AND STANDARD DEVIATION OBTAINED BY THE PROPOSED EMT-RE ALGORITHM, MEMAO, AND DE ON THE LARGE-SCALE OPTIMIZATION PROBLEMS AFTER COMPLETING 10,000 FUNCTION EVALUATIONS. SYMBOLS "+", "-", AND "≈" DENOTE THE COMPARED METHOD IS SIGNIFICANTLY BETTER THAN, SIGNIFICANTLY WORSE THAN, AND STATISTICALLY TIED BY EMT-RE.

Funcs.		Sphere	Ackley	Rastrigin	Weierstrass	Rosenbrock	Griewank	Schwefel	Levy
EMT-RE	Mean	1.91E+00	6.01E+00	1.66E+02	1.17E+01	3.97E+02	6.05E+00	4.35E+03	3.07E+00
	Std.	6.61E-01	7.07E-01	3.48E+01	1.39E+00	1.22E+02	1.79E+00	1.10E+03	1.10E+00
MeMAO	Mean	1.05E+01-	1.13E+01-	2.42E+02-	2.39E+01-	1.47E+03-	2.59E+01-	4.90E+03≈	1.45E+01-
	Std.	2.84E+00	1.54E+00	1.27E+01	1.23E+00	7.34E+02	9.63E+00	1.15E+03	3.71E+00
DE	Mean	3.09E+01-	1.57E+01-	2.59E+02-	3.41E+01-	9.69E+03-	7.64E+01-	6.42E+03-	4.52E+01-
	Std.	4.67E+00	5.87E-01	1.35E+01	9.53E-01	1.82E+03	8.65E+00	5.99E+02	7.15E+00



(a) Sphere (b) Ackley (c) Rastrigin (d) Weierstrass

(e) Rosenbrock (f) Griewank (g) Schwefel (h) Levy

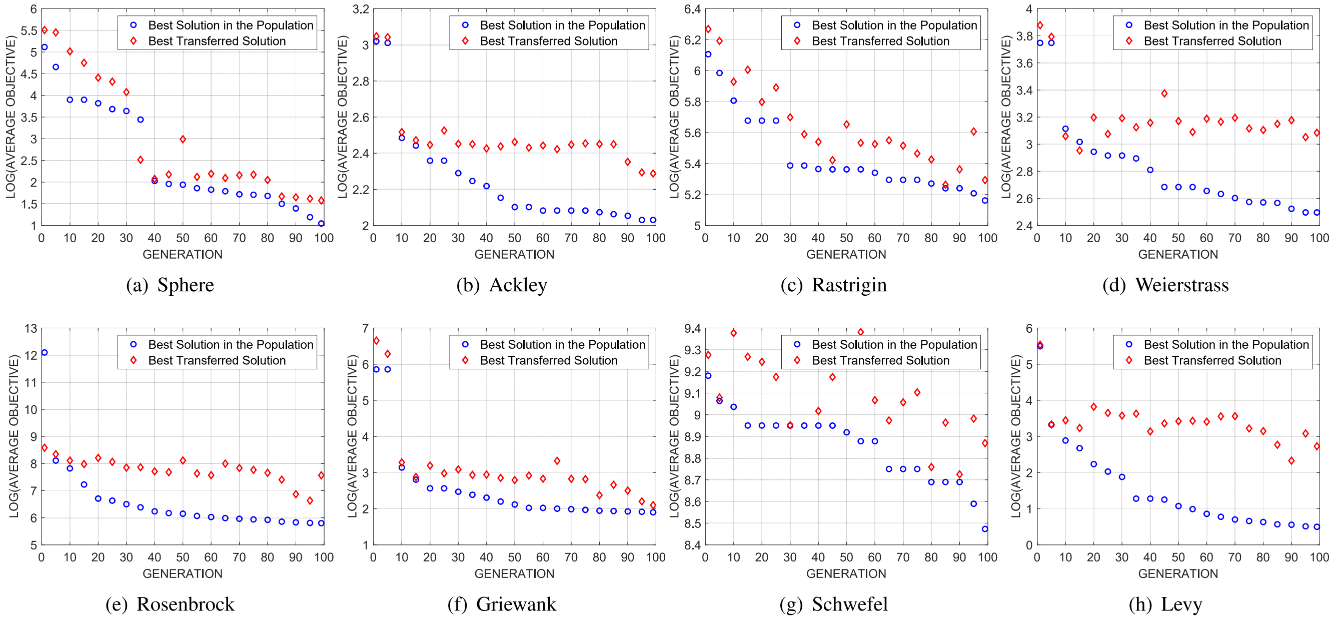
Fig. 2. Convergence traces of the proposed EMT-RE, MeMAO and single-task DE on eight high-dimensional problems. y-axis: log(objective value); x-axis: generation.

with many local minima closed to the global minimum [42]. On the other hand, on functions such as Rastrigin, Weierstrass, Levy, etc., an obvious drop of objective value can be observed in the convergence traces of the proposed EMT-RE method, which could be contributed to the effectiveness of the search on problems without a low effective dimensionality.

Further, to depict intuitive understanding of the improving performance of the EMT-RE for large-scale optimization problems, we present the best transferred solution across tasks of sampled generations and the corresponding best existing solution in the population on 8 target high-dimensional optimization problems, shown in Fig. 3. As can be observed, the transfer of useful solutions across tasks happens along the whole evolutionary search process in EMT-RE, and the transferred solutions are generally close to the best existing solutions in the population on each problem. Therefore, the proposed EMT-RE performs competitively against independent evolutionary methods. On the other hand, the high-quality transferred solutions, which possess lower objective values than the existing best solutions in the population, maybe precisely the point of the superior convergence speed of the proposed EMT-RE. For example, in Fig. 3(d) and Fig. 3(e), the superior solutions obtaining better objective values are transferred across tasks at the beginning of the evolution, thereby leading a clear drop of the objective value in the convergence trend.

V. CONCLUSION

In contrast to MeMAO, which is a memetic multi-agent optimization paradigm for high-dimensional optimization problems with the low effective dimensions, this paper has proposed an evolutionary multitasking assisted random embedding method (EMT-RE) for solving large-scale optimization problems. Instead of conducting a solution searching on the randomly embedded space directly, we treat the embedded subproblem as a helper task for the given problem and perform the implicit EMT on both the given problem and the randomly embedded task. Not only the useful solutions found along the search can be transferred across tasks for efficient problem solving, but also the effectiveness of the search on problems without a low effective dimensionality can be guaranteed. To verify the effectiveness of the proposed EMT-RE, comprehensive empirical studies on 8 synthetic



(a) Sphere (b) Ackley (c) Rastrigin (d) Weierstrass

(e) Rosenbrock (f) Griewank (g) Schwefel (h) Levy

Fig. 3. Illustration of the best transferred solution across tasks and the existing best solutions in the population on eight benchmarks. y-axis: log(objective value); x-axis: generation.

optimization problems have been conducted. The obtained results show that the embedded task assisted the original problem, and the knowledge transferred across tasks in EMT-RE provides a significant improvement in problem-solving, which has confirmed the efficacy of the proposed algorithm.

In the future, besides random embedding, there are many ways for the problem reformulation, either mathematically or other efficient learning approaches, which can be further explored. To improve the implicit genetic transfer in EMT, we would like to borrow ideas of explicit autoencoding or more autonomous knowledge sharing methods. Further, how to apply EMT-RE to solve large-scale optimization problems in real life, such as hyperparameter optimization [43], is also a research interest with great promise.

VI. Acknowledgement

This work is partially supported by the National Natural Science Foundation of China (NSFC) under grant No. 61876025, No. 61906032, and No. 61876162, by the Shenzhen Scientific Research and Development Funding Program under grant JCYJ20180307123637294, and by the Research Grants Council of the Hong Kong SAR under grant No. CityU11202418 and No. CityU11209219.

References

[1] R. Storn and K. Price, "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces," *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[2] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE transactions on evolutionary computation*, vol. 15, no. 1, pp. 4–31, 2010.

[3] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[4] J. Roubos, G. Van Straten, and A. Van Boxtel, "An evolutionary strategy for fed-batch bioreactor optimization; concepts and performance," *Journal of Biotechnology*, vol. 67, no. 2-3, pp. 173–187, 1999.

[5] D. Simon, *Evolutionary optimization algorithms*. John Wiley & Sons, 2013.

[6] Z. Yang, K. Tang, and X. Yao, "Large scale evolutionary optimization using cooperative coevolution," *Information Sciences*, vol. 178, no. 15, pp. 2985–2999, 2008.

[7] J.-R. Jian, Z.-H. Zhan, and J. Zhang, "Large-scale evolutionary optimization: a survey and experimental comparative study," *International Journal of Machine Learning and Cybernetics*, pp. 1–17, 2019.

[8] R. Descartes, *Discourse on method and meditations on first philosophy*. Hackett Publishing, 1999.

[9] X. Li and X. Yao, "Cooperatively coevolving particle swarms for large scale optimization," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 2, pp. 210–224, 2011.

[10] M. N. Omidvar, X. Li, and X. Yao, "Cooperative co-evolution with delta grouping for large scale non-separable function optimization," in *IEEE Congress on Evolutionary Computation*. IEEE, 2010, pp. 1–8.

[11] M. N. Omidvar, X. Li, Y. Mei, and X. Yao, "Cooperative co-evolution with differential grouping for large scale optimization," *IEEE Transactions on evolutionary computation*, vol. 18, no. 3, pp. 378–393, 2013.

[12] Y. Ling, H. Li, and B. Cao, "Cooperative co-evolution with graph-based differential grouping for large scale global optimization," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*. IEEE, 2016, pp. 95–102.

[13] M. N. Omidvar, M. Yang, Y. Mei, X. Li, and X. Yao, "Dg2: A faster and more accurate differential grouping for large-scale black-box optimization," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 6, pp. 929–942, 2017.

[14] Y. Sun, M. Kirley, and S. K. Halgamuge, "A recursive decomposition method for large scale continuous optimization," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 5, pp. 647–661, 2017.

[15] P. Yang, K. Tang, and X. Yao, "Turning high-dimensional optimization into computationally expensive optimization," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 1, pp. 143–156, 2017.

[16] J.-i. Kushida, A. Hara, and T. Takahama, "Rank-based differential evolution with multiple mutation strategies for large scale global optimization," in *2015 IEEE Congress on Evolutionary Computation*. IEEE, 2015, pp. 353–360.

[17] R. Cheng and Y. Jin, "A competitive swarm optimizer for large scale optimization," *IEEE transactions on cybernetics*, vol. 45, no. 2, pp. 191–204, 2014.

[18] Y.-F. Ge, W.-J. Yu, Y. Lin, Y.-J. Gong, Z.-H. Zhan, W.-N. Chen, and J. Zhang, "Distributed differential evolution based on adaptive mergence and split for large-scale optimization," *IEEE transactions on cybernetics*, vol. 48, no. 7, pp. 2166–2180, 2017.

[19] H. Wang, S. Rahnamayan, and Z. Wu, "Parallel differential evolution with self-adapting control parameters and generalized opposition-based learning for solving high-dimensional optimization problems," *Journal of Parallel and Distributed Computing*, vol. 73, no. 1, pp. 62–73, 2013.

[20] S.-Z. Zhao, J. J. Liang, P. N. Suganthan, and M. F. Tasgetiren, "Dynamic multi-swarm particle swarm optimizer with local search for large scale global optimization," in *2008 IEEE Congress on Evolutionary Computation*. IEEE, 2008, pp. 3845–3852.

[21] D. Molina and F. Herrera, "Iterative hybridization of de with local search for the cec'2015 special session on large scale global optimization," in *2015 IEEE congress on evolutionary computation*. IEEE, 2015, pp. 1974–1978.

[22] B. Chen, R. Castro, and A. Krause, "Joint optimization and variable selection of high-dimensional gaussian processes," *arXiv preprint arXiv:1206.6396*, 2012.

[23] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, and N. De Freitas, "Bayesian optimization in high dimensions via random embeddings," in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013, pp. 1778–1784.

[24] J. Djolonga, A. Krause, and V. Cevher, "High-dimensional gaussian process bandits," in *Advances in Neural Information Processing Systems*, 2013, pp. 1025–1033.

[25] H. Qian and Y. Yu, "Scaling simultaneous optimistic optimization for high-dimensional non-convex functions with low effective dimensions," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 2000–2006.

[26] C.-L. Li, K. Kandasamy, B. Póczos, and J. Schneider, "High dimensional bayesian optimization via restricted projection pursuit models," in *Artificial Intelligence and Statistics*, 2016, pp. 884–892.

[27] Z. Wang, F. Hutter, M. Zoghi, D. Matheson, and N. De Freitas, "Bayesian optimization in a billion dimensions via random embeddings," *J. Artif. Int. Res.*, vol. 55, no. 1, p. 361–387, Jan. 2016.

[28] Z. Wang, C. Li, S. Jegelka, and P. Kohli, "Batched high-dimensional bayesian optimization via structural kernel learning," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, 2017, p. 3656–3664.

[29] H. Qian, Y.-Q. Hu, and Y. Yu, "Derivative-free optimization of high-dimensional non-convex functions by sequential random embeddings." in *Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 1946–1952.

[30] Y. Hou, N. Jiang, H. Ge, Q. Zhang, X. Qu, L. Feng, and A. Gupta, "Memetic multi-agent optimization in high dimensions using random embeddings," in *2019 IEEE Congress on Evolutionary Computation*, June 2019, pp. 135–141.

[31] Y. Hou, Y.-S. Ong, L. Feng, and J. M. Zurada, "An evolutionary transfer reinforcement learning framework for multiagent systems," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 601–615, 2017.

[32] A. Gupta, Y. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 3, pp. 343–357, June 2016.

[33] K. K. Bali, A. Gupta, L. Feng, Y. S. Ong, and T. P. Siew, "Linearized domain adaptation in evolutionary multitasking," in *2017 IEEE Congress on Evolutionary Computation*. IEEE, 2017, pp. 1295–1302.

[34] Y.-S. Ong and A. Gupta, "Evolutionary multitasking: a computer science view of cognitive multitasking," *Cognitive Computation*, vol. 8, no. 2, pp. 125–142, 2016.

[35] L. Feng, L. Zhou, J. Zhong, A. Gupta, Y.-S. Ong, K.-C. Tan, and A. Qin, "Evolutionary multitasking via explicit autoencoding," *IEEE transactions on cybernetics*, vol. 49, no. 9, pp. 3457–3470, 2018.

[36] A. Gupta, Y.-S. Ong, and L. Feng, "Insights on transfer optimization: Because experience is the best teacher," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 51–64, 2017.

[37] C. R. Cloninger, J. Rice, and T. Reich, "Multifactorial inheritance with cultural transmission and assortative mating. ii. a general model of combined polygenic and cultural inheritance." *American journal of human genetics*, vol. 31, no. 2, p. 176, 1979.

[38] L. Zhou, L. Feng, K. Liu, C. Chen, S. Deng, T. Xiang, and S. Jiang, "Towards effective mutation for knowledge transfer in multifactorial differential evolution," in *2019 IEEE Congress on Evolutionary Computation*. IEEE, 2019, pp. 1541–1547.

[39] S. Mashohor, J. R. Evans, and T. Arslan, "Elitist selection schemes for genetic algorithm based printed circuit board inspection system," in *2005 IEEE Congress on Evolutionary Computation*, vol. 2. IEEE, 2005, pp. 974–978.

[40] X. Chen, Y.-S. Ong, M.-H. Lim, and K. C. Tan, "A multi-facet survey on memetic computation," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 5, pp. 591–607, 2011.

[41] F. Wilcoxon, S. Katti, and R. A. Wilcox, "Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test," *Selected tables in mathematical statistics*, vol. 1, pp. 171–259, 1970.

[42] M. Jamil and X.-S. Yang, "A literature survey of benchmark functions for global optimization problems," *arXiv preprint arXiv:1308.4008*, 2013.

[43] S. R. Young, D. C. Rose, T. P. Karnowski, S.-H. Lim, and R. M. Patton, "Optimizing deep learning hyper-parameters through an evolutionary algorithm," in *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, 2015, pp. 1–5.