

Adaptive Geodesic Flow Kernel Transfer for Many-Task Optimization

Yang-Tao Dai, Xiao-Fang Liu*

Institute of Robotics and Automatic Information Systems *School of Computer Science and Engineering* *Hanyang University*
College of Artificial Intelligence *South China University of Technology* *Ansan, South Korea*
Nankai University *Guangzhou, China* *Nankai University*
Tianjin, China *zhanapollo@163.com* *Tianjin, China*
liuxiaofang@nankai.edu.cn *junzhang@ieee.org*

Abstract—Many-task optimization problems (MaTOP) involve more than three tasks, which can be solved simultaneously via knowledge transfer by utilizing complementary information of different tasks. Due to the biases between tasks, relevant tasks are usually selected for knowledge transfer to avoid negative effects. There are two challenging issues, i.e., source task selection and inter-task knowledge transfer. To address these issues, this paper proposes an adaptive geodesic flow kernel transfer method (AGFKTM) for MaTOP. In AGFKTM, multiple source tasks are selected based on both the similarity between tasks and the performance of tasks. In this way, similar and well-performed tasks are selected with a high priority. In addition, an adaptive geodesic flow kernel is constructed to implement knowledge transfer, in which the adopted subspaces along the geodesic flow path are adaptively controlled. Particularly, the transferred solutions are used to generate new ones using mutation operators. Integrating the AGFKTM into differential evolution, a new algorithm named AGFKT-DE is put forward. Experimental results on GECCO20MaTOP benchmark show that the new algorithm outperforms state-of-the-art algorithms.

Index Terms—evolutionary computation, many-task optimization, knowledge transfer, geodesic flow kernel, differential evolution

I. INTRODUCTION

With the successful applications of evolutionary algorithms to real-world multitask optimization problems, e.g., resource scheduling in cloud computing [1] and topology optimization [2], evolutionary transfer optimization has become popular in recent years [3]. Evolutionary transfer optimization usually utilizes the searching experience of some tasks (source tasks) to assist the optimization of other tasks (target tasks). This is called knowledge transfer between tasks. Knowledge transfer can leverage the information of different tasks to help evolutionary algorithms to improve optimization performance [4].

Multiple knowledge transfer methods are developed for multitask optimization. With the emergence of big data, many-task optimization problems (MaTOP) become a significant problem, which involve more than three tasks [5]. Due to the increase of task number, knowledge transfer between tasks is

more challenging. Especially, it is critical to select appropriate source tasks for a specific target one from a large number of tasks [6]. In addition, since the tasks usually have a different landscape and global optimum, alignment between source and target tasks is important to avoid negative transfer.

Since tasks have different domain biases, relevant tasks are usually selected as source tasks for knowledge transfer. For example, the mean discrepancy value between a task and a target one is measured, and the task with a smaller maximum mean discrepancy value is regarded as a relevant task [6]. However, some tasks are similar to the target task but may perform poorly, which are not appropriate choices. Based on this observation, some task selection methods consider the performance of tasks. For example, source tasks are selected based on the fitness improvement of a task [7], or the number of candidate solutions that successfully update individuals [8]. There is work to combine the advantages of task similarity and task performance. For example, source tasks are selected according to the similarity based on shift invariance and the ranking of tasks [7]. However, the method only simply shares the parameter settings of several DE solvers.

Based on the selected source tasks, knowledge transfer is often performed using crossover or mutation operators [8]. Considering the differences between tasks in terms of global optima and landscapes, solutions from different tasks are aligned for knowledge transfer using multiple methods, e.g., the subspace alignment methods and domain adaptation methods. For example, sample geodesic flow is adopted to sample an intermediate subspace along the geodesic flow path between source and target domains for transfer [9]; subspace alignment methods transfer solutions from a source subspace to a target subspace [10]. However, these methods use a limited number of subspaces only and do not well exploit the subspaces between source and target domains.

According to the above discussion, this paper proposes differential evolution based on adaptive geodesic flow kernel transfer (AGFKT-DE) for MaTOP. In AGFKT-DE, a new source task selection method is developed to select multiple source tasks based on the similarity between tasks and the performance of tasks. In addition, an adaptive geodesic flow kernel transfer method is proposed for knowledge transfer

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62103202, in part by the Natural Science Foundation of Tianjin under Grant 21JCQNJC00140, and in part by the Fundamental Research Funds for the Central Universities, Nankai University (Grant 63231162). (Corresponding author: Xiao-Fang Liu)

using multiple mutation strategies. Particularly, an adaptive geodesic flow kernel is constructed for solution alignment, in which the adaptive parameter controls the endpoint of geodesic flow transfer path. In this way, the subspaces between source and target domains can be better exploited. Experimental studies are carried out on GECCO20MaTOP benchmark. The results show that AGFKT-DE is significantly better than state-of-the-art MaTOP algorithms in terms of solution optimality and convergence speeds.

The remainder of this paper is organized as follows. Section II gives the background. Section III presents the proposed method in detail. Experimental results are presented in Section IV. The conclusion is drawn in Section V.

II. BACKGROUND

A. Many-Task Optimization Problem

A MaTOP contains more than three tasks. Given a MaTOP with NT single-objective optimization tasks $\{f_k : \mathbf{x} \rightarrow \mathbb{R} | k = 1, \dots, NT\}$. Denote the k -th task as t_k , the optimal solution of t_k as \mathbf{x}_k^* , and the search space of t_k as $\Gamma_k \subseteq \mathbb{R}^{D_k}$, where D_k is the dimensionality of Γ_k . Each task t_k can be formulated as:

$$\begin{aligned} \mathbf{x}_k^* &= \arg \min_{\mathbf{x}_k \in \Gamma_k} f_k(\mathbf{x}_k), \\ \text{s.t. } \mathbf{x}_k^* &\in \Gamma_k, \forall 1 \leq k \leq NT. \end{aligned} \quad (1)$$

Optimization algorithms aim to find the optimal solutions of all tasks simultaneously. Since the search spaces of multiple tasks may be different, the solutions of each task are usually encoded into a unified search space.

B. Related Work

Multiple algorithms are developed to solve MaTOP. One of the most representative algorithm is multifactorial evolutionary algorithm (MFEA) [11]. Inspired by the biocultural models of multifactorial inheritance, MFEA generates offsprings by assortative mating in a unified search space and evaluates offsprings via vertical cultural transmission.

There are two key issues, i.e., the source task selection and the knowledge transfer strategy, to address. For source task selection, improvements brought by other tasks are measured so as to control the information flow between tasks [12]. To measure the similarity between tasks, a bi-objective measurement is proposed to evaluate the differences between tasks in terms of global optima and landscapes [13]. Similarly, the Kullback-Leibler divergence between populations is adopted to measure the task similarity [5]. The task similarity can also be measured based on shift invariance, which are used as features for task grouping using K-means clustering [7]. The gradient information and location of the best solution are combined to measure the task similarity [14]. In addition, anomaly detection model is constructed to select candidate transferred individuals [15].

Based on the selected source tasks, a reasoning strategy is designed to improve the diversity of knowledge and utilize the complementarities of different tasks based on the Bayesian method [16]. The denosing autoencoder [17] and linearized

domain adaptation [18] are also adopted to learn the linear mapping between tasks for explicit knowledge transfer. Furthermore, a kernelized autoencoder is designed to construct nonlinear mapping between tasks [19]. Similarly, transfer component analysis is applied to reduce differences between distinct domains and learn a shared subspace between source and target domains [20]. Moreover, an ensemble method is proposed to take the advantages of multiple domain adaptation methods for knowledge transfer [21]. Specifically, for tasks with different dimensions, the knowledge transfer is implemented using dimension supplement strategy based on RBF neural network and dimension reduction strategy based on dynamic principal component analysis [14].

III. THE PROPOSED METHOD

A. Source Task Selection

A MaTOP contains multiple tasks with a different bias. To avoid negative knowledge transfer, it is pivotal to select appropriate source tasks for a specific target task. It is necessary to consider both the similarity between tasks and the performance of tasks. Thus, we propose a new selection strategy based on the similarity and performance of tasks, named SPS. In SPS, the similarity between two task are measured using the Wasserstein distance [22] between their populations. In addition, the performance of a task is evaluated using the successful update times of individuals during the evolutionary procedure, denoted as ST . The ST values of all tasks are calculated and the median value is denoted as ST_m .

Given a target task with a high ST value ($ST \geq ST_m$), top n_1 tasks (with smallest WD values) similar to the target one are selected first. Then, n_2 ($n_1 \geq n_2$) tasks with largest ST are selected from the n_1 tasks as source tasks. In contrast, given a target task with a relatively low ST value ($ST < ST_m$), n_2 tasks are selected via roulette wheel selection. The selection probability of j -th task is denoted as p_j and calculated by

$$p_j = \frac{1}{2} \left(\frac{1}{NT - 2} \cdot \frac{\sum_{k \neq i}^{NT} WD_{ik} - WD_{ij}}{\sum_{k \neq i}^{NT} WD_{ik} + \epsilon} + \frac{ST_j}{\sum_{k \neq i}^{NT} ST_k + \epsilon} \right) \quad (2)$$

where NT is the number of tasks in a MaTOP, i is the index of the target task, WD_{ik} is the Wasserstein distance between the target task and k -th task, ST_k is the performance of k -th task evaluated using the successful update times of individuals, and $\epsilon > 0$ is a small value to avoid division by zero. Particularly, the sum of the probabilities of all tasks equals to 1. The proposed strategy tends to maintain the good performance of target tasks by selecting similar and well-performed tasks. Besides, diversified tasks are selected to help performance improvement for poorly-performed tasks.

B. Knowledge Transfer between Tasks

1) *Geodesic Flow Kernel*: The basic assumption of geodesic flow is that data can be embedded in a low-dimension subspace. To construct geodesic flow, principal component analysis (PCA) [23] is applied to establish the linear subspaces of source and target domains. Denote the dimensionality of

Algorithm 1 AGFKT-DE

Input: The parameters for source task selection n_1, n_2, NT optimization tasks $\{t_1, t_2, \dots, t_{NT}\}$.

Output: NT solutions $\{\mathbf{x}_k^* | k = 1, 2, \dots, NT\}$ for NT tasks.

```
1: Initialize subpopulations  $\{pop_k | k = 1, 2, \dots, NT\}$ .
2:  $g \leftarrow 1, FE \leftarrow 0, STs \leftarrow zeros(NT), \alpha \leftarrow \{\alpha_k | \alpha_k = randint(1, 9)/10, k = 1, 2, \dots, NT\}$ .
3: while  $FE \leq maxFE$  do
4:   Compute geodesic flow kernel transfer matrices  $\{G^{*j \rightarrow i}(\alpha_i) | i, j = 1, 2, \dots, NT \text{ and } i \neq j\}$  between every two tasks by Eqs. (3), (7)-(9).
5:   for each task  $t_i$  do
6:      $S \leftarrow \emptyset$ .
7:     Select source tasks by SPS.
8:     for each individual  $\mathbf{x}_k$  in  $pop_i$  do
9:       if  $rand() < rmp$  then
10:        Randomly choose one mutation operator from Eqs. (10)-(13) to generate mutant vector  $\mathbf{u}_k$ .
11:       else
12:        Generate  $\mathbf{u}_k$  by individuals from  $pop_T$ .
13:       end if
14:       Generate trial vector  $\mathbf{v}_k$  by binomial crossover.
15:        $S \leftarrow S \cup \mathbf{v}_k$ .
16:     end for
17:      $S \leftarrow S \cup pop_i$ .
18:     Select the best half of solutions from  $S$  to update  $pop_i$  and count the number  $n$  of selected trial vectors.
19:      $ST_i \leftarrow ST_i + n$ , update  $FE$ .
20:   end for
21:   Adjust  $\alpha$  of each task by Eq. (5).
22:    $g \leftarrow g + 1$ .
23: end while
```

original search space as D and the dimensionality of subspace as d , where $d < D$. Denote the subspaces of source and target tasks are $P_S \in \mathbb{R}^{D \times d}$ and $P_T \in \mathbb{R}^{D \times d}$, respectively. The orthogonal complement to P_S is denoted as $R_S \in \mathbb{R}^{D \times (D-d)}$, s.t. $R_S^T P_S = \mathbf{0}$, where R_S^T is the transpose of R_S . Note that $d \leq D/2$, since the size of orthogonal matrix $U_2 \in \mathbb{R}^{(D-d) \times d}$ obtained by the SVD of $R_S^T P_T \in \mathbb{R}^{(D-d) \times d}$ should satisfy $D - d \geq d$. On the manifold, the geodesic flow $\Phi(t)$, where $\Phi(0) = P_S$ and $\Phi(1) = P_T$, connecting the subspaces of source and target domains can be parameterized as $\Phi(t) = P_S U_1 \Lambda(t) - R_S U_2 \Omega(t)$, where t is in a range of $(0, 1)$. By setting t to be a specific real number t_0 , $\Phi(t_0)$ represents a corresponding subspace on the geodesic flow. U_1, U_2, Λ and Ω are matrices obtained by singular value decomposition (SVD):

$$P_S^T P_T = U_1 \Lambda V^T, R_S^T P_T = -U_2 \Omega V^T \quad (3)$$

where $U_1 \in \mathbb{R}^{d \times d}$ and $U_2 \in \mathbb{R}^{(D-d) \times d}$ are orthogonal matrices, Λ and Ω are $d \times d$ diagonal matrices, and we simply denote them as $\Lambda = diag(\cos\theta_k)$ and $\Omega = diag(\sin\theta_k)$, where $0 \leq \theta_k \leq \pi/2$ are principal angles between $P_S, k = 1, 2, \dots, d$.

Instead of sampling a finite number of subspaces in the parameterized geodesic flow [9], this paper adopts geodesic flow kernel [24] to concatenate infinite projections of individuals into an infinite-dimensional feature vector \mathbf{z}^∞ . For two D -dimensional solutions \mathbf{x}_i and \mathbf{x}_j , the geodesic flow kernel is defined by inner product between the feature vectors:

$$\langle \mathbf{z}_i^\infty, \mathbf{z}_j^\infty \rangle = \int_0^\alpha (\Phi(t)^T \mathbf{x}_i)^T (\Phi(t)^T \mathbf{x}_j) dt = \mathbf{x}_i^T G \mathbf{x}_j \quad (4)$$

where $\alpha = 1$ represents the endpoint of the geodesic flow path for transfer, and $G \in \mathbb{R}^{D \times D}$ is a geodesic flow kernel matrix.

2) *Adaptive Geodesic Flow Kernel:* Since α is related to the subspaces adopted for knowledge transfer, the impact of α is investigated. Taking an instance I_5 in the GECCO20MaTOP as an example, we report the average value and standard deviation of the results over 20 runs in Table I. As show in Table I, a smaller α improves the algorithm performance and different problems require different α values. Based on the observation, we propose an adaptive geodesic flow kernel to diversify the transferred solutions by adaptively adjusting α according to the performance of a target task. Specifically, $0 < \alpha \leq 1$ is parameterized to adaptively control the subspaces used for domain adaptation. In addition, adapting control parameters can improve algorithm diversity [25].

In the adaptive geodesic flow kernel, α_i is adaptively controlled for each task t_i . For the i -th task, α_i is randomly initialized as $\alpha_i = randint(1, 9)/10$, where $randint(1, 9)$ is a random integer in a range of $[1, 9]$. In every generation, α_i is adjusted according to ST_i as

$$\alpha_i = \begin{cases} \beta * \alpha_j + (1 - \beta) * \alpha_i, & \text{if } ST_i > ST_m, \\ \gamma * \alpha_i, & \text{otherwise,} \end{cases} \quad (5)$$

where β is the learning rate and γ is the decay coefficient. Particularly, for a task with a high ST_i value (larger than ST_m), α_i is updated by learning from a random task t_j with $ST_j \geq ST_i$; otherwise, α_i is slightly reduced by multiplying a decay coefficient to improve optimization performance.

To compute the geodesic flow kernel matrix G , Eq. (4) is derived as follows:

$$\begin{aligned} G &\propto \int_0^\alpha \Phi(t) \Phi(t)^T dt \\ &\propto \int_0^\alpha \begin{bmatrix} diag(\frac{1+\cos(2\theta_k t)}{2}) & diag(-\frac{\sin(2\theta_k t)}{2}) \\ diag(-\frac{\sin(2\theta_k t)}{2}) & diag(\frac{1-\cos(2\theta_k t)}{2}) \end{bmatrix} dt \quad (6) \\ &\propto \begin{bmatrix} diag(\frac{\alpha}{2} + \frac{\sin(2\theta_k \alpha)}{4\theta_k}) & diag(\frac{\cos(2\theta_k \alpha) - 1}{4\theta_k}) \\ diag(\frac{\cos(2\theta_k \alpha) - 1}{4\theta_k}) & diag(\frac{\alpha}{2} - \frac{\sin(2\theta_k \alpha)}{4\theta_k}) \end{bmatrix} \end{aligned}$$

where the sign “ \propto ” represents “proportional to”. Denote the above diagonal matrices as:

$$\begin{aligned} \Sigma_1 &= diag\left(\frac{\alpha}{2} + \frac{\sin(2\theta_k \alpha)}{4\theta_k}\right), \\ \Sigma_2 &= diag\left(\frac{\cos(2\theta_k \alpha) - 1}{4\theta_k}\right), \\ \Sigma_3 &= diag\left(\frac{\alpha}{2} - \frac{\sin(2\theta_k \alpha)}{4\theta_k}\right). \end{aligned} \quad (7)$$

TABLE I
STATISTICAL RESULTS OF GFKTM WITH DIFFERENT SETTINGS OF α ON
17 RASTRIGIN TASKS, 17 GRIEWANK TASKS AND 16 WEIERSTRASS TASKS
IN I_5

α		17*rastrigin	17*griewank	16*weierstrass
1.0	mean	4.03E+02	7.08E-02	3.28E+00
	std	1.37E+01	1.25E-02	6.34E-01
0.8	mean	3.88E+02	2.28E-02	2.65E+00
	std	1.63E+01	3.47E-03	2.52E-01
0.6	mean	3.78E+02	1.84E-02	2.57E+00
	std	2.28E+01	6.65E-03	5.13E-01
0.4	mean	3.76E+02	9.55E-03	2.24E+00
	std	2.33E+01	4.01E-03	1.04E+00
0.2	mean	3.90E+02	7.32E-03	2.11E+00
	std	1.67E+01	3.96E-03	1.13E+00

Similarly, the geodesic flow kernel matrix $G(\alpha)$ with an adaptive parameter α is computed by

$$G(\alpha) = [P_S U_1 \quad R_S U_2] \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_3 \end{bmatrix} \begin{bmatrix} U_1^T P_S^T \\ U_2^T R_S^T \end{bmatrix}. \quad (8)$$

Since the the geodesic flow kernel is defined by inner product, the transfer matrix, denoted as $G^* \in \mathbb{R}^{D \times D}$, is the square root of the matrix G , i.e.,

$$G(\alpha) = G^*(\alpha)G^*(\alpha). \quad (9)$$

A solution $\mathbf{x} \in \mathbb{R}^{1 \times D}$ from the subpopulation of a source task can be transformed into a D -dimensional space between source and target domains related to α , i.e., $\mathbf{y} = \mathbf{x}G^*(\alpha)$, where $\mathbf{y} \in \mathbb{R}^{1 \times D}$ is the transferred solution.

3) *Knowledge Transfer along the Geodesic Path:* To perform knowledge transfer, the solutions from source tasks are transferred along the geodesic path by the transfer matrix $G^*(\alpha)$. In AGFKT-DE, three common mutation strategies, i.e., DE/rand/1, DE/current-to-rand/1 and DE/rand-to-best/1 [26], are adopted to generate diversified solutions for every individual in the subpopulation of target task. Denote the subpopulation of target task as pop_T , the subpopulations of n_2 selected source tasks are combined to form a population pop_S , and pop_T and pop_S are combined together to form a new population $pop_{ST} = pop_S \cup pop_T$.

$$\mathbf{u}_k = \mathbf{x}_{r1,T} + F(\tilde{\mathbf{x}}_{r2,S} - \tilde{\mathbf{x}}_{r3,S})G^*, \quad (10)$$

$$\mathbf{u}_k = \mathbf{x}_{k,T} + F(\tilde{\mathbf{x}}_{r1,ST}G^* - \tilde{\mathbf{x}}_{k,T}) + F(\tilde{\mathbf{x}}_{r2,S} - \tilde{\mathbf{x}}_{r3,S})G^*, \quad (11)$$

$$\mathbf{u}_k = \mathbf{x}_{r1,ST}G^* + F(\mathbf{x}_{best,T} - \mathbf{x}_{k,T}) + F(\tilde{\mathbf{x}}_{r2,S} - \tilde{\mathbf{x}}_{r3,S})G^*, \quad (12)$$

$$\mathbf{u}_k = \mathbf{x}_{r1,ST} + F(\mathbf{x}_{r2,S} - \mathbf{x}_{r3,S}), \quad (13)$$

For k -th individual $\mathbf{x}_{k,T}$ from the target task t_i , knowledge transfer is performed with a random mating probability (rpm). If knowledge transfer is performed, a mutation operator is randomly selected to generate a mutant vector \mathbf{u}_k by Eqs. (10)-(13), where F is a scale factor, $r1$, $r2$ and $r3$ are three distinct random integers, and G^* is the shorthand form of the transfer matrix $G^{*j \rightarrow i}(\alpha_i)$ (j is the index of a specific source task related to a solution that is transferred to i -th

task). The sign $\tilde{\mathbf{x}}$ is the resulting solution of \mathbf{x} normalized to zero-mean and unit-variance, and $\bar{\mathbf{x}}$ is the solution centered by subtracting the subpopulation center. The subscripts T , S and ST represent that the individual is selected from pop_T , pop_S and pop_{ST} , respectively, and the ranges of random integers $r1$, $r2$ and $r3$ are determined according to the size of the population from which the individual is selected. Specifically, a mutation operator in source domain Eq. (13) directly uses the individuals in the source domains to generate mutant vector to help increase solution diversity.

If knowledge transfer is not performed, a mutant vector \mathbf{u}_k is generated by three individuals which are randomly selected from pop_T using DE/rand/1 operator.

After the mutant vector \mathbf{u}_k is generated, a trail vector \mathbf{v}_k is generated via binomial crossover. The trail vectors and all the individuals of the target subpopulation are stored in a set S . The half of individuals in S with best fitness values are selected to update the target subpopulation. The complete algorithm is presented in **Algorithm 1**. FE is the number of function evaluations, and $maxFE$ is the maximum number of function evaluations. The algorithm stops until $FE > maxFE$, and outputs NT solutions for all optimization tasks.

IV. EXPERIMENTS

A. Experimental Settings

1) *Test Problems:* The widely used GECCO20MaTOP¹ benchmark is taken to test the performance of the proposed algorithm. The benchmark contains ten 50-task problems I_1 - I_{10} , which are constructed based on 7 different basic functions. In each problem, tasks are rotated and shifted, and the dimensionalities of all tasks are 50.

2) *Competing Algorithms:* To verify the performance of AGFKT-DE, seven typical and state-of-the-art MaTOP algorithms are selected for comparisons, i.e., differential evolution (DE), many-task algorithm based on biocoenosis through symbiosis using DE solver (DE/rand/1, $F = 0.5$, $CR = 0.7$) (EBSDE) [12], multifactorial DE based on adaptive intertask coordinate system (GFMFDE) [9], multitask evolutionary algorithm based on anomaly detection (MTEA-AD) [15], multitask optimization with adaptive knowledge transfer (AEMTO) [8], evolutionary many-task optimization based on multisource knowledge transfer (EMaTO-MKT) [6], and a two-stage transferable adaptive differential evolution (TRADE) [7]. These methods adopt different strategies for task selection and knowledge transfer, making the comparisons comprehensively and convincing.

3) *Parameter Settings:* The parameter settings of all competing algorithms are set the same as in the original papers. In the proposed AGFKT-DE, the population size is set to 50 for each task, F is set to 0.5, CR is set to 0.7, the dimensionality d of subspaces is set to 20, rpm is set to 0.3, γ is set to 0.9, β is set to 0.1, n_1 and n_2 are set to 6 and 3, respectively.

¹http://www.bdsc.site/websites/MTO_competition_2020/MTO_Competition_GECCO_2020.html

TABLE II
COMPARATIVE RESULTS BETWEEN AGFKT-DE AND ITS VARIANTS
noSPS, noAGFK AND noMSD ON GECCO20MATOP

AGFKT-DE vs Problems	noSPS W/T/L	noAGFK W/T/L	noMSD W/T/L
I_1	50/0/0 +	4/45/1 =	50/0/0 +
I_2	50/0/0 +	13/37/0 +	50/0/0 +
I_3	45/5/0 +	50/0/0 +	50/0/0 +
I_4	50/0/0 +	12/34/4 +	50/0/0 +
I_5	50/0/0 +	17/2/31 -	50/0/0 +
I_6	34/9/7 +	18/26/6 +	36/13/1 +
I_7	50/0/0 +	17/17/16 =	50/0/0 +
I_8	38/11/1 +	12/30/8 =	48/2/0 +
I_9	34/14/2 +	19/30/1 +	40/10/0 +
I_{10}	29/20/1 +	16/34/0 +	30/18/2 +
+/=/-	10/0/0	6/3/1	10/0/0

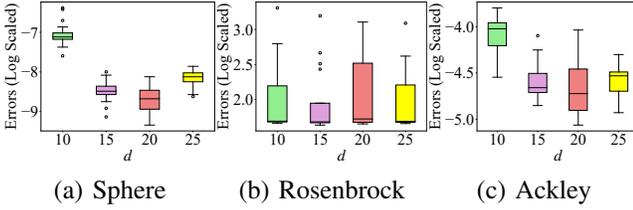


Fig. 1. Box plot of errors using AGFKT-DE with different d on 17 Sphere tasks, 17 Rosenbrock tasks and 16 Ackley tasks in I_4 .

In all experiments, the maximum number of function evaluations $maxFE$ is set to 2500000 for fair comparisons, and each algorithm runs 20 times independently on each problem. On each task, wilcoxon rank-sum tests are carried out between the proposed AGFKT-DE and each competing algorithm at a significance level of 0.05. The symbols “W/T/L” (Win/Tie/Lose) indicate that AGFKT-DE is significantly better than, equal to, and significantly worse than the competing algorithm on one task, respectively. If the number of “W” is larger than the number of “L” by 5 (10% of 50 tasks), AGFKT-DE is said to be better than the competing algorithm on a MaTOP following [7]. We use the signs “+”, “=” and “-” to represent AGFKT-DE is better than, equal to and worse than the competing algorithm on a MaTOP.

B. Parameter Investigation

The dimensionality d of subspaces is investigated in this subsection. Taking I_4 instance as an example, we test the algorithm performance by setting $d = 10, 15, 20, 25$. The statistical results are presented in Fig. 1. As shown in Fig. 1, $d = 10$ gets the worst performance on tasks constructed by sphere and ackley functions, showing that AGFKT-DE with a too low d value can not preserve enough information of the source populations for transfer. In contrast, $d = 20$ achieves the best performance on both sphere and ackley functions. Meanwhile, different d values get a similar performance on rosenbrock tasks. Therefore, $d = 20$ is a good choice.

C. Effects of Components in AGFKT-DE

To verify the effectiveness of components, i.e., source task selection strategy based on the similarity and performance of

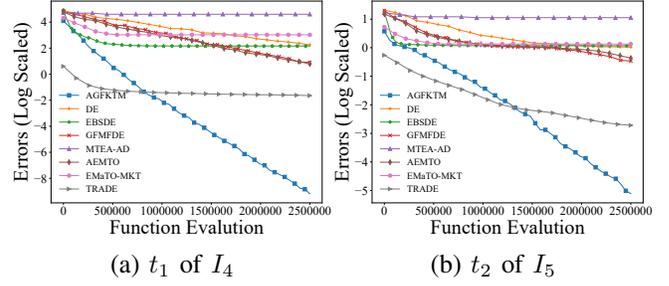


Fig. 2. Convergence curves of AGFKT-DE and competing algorithms on 2 tasks of problems I_4 and I_5 .

tasks, adaptive geodesic flow kernel, and mutation in source domains, three variants of AGFKT-DE, i.e., noSPS, noAGFK, and noMSD, are compared with AGFKT-DE. Specifically, noSPS selects n_2 source tasks randomly, noAGFK performs knowledge transfer by using the transfer matrix with a fixed $\alpha = 1$, and noMSD abandons the mutation in source domains Eq. (13). The statistical results are reported in Table II. As shown in Table II, AGFKT-DE performs better than noSPS and noMSD on all problems. It shows that the proposed source task selection strategy can choice good source tasks for transfer. In addition, mutation in source domains can also improve algorithm performance. Compared with noAGFK, AGFKT-DE performs better on 6 out of 10 problems and worse on 1 problem only. It shows that the adaptively control of α in geodesic flow kernel can improve algorithm performance. In general, the results indicate that all the three components are necessary for AGFKT-DE.

D. Compared with Other Algorithms

The statistical results of AGFKT-DE and competing algorithms are reported in Table III. As shown in Table III, AGFKT-DE is significantly better than DE, EBSDE, GFMFDE, MTEA-AD, AEMTO, and EMaTO-MKT on all problems. In addition, AGFKT-DE performs significantly better than TRADE on 8 out of 10 problems, whereas worse on 0 problem. Compared with GFMFDE using sample geodesic flow, the better performance of AGFKT-DE indicates that the adaptive geodesic flow kernel is more effective by using multiple intermediate subspaces along the geodesic flow.

To further investigate the convergence performance of all algorithms, taking I_4 and I_5 as examples, two different tasks are selected for observing the convergence curves of all algorithms. As shown in Fig. 2, among all algorithms, AGFKT-DE converges towards better results quickly on t_1 of I_4 and t_2 of I_5 . Thus, the proposed method has better performance and a faster convergence speed.

V. CONCLUSION

This paper proposes an adaptive geodesic flow kernel transfer method to solve MaTOP. Integrating the adaptive geodesic flow kernel transfer method and differential evolution, a new algorithm named AGFKT-DE is put forward. In AGFKT-DE, the source task selection is conducted to select multiple

TABLE III
COMPARATIVE RESULTS BETWEEN AGFKT-DE AND OTHER COMPETING ALGORITHMS ON GECCO20MATOP

AGFKT-DE vs Problems	DE W/T/L	EBSDE W/T/L	GFMFDE W/T/L	MTEA-AD W/T/L	AEMTO W/T/L	EMaTO-MKT W/T/L	TRADE W/T/L
I_1	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	45/5/0 +	50/0/0 +
I_2	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	46/4/0 +
I_3	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	50/0/0 +	46/0/4 +	50/0/0 +
I_4	50/0/0 +	50/0/0 +	36/14/0 +	50/0/0 +	35/15/0 +	50/0/0 +	24/26/0 +
I_5	50/0/0 +	50/0/0 +	34/0/16 +	50/0/0 +	34/1/15 +	50/0/0 +	32/2/16 +
I_6	50/0/0 +	50/0/0 +	50/0/0 +	41/9/0 +	50/0/0 +	34/8/8 +	36/14/0 +
I_7	50/0/0 +	50/0/0 +	32/2/16 +	47/3/13 +	27/10/13 +	50/0/0 +	19/15/16 =
I_8	50/0/0 +	40/8/2 +	39/1/10 +	49/0/1 +	47/3/0 +	49/0/1 +	18/17/15 =
I_9	50/0/0 +	42/7/1 +	42/0/8 +	35/12/3 +	49/1/0 +	42/0/8 +	27/12/11 +
I_{10}	50/0/0 +	40/9/1 +	40/2/8 +	46/3/1 +	50/0/0 +	40/0/10 +	35/5/10 +
+/=/-	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	8/2/0

source tasks, which are similar to the target task and also perform well. In addition, the knowledge transfer is performed along the geodesic path between source and target tasks, in which an adaptive geodesic flow kernel is constructed according to the performance of tasks. Experimental results on the GECCO20MaTOP benchmark show that AGFKT-DE is significantly better than state-of-the-art algorithms on most problems. The proposed algorithm can well transfer knowledge between tasks to improve optimization performance.

REFERENCES

- [1] L. Bao, Y. Qi, M. Shen, X. Bu, J. Yu, Q. Li, and P. Chen, "An evolutionary multitasking algorithm for cloud computing service composition," in *Services – SERVICES 2018*, A. Yang, S. Kantamneni, Y. Li, A. Dico, X. Chen, R. Subramanyan, and L.-J. Zhang, Eds. Cham: Springer International Publishing, 2018, pp. 130–144.
- [2] H. Sasaki, "Topology optimization for motor using multitask convolutional neural network under multiple current conditions," *IEEE Transactions on Magnetics*, vol. 58, no. 9, pp. 1–4, 2022.
- [3] K. C. Tan, L. Feng, and M. Jiang, "Evolutionary transfer optimization - a new frontier in evolutionary computation research," *IEEE Computational Intelligence Magazine*, vol. 16, no. 1, pp. 22–33, 2021.
- [4] T. Wei, S. Wang, J. Zhong, D. Liu, and J. Zhang, "A review on evolutionary multitask optimization: Trends and challenges," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 5, pp. 941–960, 2022.
- [5] Y. Chen, J. Zhong, L. Feng, and J. Zhang, "An adaptive archive-based evolutionary framework for many-task optimization," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 3, pp. 369–384, 2020.
- [6] Z. Liang, X. Xu, L. Liu, Y. Tu, and Z. Zhu, "Evolutionary many-task optimization based on multisource knowledge transfer," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 2, pp. 319–333, 2022.
- [7] S.-H. Wu, Z.-H. Zhan, K. C. Tan, and J. Zhang, "Transferable adaptive differential evolution for many-task optimization," *IEEE Transactions on Cybernetics*, pp. 1–14, 2023.
- [8] H. Xu, A. K. Qin, and S. Xia, "Evolutionary multitask optimization with adaptive knowledge transfer," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 2, pp. 290–303, 2022.
- [9] Z. Tang, M. Gong, Y. Wu, A. K. Qin, and K. C. Tan, "A multifactorial optimization framework based on adaptive intertask coordinate system," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 6745–6758, 2022.
- [10] Z. Tang, M. Gong, Y. Wu, W. Liu, and Y. Xie, "Regularized evolutionary multitask optimization: Learning to intertask transfer in aligned subspace," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 2, pp. 262–276, 2021.
- [11] A. Gupta, Y.-S. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 3, pp. 343–357, 2016.
- [12] R.-T. Liaw and C.-K. Ting, "Evolutionary many-tasking based on biocoenosis through symbiosis: A framework and benchmark problems," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, 2017, pp. 2266–2273.
- [13] Y. Jiang, Z.-H. Zhan, K. C. Tan, and J. Zhang, "A bi-objective knowledge transfer framework for evolutionary many-task optimization," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2022.
- [14] H. Han, X. Bai, H. Yang, Y. Hou, and J. Qiao, "Multitask particle swarm optimization with dynamic transformation," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–15, 2023.
- [15] C. Wang, J. Liu, K. Wu, and Z. Wu, "Solving multitask optimization problems with adaptive knowledge transfer via anomaly detection," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 2, pp. 304–318, 2022.
- [16] X. Wu, W. Wang, H. Yang, H. Han, and J. Qiao, "Diversified knowledge transfer strategy for multitasking particle swarm optimization," *IEEE Transactions on Cybernetics*, pp. 1–14, 2023.
- [17] L. Feng, L. Zhou, J. Zhong, A. Gupta, Y.-S. Ong, K.-C. Tan, and A. K. Qin, "Evolutionary multitasking via explicit autoencoding," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3457–3470, 2019.
- [18] 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 (CEC)*, 2017, pp. 1295–1302.
- [19] L. Zhou, L. Feng, A. Gupta, and Y.-S. Ong, "Learnable evolutionary search across heterogeneous problems via kernelized autoencoding," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 3, pp. 567–581, 2021.
- [20] X. Wang, Q. Kang, M. Zhou, S. Yao, and A. Abusorrah, "Domain adaptation multitask optimization," *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4567–4578, 2023.
- [21] W. Lin, Q. Lin, L. Feng, and K. C. Tan, "Ensemble of domain adaptation-based knowledge transfer for evolutionary multitasking," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2023.
- [22] V. M. Panaretos and Y. Zemel, "Statistical aspects of wasserstein distances," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 405–431, 2019. [Online]. Available: <https://doi.org/10.1146/annurev-statistics-030718-104938>
- [23] M. K. Jake Lever and N. Altman, "Principal component analysis," *Nat. Methods*, vol. 14, pp. 641–642, 07 2017.
- [24] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2066–2073.
- [25] Z.-H. Zhan, L. Shi, K. C. Tan, and J. Zhang, "A survey on evolutionary computation for complex continuous optimization," *Artificial Intelligence Review*, vol. 55, no. 1, pp. 59–110, Jan 2022.
- [26] X.-F. Liu, Z.-H. Zhan, Y. Lin, W.-N. Chen, Y.-J. Gong, T.-L. Gu, H.-Q. Yuan, and J. Zhang, "Historical and heuristic-based adaptive differential evolution," vol. 49, no. 12, pp. 2623–2635, 2019.