Evolutionary Sequential Transfer Optimization for Objective-Heterogeneous Problems

Xiaoming Xue, Cuie Yang, *Member, IEEE*, Yao Hu, Kai Zhang, *Member, IEEE*, Yiu-Ming Cheung, *Fellow, IEEE*, Linqi Song, *Member, IEEE*, and Kay Chen Tan, *Fellow, IEEE*

***Abstract*—Evolutionary sequential transfer optimization is a paradigm that leverages search experience from solved source optimization tasks to accelerate the evolutionary search of a target task. Even though many algorithms have been developed, they mainly focus on objective-homogeneous problems, where the source and target tasks possess a similar number of objectives. In this work, we explore objective-heterogeneous problems, in which knowledge transfers across single-objective optimization problems (SOPs), multiobjective optimization problems (MOPs), and many-objective optimization problems (MaOPs). Objective-heterogeneous problems challenge the existing methods due to the diverse search and objective spaces between the source and the target task. To address this issue, we present a decision variable analysis-based transfer method that can conduct knowledge transfer across problems with the different numbers of objectives. We first separate decision variables of MOPs and MaOPs into convergence-related variables (CVs) and diversity-related variables (DVs), according to their roles while treating variables of SOPs as CVs. Then, we propose a convergence transfer module to transfer knowledge of CVs to speed up the convergence. It aligns both solutions and fitness ranks for preserving fitness rank consistency between the source and target tasks, whereby accelerating search speed. Besides, a diversity transfer module is presented to refine the distribution of DVs to maintain the population diversity. The experimental results on objective-heterogeneous test problems and a real-world case study have demonstrated the effectiveness of the proposed algorithm.**

***Index Terms*—Convergence transfer, diversity transfer, evolutionary transfer optimization (ETO), heterogeneous problems, knowledge transfer.**

Manuscript received 8 April 2021; revised 28 July 2021 and 13 October 2021; accepted 27 November 2021. Date of publication 9 December 2021; date of current version 1 December 2022. This work was supported in part by the Hong Kong RGC under Grant ECS 21212419, in part by the Technological Breakthrough Project of Science, Technology and Innovation Commission of Shenzhen Municipality under Grant JSGG20201102162000001; in part by the Guangdong Basic and Applied Basic Research Foundation through the Key Project under Grant 2019B1515120032; and in part by the Hong Kong Laboratory for AI-Powered Financial Technologies. *(Corresponding authors: Cuie Yang; Linqi Song.)*

Xiaoming Xue, Yao Hu, and Linqi Song are with the Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China and also with the City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, China (e-mail: xminghsueh@gmail.com; y.hu@my.cityu.edu.hk; linqi.song@cityu.edu.hk).

Cuie Yang is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: cuieyang@outlook.com).

Kai Zhang is with the Civil Engineering School, Qingdao University of Technology, Qingdao 266520, China, and also with the School of Petroleum Engineering, China University of Petroleum (East China), Qingdao 266580, China (e-mail: zhangkai@upc.edu.cn).

Yiu-Ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China (e-mail: ymc@comp.hkbu.edu.hk).

Kay Chen Tan is with the Department of Computing, Hong Kong Polytechnic University, Hong Kong SAR, China (e-mail: kctan@polyu.edu.hk).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TEVC.2021.3133874.

Digital Object Identifier 10.1109/TEVC.2021.3133874

I. INTRODUCTION

EVOLUTIONARY algorithm (EA) is a population-based optimization approach based on the Darwinian theory [1]. For a given optimization task, EAs attempt to obtain a few high-quality solutions via appropriate solution representation and evolution mechanisms [2]. Over the past decades, EAs have been successfully applied to various real-world applications, including aerodynamic design [3], chemical processing [4], oil-gas field development [5], neural architecture search [6], and vehicle routing problem [7]. However, EAs always need a sufficient number of evaluations to achieve a satisfactory performance, which limits their applications to computationally expensive problems.

A number of techniques [8] have been proposed to reduce the computational burden of EAs, such as parallel or distributed computation [9], [10], evaluation relaxation [11], [12], hybridization [13], prior information incorporation [14], and learning from experience [15], [16]. Among them, knowledge transfer-based methods [17]–[21], known as evolutionary transfer optimization (ETO) [22], have attracted increasing research interests in recent years. In this study, we focus on a special case of ETO, evolutionary sequential transfer optimization (ESTO), which transfers knowledge from solved source optimization tasks to a target task to be optimized.

The seeding technique is a straightforward approach to ESTO that directly transfers solutions from source tasks. For example, the work in [14] and [23] initializes the new population of the target task by using good solutions from the source tasks. In this way, the search efficiency on the target task can be significantly improved when the target and source tasks share the similar optimal solutions. However, these methods may mislead the search if the optimal solutions differ a lot. To alleviate this issue, a number of dynamic injection methods have been presented [15], [24], which periodically put source solutions in the target population. To increase positive knowledge transfer, adaptation-based transfer methods have been developed to project source solutions to the target task [25]. For instance, high-order information that reflects the latent synergies between tasks is discovered and transferred in combinatorial optimization problems [26], [27] and estimation distribution algorithms [28]–[30]. More generally, in continuous optimization problems, an autoencoding evolutionary search (AES) paradigm has been proposed to conduct knowledge transfer across heterogeneous tasks [17]. It employs a single-layer denoising autoencoder to bridge the gap between tasks so that the source solutions can be well adapted to the target task. Following this framework, a selection strategy is proposed to address the multisource problem [31], which aims to select useful source tasks to reduce the risk of negative transfer.

Even though the existing ESTO methods have achieved great successes, to the best of our knowledge, they mainly focus on objective-homogeneous problems, where the source and target tasks share a similar number of objectives. In real-world applications, optimization problems are usually modeled with varying numbers of objectives depending on specific problem settings. A more practical scenario is that the source and the target tasks can be various among single-objective optimization problems (SOPs), multiobjective optimization problems (MOPs), and many-objective optimization problems (MaOPs), referred to as objective-heterogeneous ESTO problems. The objective-heterogeneous scenario is more challenging than the objective-homogeneous one due to the diverse search and objective spaces between the tasks. In this case, decision variable analysis [32], [33] is a promising technique for handling the objective-heterogeneous problems in a consistent way. Specifically, decision variables of MOPs and MaOPs can generally be classified into two categories, i.e., convergence-related variables (CVs) and diversity-related variables (DVs), according to their control properties [34], [35]. CVs are likely to converge to a single point as they control the distance between the population and the Pareto set (PS). In contrast, DVs are the root of objective conflicts and are encouraged to spread in the search space [36]. In this respect, decision variables of SOPs can be treated as CVs. Due to the dissimilarity between DVs and CVs, the past search experience of CVs cannot be properly harnessed to assist the optimization for DVs and vice versa. Therefore, the previous ESTO methods may hinder positive knowledge transfer across SOPs, MOPs, and MaOPs because they preserve much different number of DVs and CVs. Besides, these ESTO approaches often neglect fitness information during the adaptation process. The fitness information represents solution qualities, which can be utilized to facilitate the rank consistency between the source and target solutions in the learned mapping. Such consistency allows the good source solutions to be effectively transferred to the target problem.

Keeping the above in mind, this article proposes a decision variable analysis-based ESTO (DVA-ESTO), which aims to enhance positive knowledge transfer among problems ranging from SOPs to MaOPs. First, DVA-ESTO employs the decision variable analysis technique [33] to divide decision variables of MOPs and MaOPs into two groups, CVs and DVs. Second, DVA-ESTO transfers the knowledge of CVs and DVs independently from source tasks to the target task. The knowledge of CVs is used to accelerate the optimization of the target CVs, while the knowledge of DVs is encouraged to refine the distribution of the target DVs. The contributions of this article are summarized as follows.

1) This article attempts to address objective-heterogeneous problems, including SOPs, MOPs, and MaOPs, which greatly extends the application of ESTO in real-world problems.
2) We propose a new approach to transfer CVs and DVs separately based on decision variable analysis technique, which reduces negative knowledge transfer.
3) We propose a convergence transfer module (CTM) for CVs to improve the convergence, and a diversity transfer module (DTM) for DVs to increase their diversity.

The remainder of this article is organized as follows. Section II briefly introduces the definition of ESTO as well as the AES paradigm and discusses the challenges of objective-heterogeneous problems. The details of the proposed DVA-ESTO algorithm are described in Section III. Section IV presents the detailed experimental settings, and Section V provides a comprehensive experimental study conducted on a wide range of objective-heterogeneous problems and a practical case. Finally, Section VI concludes this article and provides several future directions.

II. Preliminaries

A. Evolutionary Sequential Transfer Optimization

Without loss of generality, an optimization problem can be formulated as follows:

$$\begin{aligned} &\text{minimize} \;\; F(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \\ &\text{subject to} \;\; \mathbf{x} \in \Omega \end{aligned} \tag{1}$$

where $\mathbf{x}$ is a decision variable, Ω denotes the decision space (also referred to as feasible space when a certain number of constraints are considered), and $F : \Omega \rightarrow \mathbb{R}^m$ consists of m real-valued objective functions to be optimized.

In particular, $F(\mathbf{x})$ is known as an SOP when $m = 1$. In this case, an optimizer attempts to find an optimal solution with the minimum objective value. $F(\mathbf{x})$ is recognized as an MOP if $m = 2$ or 3, and it is defined as an MaOP if $m > 3$. In MOPs and MaOPs, solution $\mathbf{x}_1$ is called to dominate solution $\mathbf{x}_2$ if and only if $f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2)$, $i = 1, 2, \ldots, m$ and $F(\mathbf{x}_1) \neq F(\mathbf{x}_2)$. A point that cannot be dominated by any other solutions is denoted as a Pareto optimal solution, and all nondominated solutions form a Pareto optimal set (PS). The mapping of PS in the objective space is called Pareto optimal front (PF) [37].

ESTO methods aim to improve the efficiency of an evolutionary solver by transferring knowledge of solved source tasks. Suppose a knowledge base M_b contains search experience of $K - 1$ source tasks $F_1, F_2, \ldots, F_{k-1}$, ESTO methods leverage useful knowledge from M_b to accelerate the search process on the target task F_k [25]. With the past solving experience stored in M_b, a successful ESTO is expected to converge faster than the baseline method with no transfer.

B. Autoencoding Evolutionary Search

The AES [17] is an adaptation-based ESTO method that conducts knowledge transfer across continuous optimization

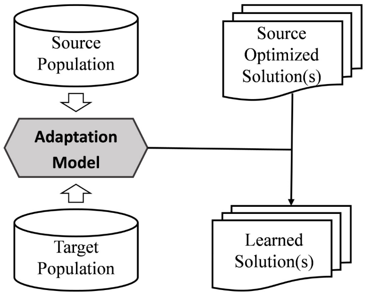


Fig. 1. Schematic diagram of the adaptation-based ESTO.

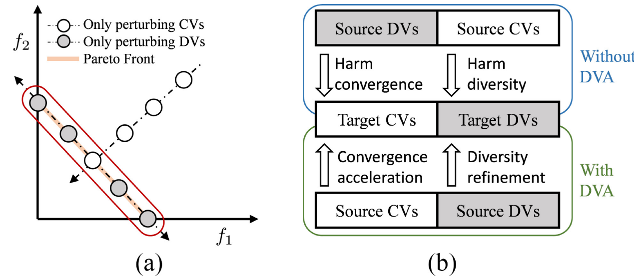


Fig. 2. DVA-based knowledge transfer: (a) illustration of CVs and DVs; (b) knowledge transfer with and without DVA.

problems via a single-layer denoising autoencoder. The autoencoder serves as an adaptation model to bridge the gap between source-target optimization instances. Fig. 1 shows the schematic diagram of the adaptation-based ESTO method.

Let P and Q represent the populations of the source and target tasks at a certain generation, respectively. Then, the mapping between the two tasks can be obtained through an autoencoder with P as inputs and Q as outputs. By simplifying the autoencoder into a single-layer version, a closed-form solution can be derived using the ordinary least square method [38]

$$M = \left(QP^T\right)\left(PP^T\right)^{-1}. \quad (2)$$

The learned mapping M can then project the elite source solutions to the target task. In the original AES paradigm, these projected source solutions are periodically injected into the target population. In the selection process, the injected solutions with competitive fitness will be reserved to form a new population, and those with low fitness will be automatically discarded. It can be seen that the adaptation model plays a crucial role in generating the transferred solutions. Most recently, Zhou *et al.* [39] proposed a kernelized AES (KAES) algorithm to model the nonlinear relationship between two heterogeneous instances. The source and target solutions used for adaptation are sorted by solution quality to facilitate the ordinal correlation toward effective knowledge transfer. In [40] and [41], the mapping is learned using a two-layer feed-forward neural network. However, the above sort-based solution pairing may encounter a severe issue called chaotic matching that obstructs the learning of intertask mapping [42], since a single fitness (or fitnesses in MOPs) always corresponds to an infinite number of solutions in the search space. Consequently, the sort-based solution pairing tends to match the source-target solutions chaotically and provide a degraded mapping, especially when the number of learning samples is large. A promising solution for addressing this issue is to relax the fitness ranks, which will be introduced in the rank adaptation model later.

C. Objective-Heterogeneous ESTO Problems

In a heterogeneous scenario, source and target problems may vary in many aspects, such as decision variable dimension, fitness landscape, objective dimension, etc. The decision space-related heterogeneities have been widely considered in the existing studies [17], [39], [43]. In contrast, few efforts have been devoted to addressing the objective-heterogeneous ESTO. Unlike the objective heterogeneity discussed in [44], the objective heterogeneity in this work refers to the intertask discrepancy with respect to the number of objectives. In real-world applications, optimization problems are usually modeled with varying numbers of objectives depending on specific problem settings. In this way, a knowledge base collected from source tasks may contain many solved instances with different numbers of objectives, including SOPs, MOPs, or even MaOPs. Without loss of generality, optimization problems with any number of objectives can be formulated by the form shown in (1). Given an n-dimensional problem with m objectives, the optimization task is to find an optimal manifold with at most $m - 1$ dimensions in the n-dimensional search space, where the manifold dimension (i.e., $m - 1$) depends on the conflicting relationship among objectives. Considering that variables may have different effects on the optimal manifold, Ma *et al.* [33] proposed to analyze the control property of decision variables and classify them into two categories: 1) DVs and 2) CVs. DVs control the diversity of solutions for depicting the PF, while CVs govern the distance between the obtained solutions and the true PS. In brief, the location and the shape of an optimal manifold are independently controlled by CVs and DVs, respectively. Variables that possess both the above two properties are recognized as mixed variables (MVs). More details about control property analysis can be referred to [33] and [35].

Next, we will show that the knowledge transfer between two optimization tasks without investigating their variable control properties may result in some undesired interferences between the source-target variables. For illustration, CVs and DVs with two knowledge transfer processes are shown in Fig. 2. In Fig. 2(a), two types of variables are represented as white and shaded circles, respectively. Optimization of the CVs is to search for an optimal solution that minimizes the distance to the PF, which is represented as a white circle in the PF. In contrast, solving the DVs aims to obtain a set of diversified solutions for achieving a good coverage over the true PF, which are shown as shaded circles in the PF. In a word, CVs are focal variables with an optimal solution, while DVs are distributional variables with multiple diversified solutions.

Two pairs of source and target instances for conducting knowledge transfers are shown in Fig. 2(b). In the upper part,

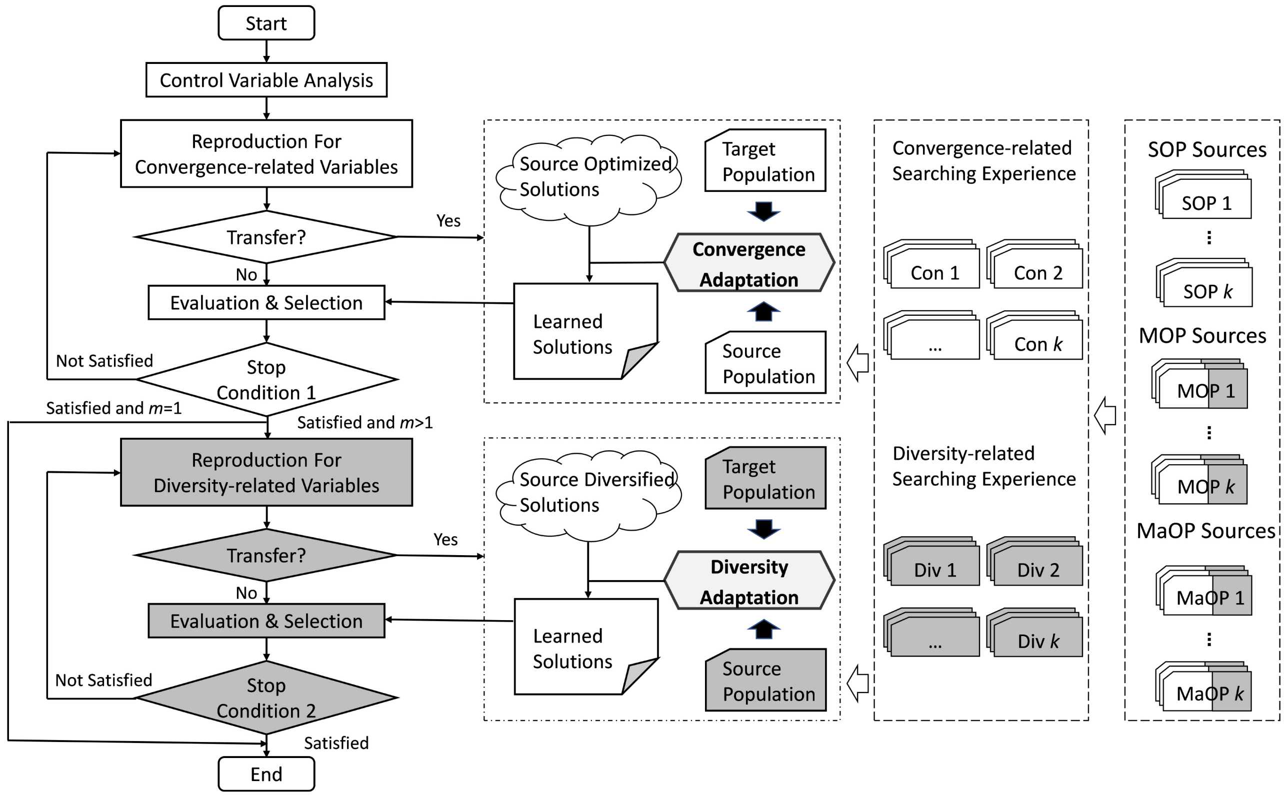


Fig. 3. Flow chart of the proposed DVA-ETSO.

the first half of the source decision variables are diversity-related, while the target variables at the corresponding position are convergence related. Without the decision variables analysis, improper transfer of source DVs for target CVs may result in convergence slow downs. Similarly, the transfer of source CVs that possess a convergence trend may harm the diversity of target DVs. Ideally, we hope that the convergence-related search experience on a source problem can be used to accelerate the convergence of target CVs, and the diversity-related experience can be utilized to refine the distribution of target DVs, as shown in the bottom part of Fig. 2(b).

However, without analyzing the control property of decision variables, it is hard to ensure such consistency of knowledge transfer. Thus, the improper transfer in Fig. 1(b) is more likely to occur if the control property of source-target variables is unclear. With the above issues in mind, we attempt to design a DVA-ESTO solver in this study. The benefit of using the control variable analysis is twofold: 1) an SOP can be treated as an MOP with only CVs. Such treatment allows us to understand SOPs, MOPs, and MaOPs from a consistent view and solve them in a unified paradigm and 2) The undesired interferences caused by the improper knowledge sharing between CVs and DVs can be greatly alleviated. To deal with the objective heterogeneity, multiobjective EA based on decomposition (MOEA/D) methods [37] can also transfer knowledge easily as they transform MOPs or MaOPs to a series of SOPs. However, the diversity in MOEA/D relies on the distribution of component SOPs, while the convergence depends on solving of component SOPs. For a given series of decomposed SOPs, how to balance the knowledge transfer among SOPs within a task and between tasks needs to be further investigated [45].

III. Proposed Algorithm

To conduct knowledge transfer between optimization problems with any number of objectives, we will develop a DVA-ESTO algorithm in this section, whose flow chart is illustrated in Fig. 3. Unlike the existing methods that learn a single mapping between the source and target tasks, DVA-ESTO transfers knowledge related to different types of variables independently. Specifically, given a knowledge base (solutions) from source tasks, DVA-ESTO separates the decision variables into CVs and DVs via decision variable analysis. Thereafter, it transfers knowledge of CVs and DVs via two phases: 1) convergence transfer phase and 2) diversity transfer phase. The convergence transfer aims to make the CVs converge faster using the past convergence searching experience. The purpose of diversity transfer is to depict the distribution of the DVs using the past distribution searching experience. It is noted that the later phase will not be executed for SOPs as they have no DVs. This condition is examined by the stop condition 1 shown in the flow chart. In this work, we propose to use different transfer mechanisms for CVs and DVs due to their distinct roles in a problem. Following the high-level structure of the proposed algorithm, we describe the two transfer methods in Sections III-A and III-B, respectively. Finally, Section III-C presents the detailed algorithmic implementation.

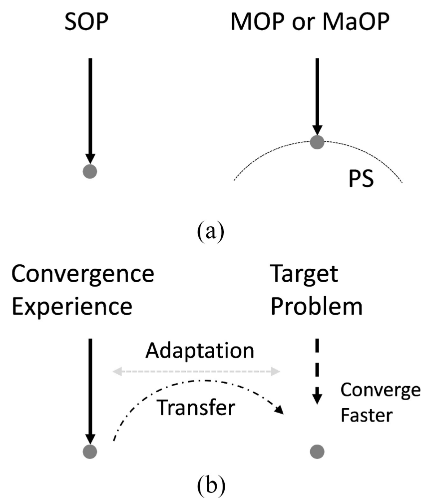


Fig. 4. Illustrations of the convergence processes and their experience reuse: (a) convergence processes of SOP, MOP, and MaOP under a specific objective preference; (b) accelerating convergence of a target problem using the past searching experience.

A. Convergence Transfer Module

In SOPs, CVs govern the convergence of the current population toward the optimum. While in MOPs and MaOPs, these variables control the distance between the current solution set and the PS. Particularly, given a specific solution with DVs fixed, the distance measures the convergence of an aggregated function toward a point in the Pareto set. An example to illustrate the above convergence processes is shown in Fig. 4(a), where the black arrows represent the convergence paths with increased fitness (or aggregated fitness), and the gray dots denote an optimum of SOP or a Pareto solution of MOP (or MaOP). Thus, optimization of the CVs is to search for an optimal solution to achieve the highest fitness (or aggregated fitness). To achieve knowledge transfer, convergence transfer optimization attempts to accelerate the convergence of target CVs using the past convergence searching experience, as shown in Fig. 4(b). Now, two key issues are required to be addressed under the adaptation-based ESTO framework. One is how to adapt the source instances from the perspective of convergence transfer, and the other is how to obtain the learned solutions for guiding the target instance after the adaptation.

1) Convergence Mapping Learning: To adapt source instances to the target task, we first learn a mapping between the source and target tasks according to the collected solutions. Suppose individuals and corresponding fitness values of the source and target tasks are $\{(\mathbf{x}_i^s, y_i^s)|_{i=1}^{n_s}\}$ and $\{(\mathbf{x}_i^t, y_i^t)|_{i=1}^{n_t}\}$, respectively. In a heterogeneous scenario, apart from the landscape discrepancy, the dimension d_s of the source instance can be different from that of the target instance d_t. Let the individual be a column vector and

$$X_s = \left[\mathbf{x}_1^s, \ldots, \mathbf{x}_{n_s}^s\right], \quad X_t = \left[\mathbf{x}_1^t, \ldots, \mathbf{x}_{n_t}^t\right] \tag{3}$$

and let the standardized matrices of X_s and X_t be $\tilde{X}_s$ and $\tilde{X}_t$, where individuals are standardized to have zero mean and unit standard deviation.

To overcome the problem gap, we project the source and target samples into a common low-dimensional space $\mathcal{V} \subset \mathbb{R}^l$

$$\mathbf{u}_i^s = P_s^T \tilde{\mathbf{x}}_i^s \in \mathbb{R}^l, \quad \mathbf{u}_i^t = P_t^T \tilde{\mathbf{x}}_i^t \in \mathbb{R}^l \tag{4}$$

where $P_s \in \mathbb{R}^{d_s \times l}$ and $P_t \in \mathbb{R}^{d_t \times l}$. The matrices of source and target projected samples are represented as $U_s \in \mathbb{R}^{l \times n_s}$ and $U_t \in \mathbb{R}^{l \times n_t}$, respectively.

The moment matching is a common approach since the moment information is a good measurement for characterizing a distribution [46]. According to the Hausdorff moment problem [47], the collection of all the moments uniquely determines a probabilistic distribution on a bounded interval. Thus, adaptation between two unknown distributions can be realized by adapting all their moment information (i.e., zero-order to infinite-order moments). However, it is computationally impractical for considering an infinite number of moments. In this study, we employ the first two moments (i.e., mean and covariance) as main moment information to represent an individual distribution. For adaptation purpose, we hope the moment measurements of source-target individuals in the common space (i.e., $\{\mathbf{u}_i^s|_{i=1}^{n_s}\}$ and $\{\mathbf{u}_i^t|_{i=1}^{n_t}\}$) can be as close as possible. Thus, an objective function to be minimized for bridging the gap between the source-target individual distributions in the common space can be formulated as follows:

$$\arg\min_{P_s, P_t} \mathrm{I}(U_s, U_t) = \left\|\bar{\mathbf{u}}^s - \bar{\mathbf{u}}^t\right\|_F^2 + \left\|\frac{\bar{U}_s \bar{U}_s^T}{n_s} - \frac{\bar{U}_t \bar{U}_t^T}{n_t}\right\|_F^2 \tag{5}$$

where $\bar{u} = (1/n)\sum_i \mathbf{u}_i$. It can be easily proved that the first moments of $\{\mathbf{u}_i^s\}$ and $\{\mathbf{u}_i^t\}$ are zeros (i.e., $\bar{\mathbf{u}}^s = \bar{\mathbf{u}}^t = \mathbf{0}_{l\times 1}$), as the original sample data X_s and X_t are standardized. In this case, the optimization model in (5) can be rewritten as follows:

$$\arg\min_{P_s, P_t} \mathrm{I}(U_s, U_t) = \left\|\frac{P_s^T \tilde{X}_s \tilde{X}_s^T P_s}{n_s} - \frac{P_t^T \tilde{X}_t \tilde{X}_t^T P_t}{n_t}\right\|_F^2. \tag{6}$$

Meanwhile, we hope that the fitness ranks of the source and target tasks can be aligned in the common space. Thus, we incorporate the fitness rank into mappings to preserve fitness rank consistency between the source and target instances. In this way, the elite source solution(s) are more likely to be helpful to guide the convergence of the target CVs. First, the source and target data are classified into many rank-labeled individuals according to their fitness values. To this end, we sort the individuals based on fitness values and then divide them into several groups, where the individuals in each group have the same fitness rank. The rank labels of the source-target instances are represented as r_i^s and $r_i^t \in \{1, 2, \ldots, C\}$, where $r_i^s = j$ indicates that the fitness rank of ith source individual is j. Next, we define a variable set S_i used to store all the indexes of target solution(s) that have the same fitness rank with ith source individual. Therefore, the fitness rank adaptation can be achieved by minimizing the following fitness rank matching function:

$$\arg\min_{P_s, P_t} \mathrm{F}(U_s, U_t) = \frac{1}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j \in S_i} \left\|\mathbf{u}_i^s - \mathbf{u}_j^t\right\|_2^2. \tag{7}$$

By substituting the expanded form of u into (7), we have

$$\arg\min_{P_s,P_t} \mathrm{F}(U_s, U_t) = \frac{1}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j\in S_i} \left\| P_s^T \tilde{\mathbf{x}}_i^s - P_t^T \tilde{\mathbf{x}}_j^t \right\|_2^2. \quad (8)$$

To adapt the solution distributions and the fitness ranks simultaneously, a weighted model is given as follows:

$$\arg\min_{P_s,P_t} f_c(P_s, P_t) = \mathrm{I}(U_s, U_t) + \alpha \mathrm{F}(U_s, U_t) \quad (9)$$

where α is a tradeoff coefficient.

In fact, the problem here can be seen as a supervised domain adaptation task discussed in [46], where the individual solutions represent feature inputs and the fitness ranks correspond to class labels. Another application of this adaptation technique can be found in EMT problems [48], where class labels are determined by the Pareto domination relationship. Using the similar derivation in [46], we proved that the projections that minimize the objective function defined in (9) can be determined by the eigenvectors of a nonlinear eigenvalue decomposition problem.[1] With the obtained projections P_s and P_t, the source-target instances are adapted in the common low-dimensional space, which can be expressed as follows:

$$P_s^T \tilde{\mathbf{x}}^s = P_t^T \tilde{\mathbf{x}}^t. \quad (10)$$

A mapping to bridge the problem gap can be obtained as follows:

$$M_c = \left(P_t P_t^T\right)^{-1} P_t P_s^T \quad (11)$$

where M_c denotes the source-target mapping used for convergence adaptation.

2) Adaptation of Source Convergence Experience: Once the mapping is learned, we propose to select the best source solution(s) with optimal convergence state and project it into the target instance to accelerate the convergence. These elite source solutions can be randomly selected from the final population of the source instance

$$\mathbf{x}_i^{sc} = \mathrm{Pop}_{sc}^F(\mathrm{rand}), \quad i = 1, 2, \ldots, t \quad (12)$$

where $\mathbf{x}_i^{sc}$ denotes the ith selected source solution for convergence transfer, Pop_{sc}^F is the final source population of CVs, rand is a randomly generated exclusive index for retrieving source converged individuals, and t represents the number of solutions to be adapted and transferred.

Finally, multiple adapted source solutions to be injected into the target population can be obtained as follows:

$$\tilde{\mathbf{x}}_i^{\mathrm{tranc}} = M_c \tilde{\mathbf{x}}_i^{sc}, \quad i = 1, 2, \ldots, t \quad (13)$$

where $\tilde{\mathbf{x}}_i^{\mathrm{tranc}}$ denotes the ith adapted solution to be injected into the target population for convergence acceleration, $\tilde{\mathbf{x}}_i^{sc}$ represents the ith standardized source solution. It is noteworthy that the adapted solution obtained here should be standardized back to the original scale for further evaluation.

Algorithm 1 shows the implementation of the CTM. First, the source and target data are collected from a certain generation. Then, the adaptation model in (9) is turned into a nonlinear eigenproblem and solved by an iterative procedure, as shown in lines 3–8. Finally, a number of solutions to be transferred can be obtained by adapting a selected source solution set. The obtained transferred solutions can then be injected into the target population to undergo the selection process, as shown in the upper part of Fig. 3. If the target task is an SOP, then all the available computational resources will be consumed in this convergence transfer optimization phase; otherwise, a certain percentage of the recourse should be allocated to handle the DVs.

[1]A detailed derivation is provided in Section S-I of the supplementary document accompanying this article.

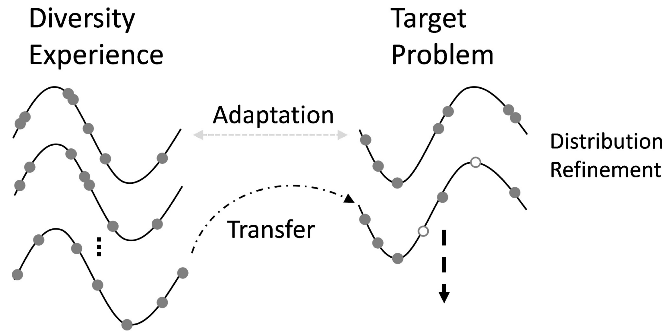


Fig. 5. Illustrations of diversity refining process and its experience reuse.

Algorithm 1: *ConvergenceTransfer*$(X_t, y^t, X_s, y^s, X_{sc}^F)$

Input: X_t (current target population), y^t (fitness of the target population), X_s (source population at the current generation), y^s (fitness of the source population), X_{sc}^F (a selected source solution set containing t solutions)
Output: X_{tranc} (solutions to be transferred for convergence acceleration)

```
// Data preparation
```
1 $[\tilde{X}_t, \tilde{X}_s, \tilde{X}_{sc}^F] \leftarrow$ Standardize the data X_t, X_s, and X_{sc}^F;
2 $[r^t, r^s] \leftarrow$ Get fitness ranks based on y^t and y^s;
```
// Source-target mapping conduction
```
3 Calculate the coefficient matrices used for adaptation;
4 **while** *stopping conditions are not satisfied* **do**
5 Calculate the loss $f_c(P_s, P_t)$ in Eq. (9);
6 $[\mathrm{V}, \Lambda] \leftarrow$Solve the corresponding eigenproblem;
7 Sort V in an ascending order based on the eigenvalues Λ;
8 Update (P_s, P_t) using the first l sorted eigenvectors;
9 $M_c \leftarrow$ Calculate the mapping using Eq. (11);
```
// Source solution adaptation
```
10 $\tilde{X}_{tranc} \leftarrow$ Obtain the adapted solution using Eq. (13);
11 $X_{tranc} \leftarrow$ Standardize $\tilde{X}_{tranc}$ back to the original scale;
12 **return** X_{tranc}.

B. Diversity Transfer Module

As mentioned earlier, DVs govern the shape of a PS. In other words, the diversity of the PS is controlled by the DVs. Thus, a number of well-distributed realizations of the DVs are required to achieve diverse solutions along the PF. Unlike the CTM, the proposed DTM transfers knowledge between DVs, so that the past diversity searching experience can be used to better refine the distribution of target DVs. As shown in Fig. 5, PSs of the source and target tasks are represented as solid lines, and the solutions are marked as gray circles. To achieve a better coverage over the true PS of the target instance, several unexplored promising regions are expected to be investigated by the transferred solutions, which are denoted as white circles in Fig. 5. Similarly, following the adaptation-based ESTO framework, we need to address two key issues: 1) adapting the source-target DVs and 2) obtaining the adapted source solutions for better refining the target DVs.

1) Diversity Mapping Learning: Unlike CVs that possess a single optimal solution with the highest fitness, DVs are distributional variables that contain multiple solutions for depicting the low-dimensional Pareto manifold. The word "distributional" here is used to describe the optimal (or final) state of decision variables. In this way, a number of diversified solutions for the DVs are required to approximate the true PS. To adapt such distributional variables, we propose to bridge the distribution gap by adapting the first- and second-order moment information of the source-target instances. Such an adaptation mapping can be analytically derived from the source-target Gaussian representation models [42], because the first- and second-order information can be fully considered by a multivariate Gaussian distribution. It can be seen that the diversity adaptation here can be regarded as the CTM with only the solution distribution adaptation term in (6).

We choose to remove the fitness adaptation term since the distribution of DVs will not degenerate according to the fitness information during the optimization process. Given the source-target data X_s and X_t used for adaptation, their representation models are represented as $\mathcal{N}(\mu_s, \Sigma_s)$ and $\mathcal{N}(\mu_t, \Sigma_t)$, respectively. A closed-form solution of the adaptation mapping $\mathbf{x}^\top A + \mathbf{b}$ is given as follows:

$$M_d = [A, \mathbf{b}] = \left[L_s L_t^{-1}, \mu_t^\top - \mu_s^\top L_s L_t^{-1}\right] \tag{14}$$

where L_s and L_t denote the lower triangular matrixes obtained from the Cholesky decomposition on the source and target inverse covariance matrixes, respectively.

Particularly, (14) does not work on problems with distinct dimensions (i.e., $d_s \neq d_t$). In this case, a randomly generated projection matrix $R_{d_s \times d_t}$ can be employed to project the source data to make it match with the target instance, since the source data structure can normally be well maintained using such a projection according to the Johnson–Lindenstrauss lemma [49]. The projected source data used for adaptation is denoted as $X_{sp} = X_s R$. Then, the source representation model is calculated based on X_{sp}. Accordingly, the adaptation mapping is modified as follows:

$$M'_d = RM_d. \tag{15}$$

2) Adaptation of Source Diversity Experience: After the adaptation, the DVs of source and target tasks share similar distributions in the mapped space. To better refine the distribution of the target DVs, we propose to project the well-distributed source solutions at the final generation into the target task and retain part of them for better depicting the distribution, as illustrated in Fig. 5. First, the selected source solutions are adapted by the obtained mapping M_d (or M'_d when $d_s \neq d_t$)

$$\tilde{\mathbf{x}}_i^{\text{ada}} = M_d \tilde{\mathbf{x}}_i^{sd}, \quad i = 1, 2, \ldots, N_s \tag{16}$$

Algorithm 2: *DiversityTransfer*(X_t, X_s, X_{sd}^F)

Input: X_t (current target population), X_s (current source population), X_{sd}^F (the final source population)
Output: X_{trand} (solutions to be transferred for diversity refinement)

```
// Source-target mapping conduction
```
1 Build the source-target representation models;
2 $M_d \leftarrow$ Calculate the mapping based on Eq. (14) or Eq. (15);
```
// Source solution adaptation
```
3 $\tilde{X}_{ada} \leftarrow$ Obtain the adapted solution using Eq. (16);
4 $X_{ada} \leftarrow$ Standardize $\tilde{X}_{ada}$ back to the original scale;
5 $[\mathrm{E_{tran}}, \mathrm{E_t}] \leftarrow$ Manifold embedding of X_t and X_{ada};
6 Calculate the distance vector d using Eq. (17);
7 $X_{trand} \leftarrow t$ adapted solutions with larger geodesic distances;
8 **return** X_{trand}.

where $\tilde{\mathbf{x}}_i^{\text{ada}}$ denotes the ith adapted solution, $\tilde{\mathbf{x}}_i^{sd}$ represents the ith standardized source solution stored in the final generation, and N_s is the source population size.

Then, a portion of transferred solutions that are further away from the known target solutions along the Pareto manifold (i.e., larger geodesic distance) will be retained for further evaluation since these solutions tend to investigate some unexplored regions for better depicting the manifold. With the Isomap algorithm [50], the geodesic distances between the transferred solutions and the target individuals can be calculated in an $(m-1)$-dimensional Euclidean space. In the Euclidean space, the transferred and target solutions are denoted as $\{\mathbf{e}_i^{\text{tran}}|_{i=1}^{N_s}\}$ and $\{\mathbf{e}_i^{t}|_{i=1}^{N_t}\}$, respectively, where N_t represents the target population size. The minimum geodesic distance between each transferred solution and the target individuals can be calculated as follows:

$$d_i = \min_{j \in \{1,\ldots,N_t\}} \left\| \mathbf{e}_i^{\text{tran}} - \mathbf{e}_j^{t} \right\|, \quad i = 1, 2, \ldots, N_s \tag{17}$$

where d_i denotes the shortest geodesic distance from the ith transferred solution to all the known target individuals in the Euclidean space.

Finally, the transferred solutions are sorted in a descending order based on the distance vector d, and the first t sorted solutions are selected to be injected into the target diversity-related population. Algorithm 2 shows the detailed implementation of the proposed DTM. Unlike the CTM that directly transfers the projected solutions via injection, the diversity transfer employs an additional prescreening module to identify promising adapted solutions using the manifold embedding, as shown in lines 5–7.

C. Algorithmic Implementation

The algorithmic implementation of the overall DVA-ESTO algorithm is presented as Algorithm 3.

Algorithm 3: Decision Variable Analysis-Based ESTO

Input: $\Omega_t \rightarrow \mathbb{R}^{m_t}$ (target instance), $\{(X_s^i, F_s^i)|_{i=1}^{g}\}$ (source data)
Output: $\mathbf{x}^{best}$ or P (Optimized solution or solution set)
// Data preparation
1 $[P_s^c, F_s^c, P_s^d] \leftarrow$ Analyze the source optimization history;
2 $[I^c, I^d] \leftarrow$ Variable analysis on the target problem;
// Convergence transfer phase
3 $FEs = 0$;
4 $P^c \leftarrow$ Initialization for the convergence-related variables;
5 $FEs \leftarrow FEs + PopSize$;
6 **while** $FEs < \gamma \cdot FEsMax$ **do**
7     $O^c \leftarrow$ Generation of convergence-related offspring;
8   **if** *transfer condition is met* **then**
9         $T^c = ConvergenceTransfer(P^c, y^c, P_s^c, y_s^c, S^c)$;
10         $O^c \leftarrow$ Inject T^c into the offspring;
11     $P^c \leftarrow$ Environmental selection on $P^c \cup O^c$;
12     $FEs \leftarrow FEs + PopSize$;
13 **if** $FEs \geq \gamma \cdot FEsMax$ **then**
14   **return** the best solution $\mathbf{x}^{best}$;
// Diversity transfer phase
15 $P^d \leftarrow$ Initialization for the diversity-related variables;
16 $FEs \leftarrow FEs + PopSize$;
17 **while** $FEs < FEsMax$ **do**
18     $O^d \leftarrow$ Generation of diversity-related offspring;
19   **if** *transfer condition is met* **then**
20         $T^d = DiversityTransfer(P^d, P_s^d, S^d)$;
21         $O^d \leftarrow$ Inject T^d into the offspring;
22     $P^d \leftarrow$ Environmental selection on $P^d \cup O^d$;
23     $FEs \leftarrow FEs + PopSize$;
24 **return** the obtained solution set $P = P^c \cup P^d$;

First, the past searching experience on general optimization problems are classified into two categories: 1) convergence-related and 2) diversity-related experience. The source population data on CVs and the associated objective values are denoted as P_s^c and F_s^c, respectively, while the populations on DVs are represented as P_s^d. Given a target optimization task, the indexes of its CVs and DVs can be obtained as I^c and I^d using the DVA method [33]. Herein, a predefined parameter γ is used to control the ratio of computational resources assigned to the convergence transfer phase. Particularly, γ will be set to 1 if the target instance is an SOP (i.e., I^d is empty); otherwise, a predefined γ that leaves a certain percentage of resources for solving DVs of MOPs and MaOPs will be adopted.

In the convergence transfer phase, a population-based SOP solver equipped with the CTM is employed to optimize the CVs, as shown in lines 3–12. Solutions for knowledge transfer (i.e., S^c) are randomly selected from the final convergence-related population stored in P_s^c. Subsequently, the diversity transfer phase will be launched if there are some DVs remained to be solved. A population-based MOP solver together with the DTM is used to solve the DVs, as shown in lines 15–23. Here, solutions selected for knowledge transfer (i.e., S^d) are the final well-distributed individuals stored in P_s^d. Once the overall optimization is completed, the population data on the given target instance can be added to the database to enrich the past searching experience further.

IV. Experimental Settings

In this section, test problems, performance metrics, state-of-the-art ESTO algorithms used for comparison, and the associated parameter settings are presented in detail.

A. Test Problems and Performance Metrics

1) Test Problems: To the best of our knowledge, there is not yet a systematic benchmark suite for investigating the efficacy of ESTOs on optimization problems with any number of objectives. With this in mind, a test suite that contains a variety of optimization instances with strong problem heterogeneities is formulated in this work. Due to the page limit, the detailed descriptions of the test problems are provided in the supplementary document.[2] Overall, there are in total 18 problems with six SOPs, six MOPs, and six MaOPs in the conducted ESTO test suite, which possess strong heterogeneities with respect to both the decision and objective spaces. By treating SOPs as general MOPs with zero-dimensional Pareto manifolds, it can be seen that the optimal manifolds of these 18 optimization problems have distinct positions and (or) shapes. For solving a specific problem in the above ESTO benchmark, the source database used to store the past searching experience will exclude this problem.

2) Performance Metrics: For SOPs, the performance metric is simply the fitness value. For MOPs or MaOPs, different metrics can be adopted to evaluate the quality of obtained Pareto sets. In the following experimental part, we need to validate the efficacy of the proposed CTM and DTM independently. The generational distance (GD) [51] is employed to evaluate the CTM, while the pure diversity (PD) [52] is used to investigate the DTM. The overall DVA-ESTO algorithm against other ESTOs is evaluated using the inverted GD (IGD) [53].

B. ESTO Algorithms Used in the Experimental Studies

In this work, the canonical EA, NSGA-II [51], and NSGA-III [54] algorithms are selected as the baseline solvers for SOPs, MOPs, and MaOPs, respectively. Two baseline ESTOs and three advanced adaptation-based ESTOs used for comparison are considered as follows.

1) *SOLVER:* Baseline solver with no transfer.
2) *SOLVER-R:* Injection of randomly generated solutions;
3) *SOLVER-ST [31]:* Injection of adapted solutions provided by the linear single-layer autoencoder using the nearest source selection strategy.
4) *SOLVER-KAE [39]:* Injection of adapted solutions provided by the kernelized single-layer autoencoder.
5) *SOLVER-NDA [40]:* A probabilistic model-based sequential transfer algorithm with nonlinear domain adaptation.

A baseline variable analysis-based solver without the transfer components contains two optimization phases. In this work, the GA optimizer is used in the convergence optimization phase, while the NSGA-II and NSGA-III solvers are adopted in the diversity optimization phase for MOPs and MaOPs,

[2]See Section S-II-A of the supplementary document.

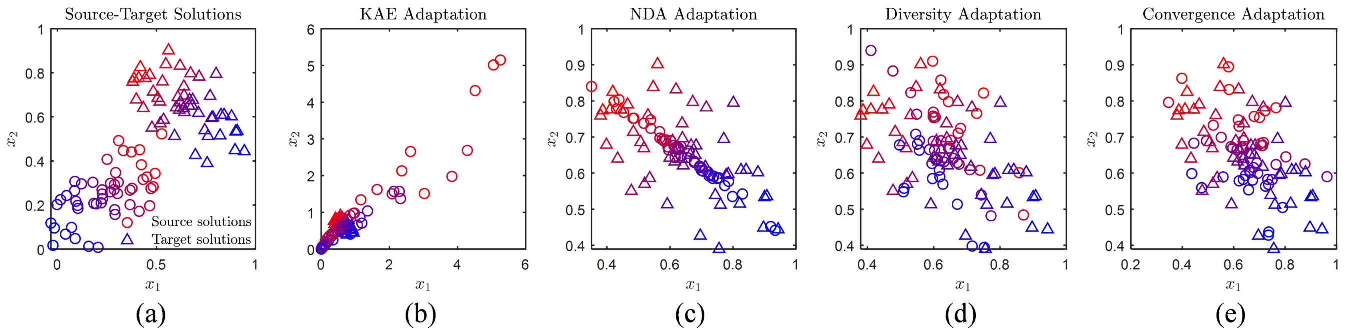


Fig. 6. Adaptation behaviors of KAE, NDA, and the proposed models on a 2-D case: (a) source-target solutions; (b) KAE adaptation; (c) NDA adaptation; (d) diversity-related adaptation; (e) convergence-related adaptation.

respectively. Two ESTOs extended from the baseline variable analysis-based solver are 1) DVA-C: injection of solutions provided by the CTM only and 2) DVA-D: injection of solutions provided by the DTM only.

C. Parameter Settings

For the sake of fairness, evolutionary operators and associated parameter settings are kept the same for all the methods. Some of these general parameters are outlined as follows.

1) Population size (N): 50.
2) Maximum function evaluations for target tasks: 10 000.
3) Maximum function evaluations for source tasks: 20 000.
4) Evolutionary operators for all the algorithms.
   a) Solutions are scaled into a standardized range.
   b) SBX crossover with probability $p_c = 1$ and distribution index $\eta_c = 15$.
   c) Polynomial mutation with probability $p_m = 1/d$ and distribution index $\eta_m = 15$.

The settings of knowledge transfer-related parameters are given as follows.

1) The transfer condition: certain generation gap ($G = 1$).
2) The number of injected solutions: $t = m$.
3) For the proposed DVA-based methods.
   a) Variable analysis method: DVA [33].
   b) The parameter for resource allocation (γ): 0.95.
   c) Dimensionality of the common space (l): 4.
   d) The contribution parameter in the convergence adaptation model (α): 0.1.
   e) The maximum number of iterations for solving the nonlinear eigenproblem: 100.

V. Experimental Results and Analysis

In this section, the efficacy of the proposed method against a number of state-of-the-art ESTO solvers is investigated in detail. First, the adaptation behaviors of a number of adaptation models used in the existing ESTO methods are analyzed on a toy example. Then, the superiorities of the overall DVA-ESTO algorithm and the two transfer modules are verified. In addition, an extended MOP test suite containing problems with more complex variable control property is considered. Finally, a practical production planning optimization problem for mineral processing is considered.

A. Investigation of Adaptation Behaviors

As discussed earlier, the adaptation model plays a crucial role in generating high-quality solutions to be transferred to a target problem. In this work, two different models with distinct adaptation goals are proposed for convergence transfer and diversity transfer, respectively. To investigate their adaptation behaviors, we compare them with two advanced methods on a 2-D toy case containing a pair of heterogeneous source-target instances.

Fig. 6(a) shows the source and target solutions with known fitness values whose magnitudes are represented as gradually varied colors from blue to red. The source solutions are represented as circles, while the target solutions are shown as triangles. A solution with redder color possesses higher fitness. Consequently, the source instance tends to search northeastward while the target instance prefers moving toward the northwest. The source and the target tasks are much different with respect to the individual distribution and the fitness rank. In such a scenario, task adaptation is required to bridge the gap between the source-target instances.

Fig. 6(b)–(e) plot the adapted source solutions provided by KAE, NDA, the diversity adaptation model in Section III-B, and the convergence adaptation model in Section III-A, and the target solutions. As shown in Fig. 6(b), the KAE method tends to provide an uncanny mapping which is difficult to interpret. Using the same sort-based solution pairing manner, the NDA model successfully adapts the source-target fitness preferences, as shown in Fig. 6(c). The elite source solutions with darker colors overlap the high-quality target solutions, while the source-target bad solutions match as well. However, the adapted source solutions tend to degrade on a line in the target domain, especially when the number of learning samples is large. This issue is called mapping degradation in [42]. In contrast, the distribution adaptation model used for diversity adaptation adapts the source-target solution distributions very well. The first and second moments of the source-target instances are well aligned after the adaptation, as shown in Fig. 6(d). This indicates that the proposed diversity adaptation model works well on adapting the moment information of distributional DVs. Unlike the distribution adaptation model that stretches the source distribution to match the target distribution, the adaptation model used for convergence transfer attempts to adapt the source-target solution distributions

TABLE I
AVERAGE OBJECTIVE OR IGD VALUES AND STANDARD DEVIATION OBTAINED BY SOLVER, SOLVER-R, SOLVER-ST, SOLVER-KAE, SOLVER-NDA, DVA, AND DVA-ESTO ON THE ESTO TEST SUITE. THE HIGHLIGHTED ENTRIES ARE SIGNIFICANTLY BETTER (WILCOXON RANK-SUM TEST WITH HOLM p-VALUE CORRECTION, $\alpha = 0.05$)

Category	Problem	EA	EA-R	EA-ST	EA-KAE	EA-NDA	DVA (EA)	DVA-ESTO
SOPs	SOP 1	6.75e+4±9.73e+3	7.00e+4±7.48e+3	7.08e+4±8.98e+3	6.93e+4±8.24e+3	6.61e+4±9.13e+3	6.75e+4±9.73e+3	5.37e+4±7.76e+3
	SOP 2	9.79e+5±7.82e+4	9.79e+5±8.96e+4	1.01e+6±9.63e+4	9.98e+5±6.77e+4	9.51e+5±1.08e+5	9.79e+5±7.82e+4	6.92e+5±5.60e+4
	SOP 3	3.56e+4±3.10e+3	3.55e+4±3.41e+3	3.62e+4±3.22e+3	3.65e+4±3.03e+3	3.37e+4±3.78e+3	3.56e+4±3.10e+3	2.47e+4±2.84e+3
	SOP 4	8.18e+0±1.14e+0	8.22e+0±9.58e-1	8.23e+0±1.10e+0	8.34e+0±8.73e-1	7.88e+0±1.09e+0	8.18e+0±1.14e+0	8.00e+0±1.32e+0
	SOP 5	5.88e+1±4.92e+0	5.81e+1±5.06e+0	5.97e+1±4.82e+0	5.85e+1±4.02e+0	5.24e+1±4.66e+0	5.88e+1±4.92e+0	3.79e+1±4.83e+0
	SOP 6	3.32e+3±4.16e+2	3.36e+3±3.66e+2	3.32e+3±4.27e+2	3.33e+3±3.96e+2	3.20e+3±4.18e+2	3.32e+3±4.16e+2	2.17e+3±3.41e+2
		NSGA-II	NSGA-II-R	NSGA-II-ST	NSGA-II-KAE	NSGA-II-NDA	DVA	DVA-ESTO
MOPs	MOP 1	3.46e+4±5.26e+3	3.25e+4±4.39e+3	3.64e+4±4.59e+3	3.32e+4±5.22e+3	.73e+4±4.01e+3	2.26e+4±4.89e+1	2.07e+4±2.26e+3
	MOP 2	8.98e+3±1.44e+3	1.02e+4±1.42e+3	1.32e+4±3.21e+3	1.11e+4±2.89e+3	9.52e+3±2.08e+3	1.83e+3±4.05e+2	1.56e+3±5.79e+2
	MOP 3	8.17e+0±1.66e+0	8.14e+0±1.29e+0	8.36e+0±1.84e+0	8.63e+0±1.20e+0	7.12e+0±1.19e+0	4.75e+0±1.97e-1	4.76e+0±2.37e-1
	MOP 4	1.92e+3±3.77e+2	2.53e+3±5.96e+2	3.64e+3±1.02e+3	3.12e+3±1.24e+3	2.67e+3±7.12e+2	1.02e+3±1.53e+2	6.64e+2±8.60e+1
	MOP 5	1.77e+7±2.22e+6	1.90e+7±3.22e+6	1.70e+7±2.71e+6	1.70e+7±2.80e+6	1.40e+7±1.75e+6	1.44e+7±2.13e+4	1.16e+7±1.37e+6
	MOP 6	7.40e-1±2.50e-1	7.78e-1±2.92e-1	7.86e-1±3.22e-1	7.39e-1±2.60e-1	8.48e-1±2.96e-1	7.66e-1±1.28e-1	7.04e-1±1.81e-1
		NSGA-III	NSGA-III-R	NSGA-III-ST	NSGA-III-KAE	NSGA-III-NDA	DVA	DVA-ESTO
MaOPs	MaOP 1	1.07e+5±1.93e+4	1.11e+5±1.59e+4	1.61e+5±5.39e+4	1.64e+5±7.37e+4	1.26e+5±1.74e+4	7.09e+4±1.78e+4	4.57e+4±1.07e+4
	MaOP 2	2.70e+4±4.11e+3	2.47e+4±2.10e+3	4.23e+4±1.12e+4	3.51e+4±9.57e+3	2.48e+4±3.82e+3	1.43e+4±2.83e+3	8.82e+3±2.34e+3
	MaOP 3	1.64e+1±6.70e-1	1.65e+1±5.54e-1	1.69e+1±1.02e+0	1.68e+1±9.27e-1	1.71e+1±5.53e-1	1.18e+1±6.94e-1	1.10e+1±8.51e-1
	MaOP 4	8.09e+3±1.84e+3	9.45e+3±1.14e+3	1.83e+4±2.70e+3	2.02e+4±3.30e+3	1.22e+4±2.10e+3	2.20e+3±2.01e+2	1.47e+3±8.07e+1
	MaOP 5	1.04e+8±1.15e+7	1.00e+8±7.36e+6	1.16e+8±2.33e+7	1.13e+8±1.22e+7	8.71e+7±1.11e+7	4.67e+7±4.74e+6	3.13e+7±5.83e+6
	MaOP 6	1.01e+2±2.37e+1	9.74e+1±1.15e+1	1.01e+2±2.41e+1	1.16e+2±3.19e+1	1.03e+2±1.43e+1	2.31e+1±4.00e+0	2.20e+1±3.87e+0

and fitness preferences simultaneously. As illustrated from Fig. 6(e), the convergence adaptation model obtains a proper rotation mapping to achieve the above two adaptation goals simultaneously. For CVs, the high-quality source solutions can be safely transferred to the target population by preserving such fitness rank consistency, so as to accelerate the convergence of target CVs. In summary, the proposed adaptation models perform well on adapting CVs and DVs. The adaptation goals presented in Sections III-A and III-B can been successfully achieved.

B. Performance Comparison With the State-of-the-Art

In this section, the superiority of the overall DVA-ESTO algorithm against other state-of-the-art ESTOs on the ESTO test suite is investigated. In the overall DVA-ESTO algorithm, both the CTM and DTM will be activated. Table I presents the objective or IGD mean and standard deviation obtained by SOLVER, SOLVER-R, SOLVER-ST, SOLVER-KAE, SOLVER-NDA, DVA, and DVA-ESTO over 30 independent runs on the ESTO test suite. Superior performance is highlighted in bold based on the Wilcoxon rank-sum test with a 95% confidence level. An instance without any highlighted entry indicates that all the ESTO solvers show comparable performance in solving the problem.

From the table, we can see that the proposed DVA-ESTO algorithm outperforms the compared methods on most of the test problems, ranging from SOPs to MaOPs. The past searching experience can be effectively reused to assist the optimization of an interested target instance, whether it is an SOP, MOP, or MaOP. More importantly, as the new instances at hand to be solved, they can be used to enrich the past searching experience further. Next, to find out the effects of CTM and DTM on the overall DVA-ESTO algorithm, we investigate their performance independently.

C. Effectiveness of the Proposed CTM

The proposed DVA-ESTO algorithms with no transfer and only the convergence transfer are termed as DVA and DVA-C, respectively. Detailed optimization results are provided in the supplementary document due to the page limit. Table II summarizes the experimental results to compare the performance of DVA-C against other ESTO methods on the 18 test problems. All experimental results are based on 30 independent runs. To test the statistical significance of the results, all other ESTO solvers are compared with the baseline (DVA-C) using a two-tailed Wilcoxon rank-sum test with $\alpha = 0.05$.

On the six SOPs, the DVA-C method shows the best convergence performance among the ESTO solvers. Fig. 7(a) shows the average convergence curves on three selected SOPs. The results on SOP 1 and SOP 6 verify the convergence acceleration of the proposed CTM. However, such improvement cannot be guaranteed on a wide variety of SOPs, as can be seen from the convergence curves on SOP 4. Fortunately, the injection-based transfer manner allows an optimizer to automatically eject the transferred solutions with low fitness in the selection phase. This mechanism enhances the robustness of the optimizer in receiving transferred solutions and thus reduces the risk of encountering the negative transfer. As can be seen

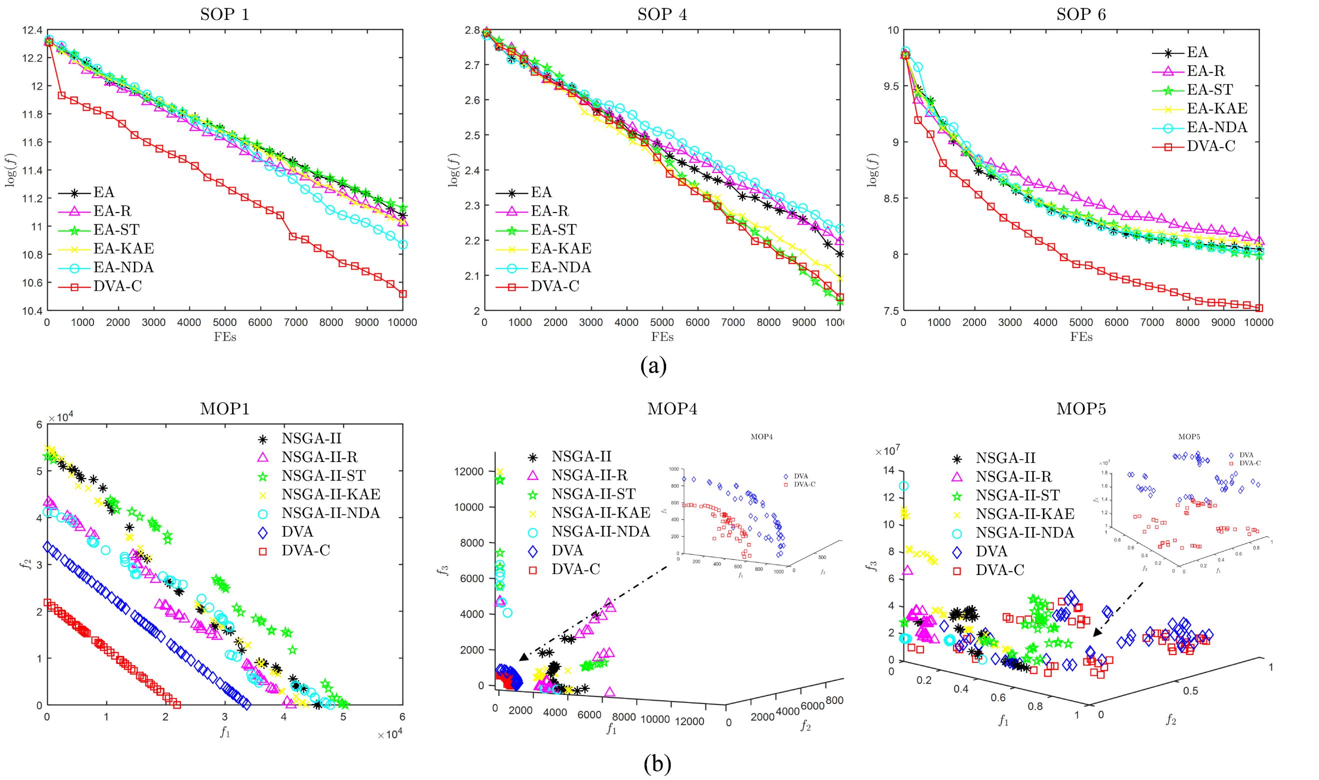


Fig. 7. Convergence curves and nondominated solutions in the objective space of SOLVER, SOLVER-R, SOLVER-ST, SOLVER-KAE, SOLVER-NDA, DVA, and DVA-C on three selected SOPs and three selected MOPs: (a) convergence curves on SOPs; (b) nondominated solutions on MaOPs.

TABLE II
PERFORMANCE COMPARISON OF DVA-C AGAINST THE REST OF THE SOLVERS ON THE 18 TEST PROBLEMS. DVA-C'S NUMBER OF WINS, TIES, AND LOSSES AGAINST OTHER METHODS IS REPORTED

	SOL w/t/l	SOL-R w/t/l	SOL-ST w/t/l	SOL-KAE w/t/l	SOL-NDA w/t/l	DVA w/t/l
SOPs	5/1/0	5/1/0	5/1/0	5/1/0	5/1/0	5/1/0
MOPs	5/1/0	5/1/0	5/1/0	5/1/0	5/1/0	4/2/0
MaOPs	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0	6/0/0
Total	16/2/0	16/2/0	16/2/0	16/2/0	16/2/0	15/3/0

from the averaged convergence curves, the EA-R shows comparable performance with the baseline solver, even though the transferred knowledge is randomly generated.

The comparison results based on the GD metric listed in Table II witness the convergence acceleration of the proposed CTM on MOPs and MaOPs.

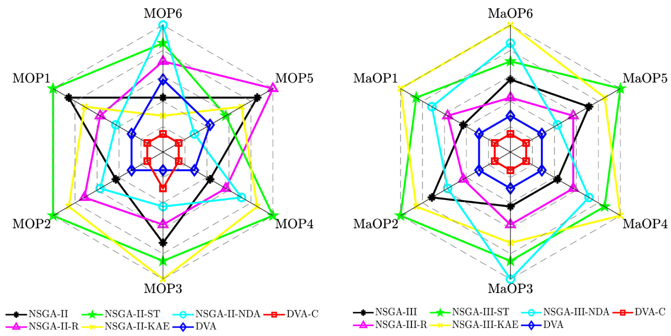
The DVA-C shows superior performance over a wide range of problems as compared to all the other ESTO solvers. Particularly, the final nondominated solutions in the objective space obtained by all the algorithms on three MOPs are graphically compared in Fig. 7(b). We can see that the DVA and DVA-C algorithms significantly outperform other solvers in terms of both convergence and diversity. The nondominated solutions provided by the DVA and DVA-C methods are uniformly distributed in the objective space. Further, the DVA-C is able to achieve better convergence state as compared to the DVA algorithm with no knowledge transfer. The solutions from the DVA-C completely dominate the solutions provided by the DVA. In summary, the proposed CTM works well on accelerating the convergence of SOPs, MOPs, and MaOPs based on the past experience drawn from heterogeneous instances.

D. Effectiveness of the Proposed DTM

When evaluating the performance of an MOP or MaOP solver, the priority should be given to convergence rather than diversity. For example, the superior diversity performance of solver A is trivial when a single solution obtained by solver B dominates all the well-distributed solutions from A. In such a scenario, solver B is significantly better than solver A despite its poor diversity. Thus, the diversity performance of different algorithms should be compared based on the same convergence state. Herein, the efficacy of the proposed DTM is verified within the DVA-ESTO framework.

Before we move on, the superiorities of the DVA and DVA-C algorithms over other ESTOs are confirmed first using the IGD metric. Due to the page limit, we compare the solver ranks measured by the IGD metric, which are shown as radar charts in Fig. 8. We can see that the DVA-based solvers show significantly better performance among the ESTO solvers under the IGD metric. Notably, benefiting from the control variable analysis, the baseline DVA solver with no knowledge transfer shows better performance than all the non-DVA solvers on most of the problems.

To verify the efficacy of the DTM, two comparison pairs are considered: 1) DVA versus DVA-D and 2) DVA-C versus DVA-ESTO. To eliminate the impact of convergence



(a) (b)

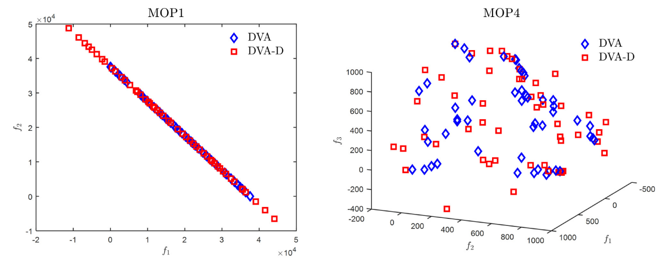
Fig. 8. Radar charts of the performance ranks measured by IGD of the seven solvers on six MOPs and six MaOPs: (a) IGD ranks on MOPs; (b) IGD ranks on MaOPs.

TABLE III
WILCOXON SIGNED-RANK TEST RESULTS FOR INVESTIGATING THE EFFICACY OF THE DTM (SIGNIFICANCE LEVEL $\alpha = 0.05$)

Problems	DVA vs DVA-D		DVA-C vs DVA-ESTO	
	p	Reject	p	Reject
MOP 1	1.92e-6	✓	1.73e-6	✓
MOP 2	1.73e-6	✓	1.73e-6	✓
MOP 3	2.80e-3	✓	9.71e-5	✓
MOP 4	1.73e-6	✓	1.73e-6	✓
MOP 5	1.48e-4	✓	4.86e-5	✓
MOP 6	1.92e-1	×	3.00e-3	✓
MaOP 1	7.16e-4	✓	9.32e-6	✓
MaOP 2	1.99e-1	×	7.04e-1	×
MaOP 3	9.80e-3	✓	1.50e-3	✓
MaOP 4	1.57e-2	✓	8.97e-2	×
MaOP 5	9.63e-4	✓	6.56e-2	×
MaOP 6	2.62e-1	×	1.32e-2	✓
Success rate	—	9/12	—	9/12

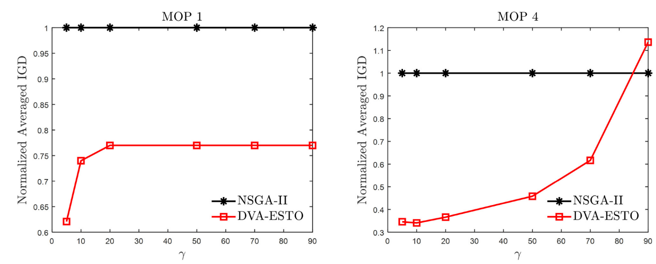
state, solvers in a comparison pair should share the same convergence-related solutions in a single run. Given the solutions sets S_1 and S_2 provided by a comparison pair (Q_1, Q_2), a nondominated solution set S_n can be obtained by selecting the nondominated solutions from $S_1 \cup S_2$. Then, the diversity of $\{Q_i\}_{i=1,2}$ can be quantified as the PD value based on S_i and S_n.

Table III summarizes the sign test results of comparison pairs to test for the efficacy of the proposed DTM. The null hypothesis H_0 is that the difference between DVA-based solvers before and after using the DTM has zero median. In other words, the null hypothesis assumes that the DTM has no positive effect on diversity performance. As can be seen from Table III, the null hypothesis is rejected on most of the problems for the DVA and DVA-C solvers. This indicates that the proposed diversity transfer method does have a significant impact on the diversity performance. To demonstrate how this module improves the diversity performance, the nondominated solutions in the objective space obtained by the DVA and DVA-D solvers on two selected MOPs are compared in Fig. 9. We can see that the solutions provided by the DVA-D algorithm possess better diversity as compared to the baseline solver DVA with no diversity transfer. The DVA solver tends to provide many crowded solution clusters with poor



(a) (b)

Fig. 9. Nondominated solutions in the objective space of DVA and DVA-D on two selected MOPs: (a) MOP1 and (b) MOP 4.



(a) (b)

Fig. 10. Normalized average IGDs obtained by DVA-ESTO against NSGA-II on two selected MOPs across 30 independent runs with different configurations of γ: (a) MOP 1 and (b) MOP 4.

diversity, while the DVA-D is able to obtain a solution set with many well-distributed solutions. In summary, with the proposed DTM, the past diversity-related searching experience can be used to better refine the distribution of target DVs.

E. Parameter Analysis

In the proposed method, γ is an important parameter that is used to allocate the computational resources for solving CVs and DVs.

In a black-box scenario, it is a challenging task to configure an optimal γ in advance. To this end, we further investigate how the configuration of γ affects the performance of the proposed DVA-ESTO.

Fig. 10 shows the normalized average IGDs obtained by DVA-ESTO against NSGA-II on MOP 1 and MOP 4 across 30 independent runs with different configurations of γ. We can see that the parameter does have a great impact on the performance of the proposed method. In particular, an extremely high value of γ may even let the proposed method perform worse than the base-line solver, as shown in Fig. 10(b). This phenomenon is caused by the imbalanced allocation of computational resources for CVs and DVs. For problems MOP 1 and MOP 4, their percentages of DVs are 0.03 and 0.06. In this case, more computational resources should be allocated to CVs rather than DVs. Thus, a low value of γ is suggested, which is consistent with the results shown in Fig. 10. In practice, γ can be simply set as the percentage of CVs. Besides, the allocation task can also be completed by adopting a utility threshold [33], [55] used to indicate the termination of the convergence optimization. In this way, a prespecified γ is not required. It is noteworthy

TABLE IV
OBJECTIVE CONFIGURATIONS OF THE SIX PRODUCTION PLANNING PROBLEMS AND THEIR DECISION VARIABLE ANALYSIS RESULTS

Problems	SOP 1	SOP 2	SOP 3	MOP 1	MOP 2	MaOP 1
Objs	$\{f_1\}$	$\{f_4\}$	$\{f_5\}$	$\{f_1, f_3\}$	$\{f_4, f_5\}$	$\{f_{1\sim5}\}$
CVs	$\{x_{1\sim6}\}$	$\{x_{1\sim6}\}$	$\{x_{1\sim6}\}$	$\{x_1, x_{3\sim6}\}$	$\{x_{1\sim4}\}$	$\emptyset$
DVs	$\emptyset$	$\emptyset$	$\emptyset$	$\{x_2\}$	$\{x_{5\sim6}\}$	$\{x_{1\sim6}\}$

that the configuration of allocation strategy does not affect the implementation of the proposed DVA-ESTO method, as their optimization histories can be analyzed in the same way.

The 18 optimization problems used in Sections V-B–V-D possess clear variable control properties (i.e., either purely convergence related or diversity related), which may not be able to represent the characteristics of many complex real-world problems. Besides, most of the variables in the 12 MOPs and MaOPs are pure convergence-related variables. With this in mind, we form an extended MOP test suite containing problems with different proportions of DVs to verify the efficacy of the proposed DVA-ESTO further.[3]

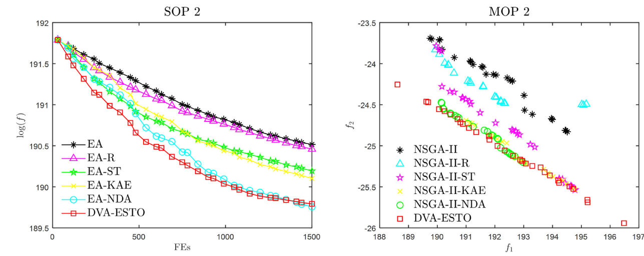
F. A Real-World Case Study

The production planning optimization is one of the most important problems in the mineral processing, which can greatly improve the nonrenewable raw mineral resource utilization. The goal of production planning optimization is to optimize production resources so as to maximize (or minimize) some desired production targets [56]. Generally, a production planning optimization problem can be modeled with an arbitrary number of objectives depending on specific problem settings, ranging from SOPs to MaOPs. Hopefully, the past planning experience stored can be used to assist the optimization of incoming tasks.

According to [56], in this work, we consider five production targets, namely, concentrate output, concentrate grade, concentration ratio, metal recovery, and cost indicators, which are denoted by f_1 to f_5, respectively. Three SOPs, two MOPs, and one MaOP are conducted to form the practical test problems. Table IV provides the detailed objective configurations of these problems and their associated variable analysis results. For solving a specific problem, the source database that contains the past searching experience will exclude this problem.

Due to the page limit, the objective or IGD mean and standard deviation obtained by the solvers in comparison over 30 independent runs are provided in the supplementary document. By leveraging the convergence-related and diversity-related searching experience separately, the proposed DVA-ESTO shows competitive optimization performance among the algorithms in comparison. To illustrate, the convergence curves on SOP 2 and the solution sets obtained on MOP 2 are compared in Fig 11. We can see that the DVA-ESTO and the solver based on NDA show superior convergence acceleration on SOP 2, while the proposed DVA-ESTO obtains a well-distributed solution set with competitive convergence performance on MOP 2.

[3]See Section S-II of the supplementary document due to the page limit.



(a) (b)

Fig. 11. Comparison of convergence curves on SOP 2 and solution sets obtained on MOP 2: (a) SOP 2 and (b) MOP 2.

However, on the MaOP problem whose every variable is recognized as diversity related, the proposed method performs poorly as compared to the other state-of-the-art ESTO solvers. This observation is consistent with the results in the extended test suite, which again suggests that the study of knowledge transfer among problems with a large portion of DVs (or mixed variables) is worth studying in the future.

VI. CONCLUSION

This article has attempted to address objective-heterogeneous ESTO problems, in which the optimization tasks vary among SOPs, MOPs, and MaOPs. To this end, this article has proposed a DVA-ESTO approach. It first divides decision variables of MOPs and MaOPs into CVs and DVs, and treats variables of SOPs as CVs from the perspective of variable control property. A CTM is then proposed to make the optimizer approach faster toward the optimal solution(s) using the past convergence experience. In the CTM, the fitness rank is considered when learning the source-target mapping. Meanwhile, for MOPs or MaOPs, a DTM is used to better refine the distribution of the solution set. Comprehensive experiments on a set of heterogeneous test problems have demonstrated the superiority of DVA-ESTO for boosting the optimizer performance via knowledge transfer.

To the best of our knowledge, this article is the first attempt toward an objective-heterogeneous ESTO. More research efforts should be devoted to this topic because the proposed DVA-ESTO is only applicable for MOPs or MaOPs with a certain portion of CVs, as observed from the results on the extended test problems. Successful sequential knowledge transfer across general MOPs or MaOPs with complex variable control properties is still an open issue. A promising solution for achieving this is to handle general optimization problems from a new perspective without the need for analyzing control property. In addition, some more complex test problems with strong heterogeneities with respect to both decision and objective spaces are required for analysis and comparisons of various ESTO methods, since such features are widespread in many real-world problems.

REFERENCES

[1] T. Back, U. Hammel, and H.-P. Schwefel, "Evolutionary computation: Comments on the history and current state," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 3–17, Apr. 1997.

[2] J. H. Holland, "Genetic algorithms and the optimal allocation of trials," *SIAM J. Comput.*, vol. 2, no. 2, pp. 88–105, 1973.
[3] Y. Ong, P. B. Nair, and K. Y. Lum, "Max–min surrogate-assisted evolutionary algorithm for robust design," *IEEE Trans. Evol. Comput.*, vol. 10, no. 4, pp. 392–404, Aug. 2006.
[4] Y. Cui, Y. Han, Z. Geng, Q. Zhu, and J. Fan, "Production optimization and energy saving of complex chemical processes using novel competing evolutionary membrane algorithm: Emphasis on ethylene cracking," *Energy Convers. Manage.*, vol. 196, pp. 311–319, Sep. 2019.
[5] X. Xue *et al.*, "A topology-based single-pool decomposition framework for large-scale global optimization," *Appl. Soft Comput.*, vol. 92, no. 106295, Jul. 2020.
[6] Y. Zhou, Y. Jin, and J. Ding, "Surrogate-assisted evolutionary search of spiking neural architectures in liquid state machines," *Neurocomputing*, vol. 406, pp. 12–23, Sep. 2020.
[7] K. C. Tan, L. H. Lee, Q. L. Zhu, and K. Ou, "Heuristic methods for vehicle routing problem with time windows," *Artif. Intell. Eng.*, vol. 15, no. 3, pp. 281–295, 2001.
[8] M. W. Hauschild, M. Pelikan, K. Sastry, and D. E. Goldberg, "Using previous models to bias structural learning in the hierarchical BOA," *Evol. Comput.*, vol. 20, no. 1, pp. 135–160, 2012.
[9] E. Cantú-Paz, *Efficient and Accurate Parallel Genetic Algorithms*. Boston, MA, USA: Kluwer Acad. Publ., 2000.
[10] M. G. Arenas *et al.*, "A framework for distributed evolutionary algorithms," in *Proc. Int. Conf. Parallel Problem Solving Nat.*, 2002, pp. 665–675.
[11] R. E. Smith, B. A. Dike, and S. A. Stegmann, "Fitness inheritance in genetic algorithms," in *Proc. ACM Symp. Appl. Comput.*, 1995, pp. 345–350.
[12] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *AIAA J.*, vol. 41, no. 4, pp. 687–696, 2003.
[13] A. Isaacs, T. Ray, and W. Smith, "A hybrid evolutionary algorithm with simplex local search," in *Proc. IEEE Congr. Evol. Comput.*, Singapore, 2007, pp. 1701–1708.
[14] J. Schwarz and J. Ocenasek, "A problem knowledge-based evolutionary algorithm KBOA for hypergraph bisectioning," in *Proc. 4th Joint Conf. Knowl. Based Softw. Eng.*, 2000, pp. 51–58.
[15] S. J. Louis and J. McDonnell, "Learning with case-injected genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 8, no. 4, pp. 316–328, Aug. 2004.
[16] A. Gupta, Y.-S. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Trans. Evol. Comput.*, vol. 20, no. 3, pp. 343–357, Jun. 2016.
[17] L. Feng, Y.-S. Ong, S. Jiang, and A. Gupta, "Autoencoding evolutionary search with learning across heterogeneous problems," *IEEE Trans. Evol. Comput.*, vol. 21, no. 5, pp. 760–772, Oct. 2017.
[18] B. Da, A. Gupta, and Y.-S. Ong, "Curbing negative influences online for seamless transfer evolutionary optimization," *IEEE Trans. Cybern.*, vol. 49, no. 12, pp. 4365–4378, Dec. 2019.
[19] Z. Tang, M. Gong, Y. Wu, W. Liu, and Y. Xie, "Regularized evolutionary multitask optimization: Learning to intertask transfer in aligned subspace," *IEEE Trans. Evol. Comput.*, vol. 25, no. 2, pp. 262–276, Apr. 2021.
[20] M. Jiang, Z. Wang, S. Guo, X. Gao, and K. C. Tan, "Individual-based transfer learning for dynamic multiobjective optimization," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4968–4981, Oct. 2021, doi: 10.1109/TCYB.2020.3017049.
[21] F. Zhang, Y. Mei, S. Nguyen, M. Zhang, and K. C. Tan, "Surrogate-assisted evolutionary multitask genetic programming for dynamic flexible job shop scheduling," *IEEE Trans. Evol. Comput.*, vol. 25, no. 4, pp. 651–665, Aug. 2021, doi: 10.1109/TEVC.2021.3065707.
[22] K. C. Tan, L. Feng, and M. Jiang, "Evolutionary transfer optimization—A new frontier in evolutionary computation research," *IEEE Comput. Intell. Mag.*, vol. 16, no. 1, pp. 22–33, Feb. 2021.
[23] P. Cunningham and B. Smyth, "Case-based reasoning in scheduling: Reusing solution components," *Int. J. Prod. Res.*, vol. 35, no. 11, pp. 2947–2962, 1997.
[24] S. J. Louis and C. Miles, "Playing to learn: Case-injected genetic algorithms for learning to play computer games," *IEEE Trans. Evol. Comput.*, vol. 9, no. 6, pp. 669–681, Dec. 2005.
[25] A. Gupta, Y.-S. Ong, and L. Feng, "Insights on transfer optimization: Because experience is the best teacher," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 51–64, Feb. 2018.
[26] L. Feng, Y.-S. Ong, A.-H. Tan, and I. W. Tsang, "Memes as building blocks: A case study on evolutionary optimization + transfer learning for routing problems," *Memetic Comput.*, vol. 7, no. 3, pp. 159–180, 2015.
[27] L. Feng, Y.-S. Ong, M.-H. Lim, and I. W. Tsang, "Memetic search with interdomain learning: A realization between CVRP and carp," *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 644–658, Oct. 2015.
[28] M. W. Hauschild and M. Pelikan, "Intelligent bias of network structures in the hierarchical BOA," in *Proc. 11th Annu. Conf. Genet. Evol. Comput.*, 2009, pp. 413–420.
[29] R. Santana, A. Mendiburu, and J. A. Lozano, "Structural transfer using EDAs: An application to multi-marker tagging SNP selection," in *Proc. IEEE Congr. Evol. Comput.*, Brisbane, QLD, Australia, 2012, pp. 1–8.
[30] S. Friess, P. Tiňo, S. Menzel, B. Sendhoff, and X. Yao, "Improving sampling in evolution strategies through mixture-based distributions built from past problem instances," in *Proc. Int. Conf. Parallel Problem Solving Nat.*, 2020, pp. 583–596.
[31] J. Zhang, W. Zhou, X. Chen, W. Yao, and L. Cao, "Multisource selective transfer framework in multiobjective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 24, no. 3, pp. 424–438, Jun. 2020.
[32] H. Wang, L. Jiao, R. Shang, S. He, and F. Liu, "A memetic optimization strategy based on dimension reduction in decision space," *Evol. Comput.*, vol. 23, no. 1, pp. 69–100, 2015.
[33] X. Ma *et al.*, "A multiobjective evolutionary algorithm based on decision variable analyses for multiobjective optimization problems with large-scale variables," *IEEE Trans. Evol. Comput.*, vol. 20, no. 2, pp. 275–298, Apr. 2016.
[34] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Trans. Evol. Comput.*, vol. 10, no. 5, pp. 477–506, Oct. 2006.
[35] X. Zhang, Y. Tian, R. Cheng, and Y. Jin, "A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 22, no. 1, pp. 97–112, Feb. 2018.
[36] C. Yang, J. Ding, K. C. Tan, and Y. Jin, "Two-stage assortative mating for multi-objective multifactorial evolutionary optimization," in *Proc. IEEE 56th Annu. Conf. Decis. Control (CDC)*, Melbourne, VIC, Australia, 2017, pp. 76–81.
[37] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
[38] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[39] L. Zhou, L. Feng, A. Gupta, and Y.-S. Ong, "Learnable evolutionary search across heterogeneous problems via kernelized autoencoding," *IEEE Trans. Evol. Comput.*, vol. 25, no. 3, pp. 567–581, Jun. 2021, doi: 10.1109/TEVC.2021.3056514.
[40] R. Lim, A. Gupta, Y.-S. Ong, L. Feng, and A. N. Zhang, "Non-linear domain adaptation in transfer evolutionary optimization," *Cogn. Comput.*, vol. 13, no. 2, pp. 290–307, 2021.
[41] R. Lim, L. Zhou, A. Gupta, Y.-S. Ong, and A. N. Zhang, "Solution representation learning in multi-objective transfer evolutionary optimization," *IEEE Access*, vol. 9, pp. 41844–41860, 2021.
[42] X. Xue *et al.*, "Affine transformation-enhanced multifactorial optimization for heterogeneous problems," *IEEE Trans. Cybern.*, early access, Dec. 15, 2020, doi: 10.1109/TCYB.2020.3036393.
[43] Z. Liang, H. Dong, C. Liu, W. Liang, and Z. Zhu, "Evolutionary multitasking for multiobjective optimization with subspace alignment and adaptive differential evolution," *IEEE Trans. Cybern.*, early access, Jun. 24, 2020, doi: 10.1109/TCYB.2020.2980888.
[44] R. Allmendinger and J. Knowles, "Heterogeneous objectives: state-of-the-art and future research," 2021, *arXiv:2103.15546*.
[45] J. Luo, A. Gupta, Y.-S. Ong, and Z. Wang, "Evolutionary optimization of expensive multiobjective problems with co-sub-Pareto front Gaussian process surrogates," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1708–1721, May 2019.
[46] L. Li and Z. Zhang, "Semi-supervised domain adaptation by covariance matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2724–2739, Nov. 2019.
[47] F. Hausdorff, "Summationsmethoden und momentfolgen. I," *Mathematische Zeitschrift*, vol. 9, nos. 1–2, pp. 74–109, 1921.
[48] Z. Chen, Y. Zhou, X. He, and J. Zhang, "Learning task relationships in evolutionary multitasking for multiobjective continuous optimization," *IEEE Trans. Cybern.*, early access, Nov. 18, 2022, doi: 10.1109/TCYB.2020.3029176.
[49] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemp. Math.*, vol. 26, no. 1, pp. 189–206, 1984.

[50] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[51] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[52] H. Wang, Y. Jin, and X. Yao, "Diversity assessment in many-objective optimization," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1510–1522, Jun. 2017.

[53] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, Apr. 2003.

[54] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 577–601, Aug. 2014.

[55] H. Chen, R. Cheng, W. Pedrycz, and Y. Jin, "Solving many-objective optimization problems via multistage evolutionary search," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3552–3564, Jun. 2021.

[56] G. Yu, T. Chai, and X. Luo, "Multiobjective production planning optimization using hybrid evolutionary algorithms for mineral processing," *IEEE Trans. Evol. Comput.*, vol. 15, no. 4, pp. 487–514, Aug. 2011.

Xiaoming Xue received the B.S. degree in petroleum engineering and the M.Sc. degree in oil and gas field development engineering from the China University of Petroleum (East China), Qingdao, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, City University of Hong Kong, Hong Kong.

His research interests include large-scale optimization, evolutionary transfer optimization, transfer learning, and their applications on engineering design optimization problems.

Cuie Yang (Member, IEEE) received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2019.

She is currently a Postdoctoral Research Fellow of the School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand. Her research interests include transfer learning, domain adaptation, concept drift learning, multitasking evolutionary optimization, and data-driven evolutionary optimization.

Yao Hu received the master's degree from the China University of Mining and Technology, Xuzhou, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Science, City University of Hong Kong, Hong Kong.

His current research interests include machine learning, transfer learning, and federated learning.

Kai Zhang (Member, IEEE) received the degree from the University of Tulsa, Tulsa, OK, USA, in May 2008, and the Ph.D. degree from the China University of Petroleum (East China), Qingdao, China, in 2008.

Since 2008, he has been teaching with the China University of Petroleum (East China), where he is a Professor and the Director of Talent Office. He teaches courses, including Fluid flow in porous media and reservoir engineering. He has already published more than 60 papers. As a Project Leader, he has been in charge of three projects supported by the Natural Science Foundation of China, one project supported by the National Natural Science Foundation of Shandong Province, one project supported by the National Ministry of Education, 20 projects supported by SINOPEC, CNOOC, and CNPC. His research interests focus on reservoir simulation, production optimization, history matching, and development of nonconventional reservoir.

Prof. Zhang is an SPE member and an InterPore Member.

Yiu-Ming Cheung (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2000.

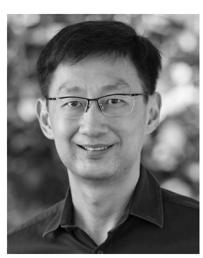
He is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, SAR, China. His research interests include machine learning, computer vision, pattern recognition, data mining, multiobjective optimization, and information hiding.

Prof. Cheung serves as an Associate Editor of IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2014 to 2020, *Pattern Recognition*, and *Neurocomputing*. He is a Fellow of IET, BCS, and RSA, as well as IETI Distinguished Fellow. For details, please refer to: http://www.comp.hkbu.edu.hk/ ymc.

Linqi Song (Member, IEEE) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in electrical engineering from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2017.

He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong and a Research Scientist with the City University of Hong Kong, Shenzhen Research Institute, Shenzhen, China. He was a Postdoctoral Scholar with the Department of Electrical and Computer Engineering, UCLA. His research interests include information theory, machine learning, and big data.

Dr. Song has received the Hong Kong RGC Early Career Scheme in 2019 and the Best Paper Award from IEEE MIPR 2020.



Kay Chen Tan (Fellow, IEEE) received the B.Eng. (First Class Hons.) and Ph.D. degrees from the University of Glasgow, Glasgow, U.K., in 1994 and 1997, respectively.

He is currently the Chair Professor (Computational Intelligence) with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. He has published over 300 refereed articles and seven books.

Prof. Tan is currently the Vice-President of IEEE Computational Intelligence Society, Kingston, RI, USA. He served as the Editor-in-Chief of the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE from 2010 to 2013 and the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION from 2015 to 2020. He currently serves as an Editorial Board Member for more than ten journals. He is an IEEE Distinguished Lecturer Program DLP Speaker and the Chief Co-Editor of Springer Book Series on Machine learning: Foundations, Methodologies, and Applications.