

Surrogate-Assisted Search With Competitive Knowledge Transfer for Expensive Optimization

Xiaoming Xue^{ID}, Member, IEEE, Yao Hu^{ID}, Member, IEEE, Liang Feng^{ID}, Senior Member, IEEE, Kai Zhang^{ID}, Member, IEEE, Linqi Song^{ID}, Senior Member, IEEE, and Kay Chen Tan^{ID}, Fellow, IEEE

Abstract—Expensive optimization problems (EOPs) have attracted increasing research attention over the decades due to their ubiquity in a variety of practical applications. Despite many sophisticated surrogate-assisted evolutionary algorithms (SAEAs) that have been developed for solving such problems, most of them lack the ability to transfer knowledge from previously-solved tasks and always start their search from scratch, making them troubled by the notorious cold-start issue. A few preliminary studies that integrate transfer learning into SAEAs still face some issues, such as defective similarity quantification that is prone to underestimate promising knowledge, surrogate-dependency that makes the transfer methods not coherent with the state-of-the-art in SAEAs, etc. In light of the above, a plug and play competitive knowledge transfer (CKT) method is proposed to boost various SAEAs in this article. Specifically, both the optimized solutions from the source tasks and the promising solutions acquired by the target surrogate are treated as task-solving knowledge, enabling them to compete with each other to elect the winner for expensive evaluation, thus boosting the search speed on the target task. Moreover, the lower bound of the convergence gain brought by the knowledge competition is mathematically analyzed, which is expected to strengthen the theoretical foundation of sequential transfer optimization. Experimental studies conducted on a series of benchmark problems and a practical application from the petroleum industry verify the efficacy of the proposed method. The source code of the CKT is available at <https://github.com/XmingHsueh/SAS-CKT>.

Received 7 September 2023; revised 9 January 2024, 15 March 2024, 30 May 2024, and 20 August 2024; accepted 23 September 2024. Date of publication 11 October 2024; date of current version 3 December 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFC3801700; in part by the National Natural Science Foundation of China under Grant U21A20512 and Grant 62371411; in part by the Research Grants Council of the Hong Kong, SAR, under Grant C5052-23G, Grant PolyU 11211521, Grant PolyU 15218622, Grant PolyU 15215623, Grant GRF 11217823, and Collaborative Research Fund Grant C1042-23GF; in part by InnoHK Initiative; and in part by the Laboratory for AI-Powered Financial Technologies, the Government of the HKSAR. This article was approved by Associate Editor C. Sun. (*Corresponding authors:* Liang Feng; 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).

Liang Feng is with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: liangf@cqu.edu.cn).

Kai Zhang is with the Civil Engineering School, Qingdao University of Technology, Qingdao 266520, China (e-mail: zhangkai@qut.edu.cn).

Kay Chen Tan is with the Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong, SAR, China (e-mail: kctan@polyu.edu.hk).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TEVC.2024.3478732>, provided by the authors.

Digital Object Identifier 10.1109/TEVC.2024.3478732

Index Terms—Evolutionary algorithms (EAs), expensive optimization problems (EOPs), knowledge competition, surrogate-assisted search (SAS), transfer optimization.

I. INTRODUCTION

EXPENSIVE optimization problems (EOPs) [1], [2], [3] refer to the problems whose objective functions or constraints involve expensive or even unaffordable evaluations, which widely exist in many real-world applications. Hyperparameter tuning of deep learning models [4], high-fidelity numerical simulation [5], and physical experiment-based optimization [6] are just a few representative examples. Despite the popularity of evolutionary algorithms (EAs) in solving various optimization problems due to their superior global search capabilities and domain information-agnostic implementations [7], they typically assume that evaluating the objectives or constraints of candidate solutions is not expensive and always require a large number of evaluations to obtain satisfactory solutions, making them computationally unaffordable for many EOPs.

Over the past decades, a variety of techniques have been developed to improve EAs on EOPs, which can be divided into three categories according to the cost reduction mechanism in [3]: 1) problem approximation and substitution [8]; 2) algorithm enhancement [9]; and 3) parallel and distributed computing. In particular, surrogate-assisted search (SAS) in the first category has been gaining great popularity due to its ease of implementation and good generality [10]. Polynomial response surface [11], Gaussian process regression [12], and radial basis function [13] are a few representative surrogates. Over the years, a growing number of surrogate-assisted EAs (SAEAs) have been proposed to combat different complexities of EOPs, including objective conflicts in multiobjective problems [14], high-dimensional variables in large-scale optimization [15], and robust optimization for problems with uncertainties [16], to name a few. However, despite numerous promising results reported, most SAEAs lack the capability to transfer knowledge from potentially similar tasks toward enhanced problem-solving efficiency. Oftentimes, they consider potentially related tasks in isolation and start the search of each individual task from scratch [17], which often require a considerable number of expensive evaluations to achieve high-quality solutions. This phenomenon is also known as the cold-start issue in [18].

To alleviate the cold-start issue, a growing number of studies propose to endow SAEAs with the ability to transfer knowledge from possibly related tasks for achieving better performance [19]. According to the availability of source tasks, knowledge transfer can be divided into three categories [20]: 1) sequential transfer [21], [22], [23]; 2) multitasking [24], [25], [26]; and 3) multiform optimization [27], [28]. In cases where no source tasks are available, one can generate alternate formulations of a target task of interest and employ them as helper tasks to better solve the target task [29], [30], which is known as multiform optimization. By contrast, if a certain number of unoptimized tasks that are potentially related to the target task are available, one can solve them with the target task in a multitasking manner to exploit intertask synergies for improved search performance [31]. Moreover, when a number of previously-solved tasks are available as the source tasks, one can extract transferable cues from them to guide the target optimization [32], [33], which is termed sequential transfer.

In this study, we focus on improving SAEAs with sequential transfer, due to its widespread application scenarios in the realm of engineering design, where many problems necessitate frequent reoptimization in view of changing environments or problem features, such as aerodynamic shape design with new conditions and constraints [34], well placement optimization of newfound reservoirs [35], geological parameter estimation of a series of core plugs from nearby blocks [36], and heat extraction optimization of geothermal systems [37]. As time goes by, an increasing number of optimization tasks accumulate in a database, providing an opportunity for the target task at hand to achieve better optimization performance via knowledge transfer. In the literature, some attempts have been made to improve SAEAs through sequential knowledge transfer. For instance, in [32], sequential transfer is conducted by aggregating the source surrogate models with the target one, where the weights for capturing source-target similarities are estimated by minimizing the squared error of out-of-sample predictions. However, the aggregation-based similarity is based on the objective values of solutions instead of their ranks, which tends to underestimate the similarity between two tasks with quantitatively different objective responses but similar ranks standing for high solution transferability [38]. Besides, the lack of task adaptation in the aggregation model makes it less effective in dealing with heterogeneous tasks. To address this issue, a generalized transfer Bayesian optimization algorithm that employs neural networks to overcome the task heterogeneity is developed in [33]. However, this algorithm shows poor portability due to its dependency on the Bayesian models and specific applicability to single-source problems.

With the above in mind, this article proposes a plug and play knowledge transfer module for enhancing the problem-solving capability of SAEAs. A novel mechanism named competitive knowledge transfer (CKT) is proposed, which assesses the promising solutions obtained by SAS on the target task and the optimized solutions from the source tasks from a consistent view. By treating both of these two types of solutions as task-solving knowledge for competition, it selects the winner to undergo the real function evaluation (FE). Overall, the

proposed SAS with competitive knowledge transfer (SAS-CKT) exhibits three main merits: 1) portability that makes it convenient to be used for improving different SAEAs and thus coherent with the state-of-the-art; 2) reliability of similarity quantification that enables knowledge transfer across two tasks with high solution transferability regardless of their different orders of magnitude; and 3) adaptivity that allows it to deal with the task heterogeneity. Experimental studies conducted on a series of benchmark problems and a practical application demonstrate the efficacy of SAS-CKT and its superiority against a few state-of-the-art algorithms. The main contributions of this article are summarized as follows.

- 1) A plug and play CKT method is proposed to enable prompt performance improvement of different SAEAs. Specifically, by treating both the source and target solutions as task-solving knowledge and assessing them from a consistent view, the winner is identified for expensive evaluation.
- 2) We conduct theoretical analyses to prove that the lower bound of the convergence gain brought by the proposed method is bounded by zero regardless of the source-target similarity. Moreover, important conditions that could lead to positive convergence gain are analyzed in detail.
- 3) The efficacy of the proposed method is empirically validated on a series of benchmark problems, as well as a practical case study from the petroleum industry.

The remainder of this article is organized as follows. Section II presents the definitions of SAS and sequential transfer. After that, a few preliminary studies that improve SAS with sequential transfer are reviewed and analyzed, followed by the motivations behind this work. Then, Section III introduces the proposed SAS-CKT in detail and Section IV presents its theoretical analyses. With the experimental settings provided in Section V, we conduct our experiments and analyze the corresponding results in Section VI. Lastly, Section VII concludes this article.

II. PRELIMINARIES

In this section, we first briefly introduce the canonical SAS paradigm and the sequential transfer-enhanced optimization. After that, the limitations of a few preliminary studies that improve SAS with sequential transfer are analyzed. Lastly, the motivations behind this work are presented.

A. Surrogate-Assisted Search

SAS is a widely used optimization paradigm for computationally expensive problems, which consists of the following three key parts [10].

- 1) *Approximation*: Construct (or update) a surrogate model to approximate the objective function of interest.
- 2) *Acquisition*: Conduct the global or local search on the surrogate model with specific optimizers (e.g., EAs or nonlinear programming techniques) to acquire a promising solution.

TABLE I
LIST OF STATE-OF-THE-ART SAS ALGORITHMS

Algorithm	Year	Surrogate Model		Search Mechanism		Infill Criterion	
		Type	Model	Prescreening or Iteration	Optimizer		
IKAEA [39]	2021	Global	Gaussian process regression (GPR)	Prescreening	Differential evolution (DE)	Expected improvement (EI)	
TLRBF [40]	2022	Global	Radial basis function (RBF)	Prescreening	Random sampling		
		Subregion		Iteration	Adaptive DE (JADE)	Predicted objective value (POV)	
		Local					
GL-SADE [41]	2022	Global	RBF	Iteration	DE	POV	
		Local	GPR	Prescreening		Lower confidence bound (LCB)	
DDEA-MESS [42]	2022	Global	RBF	Prescreening	DE		
		Local		Iteration		POV	
		Region		Trust region search			
LSADE [43]	2023	Global	RBF	Prescreening	DE	POV	
		Region	Lipschitz surrogate				
		Local	RBF				
AutoSAEA [44]	2024	Local	GPR	Prescreening	DE	{LCB, EI}	
			RBF	{Prescreening, Iteration}		POV	
			Polynomial response surface (PRS)				
		K-nearest neighbor (KNN)	Prescreening			{L1-exploitation, L1-exploration}	

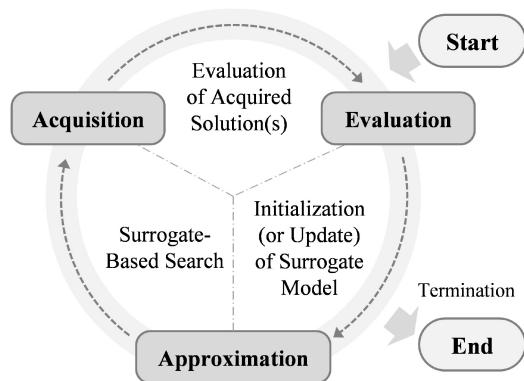


Fig. 1. High-level structure of SAS.

- 3) *Evaluation*: Evaluate the promising solution by calling the real FE and update the database with the newly evaluated solution accordingly.

Fig. 1 shows the high-level structure of SAS. It starts with the evaluation of a number of uniformly distributed solutions for initializing a surrogate model, which can be generated by a random sampling technique (e.g., the Latin hypercube sampling, LHS). Polynomial response surface [11], Gaussian process regression [12], and radial basis function [13] are a few commonly used surrogate models. Then, with a specific acquisition function, also known as infill criterion, the global or local search with particular optimizers is performed on the surrogate model instead of the real function to acquire a promising solution. Subsequently, the promising solution is evaluated by the real function and added into the database for updating the surrogate model. The above three phases are

repeatedly executed until the termination condition is met, e.g., the maximum number of expensive evaluations is reached.

Despite a wide variety of SAS algorithms that have been developed over the decades, the majority of them follow the high-level structure shown in Fig. 1. To demonstrate, we select a number of recently published SAS algorithms from different peer-reviewed journals and compare their components in Table I. It can be seen that the advances in SAS mainly come from three aspects: 1) appropriate surrogate models for approximating the problem of interest; 2) novel search mechanisms for optimizing the surrogates; and 3) effective infill criteria for acquiring promising solutions. Most recently, a sophisticated SAS framework with the auto-configuration of surrogate model and infill criterion is developed in [44], whose superiority against many other state-of-the-art SAEAs is validated on a number of complex benchmark problems and practical applications. For more detailed introductions of SAS and comprehensive reviews of various SAS algorithms, interested readers may refer to [2], [3], and [10].

B. Sequential Transfer

With a number of previously-optimized source tasks stored in a knowledge base \mathcal{M} , a target task of interest is supposed to achieve better task-solving efficiency by transferring the knowledge extracted from \mathcal{M} [20], as can be formulated by

$$Q(\mathcal{T}|\mathcal{M}) - Q(\mathcal{T}) > 0 \quad (1)$$

where \mathcal{T} is the target task, $Q(\mathcal{T})$ indicates the efficiency of an optimizer without transfer, and $Q(\mathcal{T}|\mathcal{M})$ denotes the efficiency of the optimizer with the knowledge transferred from \mathcal{M} .

The sequential transfer in (1) has been widely employed to improve the evolutionary search as a means of convergence speedup [45], [46], [47]. Likewise, it provides a promising solution for alleviating the cold-start issue in SAS, which has fostered several relevant studies presented in what follows.

C. Surrogate-Assisted Search Boosted by Sequential Transfer

When solving closed-box EOPs with SAS, the available information of the previously-optimized source tasks is simply their surrogate models with the associated evaluated solutions. Given a target task to be optimized, its task-solving efficiency is expected to be improved with the knowledge extracted from the source data, as can be formulated by

$$\min_{x \in \Omega} \left[f(x) | \{ \hat{f}_i^s(x|\mathcal{D}_i^s), 1 \leq i \leq k \} \right] \quad (2)$$

where x denotes the decision vector, Ω is the decision space, $f(x)$ is the objective function of the target task, $\hat{f}_i^s(x|\mathcal{D}_i^s)$ denotes the surrogate model of the i th source task built upon the source data $\mathcal{D}_i^s = [X_i, y_i]$, X_i and y_i denote the evaluated solutions and the associated objective values of the i th source task, and k is the number of source tasks.

Bayesian optimization is a representative SAS framework that uses Bayesian inferences (e.g., Gaussian process regression and Bayesian neural networks) as its surrogate models [48], which has been gaining increasing research efforts on endowing it with the ability to transfer knowledge from potentially related tasks for better optimization performance. Prior incorporation [49], kernelization of cross-task samples [33], and ensemble inference [50] are three knowledge transfer techniques commonly used for improving Bayesian optimization. However, all these methods are undeniably bonded to the associated Bayesian surrogate models, making them suffer from the downsides of those models, e.g., the poor scalability and the cubic complexity to the data size of Gaussian process [51]. Besides, the model-dependency of these knowledge transfer methods makes them disconnected from the advances made in general SAS methods.

In [32], a general ensemble inference model named multi-problem surrogates (MPSs) is proposed, in which the weights that capture source-target similarities are estimated by minimizing the squared error of out-of-sample predictions. It is noted that this ensemble inference method is fairly general although it was instantiated by Gaussian process only in [32], enabling it to be used in conjunction with other SAS algorithms with ease. However, for two optimization tasks with quantitatively different objective values but similar rank responses standing for high solution transferability [38], the weight estimation approach in [32] is prone to underestimate their similarity and miss out on the potentially valuable knowledge accordingly. Besides, the lack of task adaptation makes the MPSs less effective in solving problems with strong heterogeneity. To address this issue, Min et al. [33] proposed a generalized transfer Bayesian optimization algorithm, in which neural networks are employed to bridge the gap between the source and target tasks. A cross-task covariance term is used for utilizing the evaluated samples from a source task to achieve better prediction performance on the target task.

However, it is only applicable to problems with a single source task and still suffers from the issue of model-dependency.

D. Motivations Behind This Work

In light of the above, we aim to develop a general sequential transfer technique for improving SAS, which is supposed to have the following three favorable merits.

- 1) *Portability*: The knowledge transfer method should be plug and play (i.e., surrogate model-independent), enabling it to be coherent with the state-of-the-art in SAS.
- 2) *Reliability*: It should be equipped with rank-based similarity instead of the objective value-based one, as the latter is prone to underestimate the promising knowledge for problems containing tasks with heterogeneous orders of magnitude (HOM) but high ordinal correlations.
- 3) *Adaptivity*: It should be equipped with task adaptation, enabling it to solve problems with heterogeneous tasks.

Moreover, we show particular interest in the lower bound of the convergence gain brought by the proposed method and the important conditions leading to positive convergence gain, which will be theoretically derived in Section IV.

III. PROPOSED METHODS

Fig. 2 shows the flow chart of the proposed SAS-CKT, which aims to improve the optimization performance of SAS with a plug and play module named CKT. Unlike the traditional SAS that acquires promising solutions using the target surrogate model only, SAS-CKT has a knowledge competition module that treats both the source and target solutions as task-solving knowledge and compares them from a consistent view.

As shown in Fig. 2, the CKT module consists of three key parts: 1) *similarity quantification*; 2) *task adaptation*; and 3) *knowledge competition*. Specifically, similarity quantification is responsible for measuring the similarities between the source and target tasks while task adaptation is for improving the transferability of knowledge in case of task heterogeneity. Lastly, knowledge competition makes the promising solutions acquired by the target surrogate compete with the optimized solutions of the source tasks, allowing one to identify the winner for true evaluation. In what follows, the three key parts are introduced in detail.

A. Similarity Quantification

Suppose the database of a target EOP is denoted by $\mathcal{D} = [X, y]$ and a knowledge base with k previously-optimized EOPs is available, we denote the source surrogate models as $\hat{f}_i^s(x|X_i, y_i)$, $1 \leq i \leq k$. Without loss of generality, we employ the Spearman's rank correlation to measure the similarity between the source and target tasks due to its efficient utilization of the solution data and widespread application in transfer optimization, whose superiority against other metrics will be empirically validated in Section VI-B. Since the source surrogates normally have wider coverage over the search space¹ as compared to the target surrogate built upon

¹For source and target tasks with distinct decision spaces, one can transform them into a common search space $\Omega_c = [0, 1]^d$ for knowledge transfer.

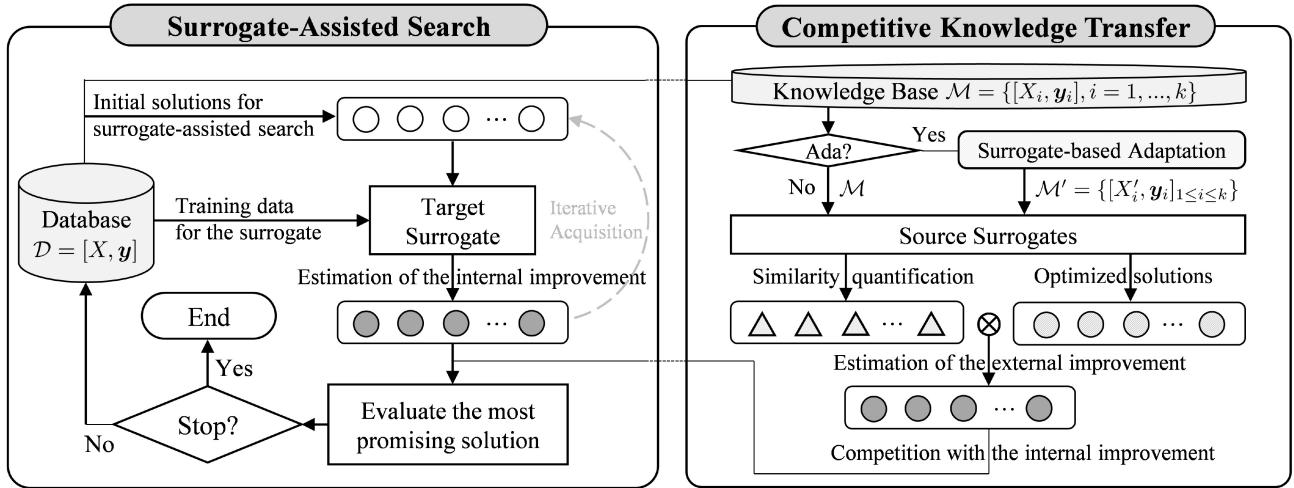


Fig. 2. Flow chart of the proposed SAS-CKT.

limited data, their prediction performance is generally better than the target one. Therefore, we propose surrogate-based Spearman's rank correlation (SSRC) to measure the source-target similarities based on the true responses of the evaluated solutions in \mathcal{D} and their predicted responses on the source surrogates, as given by

$$s_i = \frac{\text{cov}(\mathcal{R}[\hat{y}_i], \mathcal{R}[y])}{\text{std}(\mathcal{R}[\hat{y}_i])\text{std}(\mathcal{R}[y])} \quad (3)$$

where $s_i \in [-1, 1]$ is the similarity between the i th source task and the target task, $\mathcal{R}[\mathbf{a}]$ is the rank vector of \mathbf{a} , \hat{y}_i represents the predicted responses of X on the i th source surrogate.

B. Task Adaptation

For problems with heterogeneous tasks, task adaptation provides a solution for bridging the task gap and thus improving the transferability of knowledge. Without loss of generality, we propose to adapt the source solutions using a transformation $\phi(\mathbf{x}) : \Omega_s \rightarrow \Omega_t$, such that $\mathbf{x}' = \phi(\mathbf{x})$. Suppose ϕ is bijective and its inverse is $\phi^{-1} : \Omega_t \rightarrow \Omega_s$, the inversely adapted target solutions are denoted as $X^\circ = \phi^{-1}(X)$. It is noted that improving the transferability depends on what is used for representing (or indicating) the transferability [47]. Following this clue, we propose surrogate-based domain adaptation (SDA) to improve the transferability reflected by (3) with an optimal mapping, as formulated by:

$$\tilde{\phi}_i = \max_{\phi} s_i^a = \max_{\phi} \alpha_{\phi} \frac{\text{cov}(\mathcal{R}[\hat{y}_i^\circ], \mathcal{R}[y])}{\text{std}(\mathcal{R}[\hat{y}_i^\circ])\text{std}(\mathcal{R}[y])} \quad (4)$$

where $\tilde{\phi}_i$ denotes the optimal mapping for adapting the i th source task, s_i^a is the adaptation-based similarity, α_{ϕ} is a regularization term that prevents the mapping ϕ from overfitting, and \hat{y}_i° denotes the predicted responses of X° on the i th source surrogate. Then, the optimized solution $\tilde{x}_i \in X_i$ is adapted and prepared for knowledge transfer, i.e., $\tilde{x}'_i = \tilde{\phi}_i(\tilde{x}_i)$.

Without loss of generality, we consider the translation transformation in this work. Then, the mapping can be explicitly parameterized as $\phi_{\theta}(\mathbf{x}) = \mathbf{x} + \theta$, where θ is a translation vector.

The parameterized mapping ϕ_{θ} is bijective and its inverse is $\phi_{\theta}^{-1}(\mathbf{x}') = \mathbf{x}' - \theta$. Therefore, we can formulate the adaptation model in (4) into the following form:

$$\tilde{\theta}_i = \max_{\theta} \alpha_{\theta} \frac{\text{cov}(\mathcal{R}[\hat{f}_i^s(X - \Theta)], \mathcal{R}[y])}{\text{std}(\mathcal{R}[\hat{f}_i^s(X - \Theta)])\text{std}(\mathcal{R}[y])} \quad (5)$$

subject to $\theta \in [-1, 1]^d$

where $\tilde{\theta}_i$ is the estimated translation vector for the i th source task, and Θ is the broadcast matrix of θ whose shape is compatible with X . Without loss of generality, the regularization term α_{θ} is set to be $1 - \|\theta\|_{\infty}$. The optimized solution from the i th source task can be adapted as $\tilde{x}'_i = \tilde{x}_i + \tilde{\theta}_i$.

It should be noted that solving the problem in (5) requires some evaluations of the source surrogate models. For most EOPs, such surrogate evaluations are typically computationally affordable, making it worth spending a certain number of surrogate evaluations solving the adaptation problem. Moreover, it is worth noting that the task adaptation can be fulfilled in either an online or an offline fashion. In the online manner, the task adaptation is supposed to be executed once the database of the target task is updated. Alternatively, one can execute the task adaptation only once with the initial target data, which is computationally efficient but at the expense of less accurate adaptations. This tradeoff depends on the available computational resources allocated to the task adaptation. Nevertheless, in Section IV, we show that the convergence gain brought by the CKT module without the task adaptation is bounded by zero regardless of source-target similarities. In this sense, the task adaptation is more of an auxiliary component than an essential part of SAS-CKT, which is also empirically verified in Section VI-C.

C. Competitive Knowledge Transfer

From the standpoint of SAS on the target task, we define internal and external improvements as follows.

- 1) *Internal Improvement:* The estimated (or predicted) improvement brought by a promising solution acquired

by SAS on the target task under a particular infill criterion.

- 2) *External Improvement*: The estimated (or predicted) improvement made by the optimized solution from a specific source task on the target task.

For a single run of the SAS, the interval improvement can be estimated as follows:

$$\Delta_{\text{in}} = \min[\mathcal{I}(\mathcal{D})] - \mathcal{I}(x_p) \quad (6)$$

where Δ_{in} is the internal improvement, $\mathcal{I}(\cdot)$ represents the infill criterion (i.e., acquisition function) to be minimized on the target surrogate, and x_p is the most promising solution obtained by SAS with a particular search algorithm on the target surrogate. It is noted that the estimation in (6) is fairly general, which is compatible with many state-of-the-art SAS algorithms,² such as those listed in Table I.

For the external improvement, one may simply calculate it with (6) by replacing x_p with the optimized solution from a source task. However, this strategy is problematic as the source optimized solution is very likely to be mistakenly estimated by the target surrogate trained on limited data, which is especially the case at the early stage of SAS. In view of the fact that the source tasks had been optimized and their surrogates trained on a large number of evaluated solutions are relatively more reliable, we propose a convergence analogy-based estimation (CAE) method to estimate the external improvement. Unlike the replacement strategy that relies on a potentially inaccurate target surrogate, CAE calculates improvements on the source surrogates and then analogizes them to the external improvements on the target task. To enable the analogy, we postulate that the best objective values found by SAS on an optimization task are subject to the exponential decay due to its monotonicity and asymptotic convergence

$$\gamma_\tau = \gamma_o + \gamma_i e^{-\lambda\tau} \quad (7)$$

where γ_τ is the best objective value ever found till time τ , γ_i and γ_o represent the initial and optimal objective values, respectively, λ denotes the decay constant. For a specific task, its optimization trace can be modeled by (7) with particular parameters estimated by the ordinary least squares method.

For the i th source task, the improvement brought by its optimized solution against the target solutions is estimated as

$$\Delta_y^i = \min(\hat{y}_i) - \min(y_i) \quad (8)$$

where \hat{y}_i denotes the predicted responses of X on the i th source surrogate, y_i represents the objective values of all the evaluated solutions from the i th source task. It should be noted that the improvement in (8) is in the context of the i th source task and cannot be used as the improvement on the target task, as their orders of magnitude may vary a lot. To address this issue, CAE employs the exponential decay in (7) to analogize the improvement using the time interval corresponding to Δ_y^i , which is given by

²Detailed calculations of the internal improvement under a variety of infill criteria are provided in Section S-I of the supplementary document.

$$\begin{aligned} \Delta_\tau^i &= \tau[\min(y_i)] - \tau[\min(\hat{y}_i)] \\ &= \tau_{\max}^i - \ln\left(\frac{\gamma_i^i}{\min(\hat{y}_i) - \gamma_o^i}\right)/\lambda^i \\ &= \tau_{\max}^i - \tau_v^i \end{aligned} \quad (9)$$

where τ_{\max}^i denotes the time required by SAS to achieve the objective value of $\min(y_i)$ on the i th source task, which is approximately equal to the maximum number of evaluations, γ_i^i and γ_o^i are the estimated initial and optimal objective values of the i th source exponential decay, λ^i is the estimated decay constant, and τ_v^i is the estimated time required by SAS to achieve the objective value of $\min(\hat{y}_i)$ on the i th source task.

When analogizing the improvement on the target task with the improvement on the i th source task, their similarity should be taken into account. When s_i is 0, the source optimized solution (i.e., \tilde{x}_i) can be seen as a randomly generated solution for the target task. In this case, the expected time required by SAS to achieve the quality of \tilde{x}_i is simply 0. When s_i is 1, it can be concluded that the two tasks are very similar, enabling one to use the improvement on the source task to infer the external improvement brought by \tilde{x}_i . In summary, the confidence in the analogized time of \tilde{x}_i on the target task (i.e., $\tau + \Delta_\tau^i$) grows with the similarity s_i , where τ is the current time of the target search, which is simply the number of used FEs. Without loss of generality, CAE uses the linear correlation between s_i and $\tau + \Delta_\tau^i$ to estimate the external improvement brought by \tilde{x}_i , which is given by

$$\Delta_{\text{ex}}^i = s_i^+ [\min(y) - \gamma_{s_i^+(\tau+\Delta_\tau^i)}^t] \quad (10)$$

where Δ_{ex}^i denotes the external improvement brought by the optimized solution from the i th source task (i.e., \tilde{x}_i), s_i^+ is the rectified value of s_i , i.e., $s_i^+ = \max(0, s_i)$, which is used for suppressing the knowledge transfer when $s_i < 0$, $\gamma^t(\cdot)$ denotes the exponential decay model of the target task.

Finally, the knowledge competition kicks in by comparing the internal and external improvements, enabling the winner to undergo the real evaluation, which is given by

$$x_e = \begin{cases} \tilde{x}_{\max}, & \Delta_{\text{in}} < \Delta_{\text{ex}}^{\max} \\ x_p, & \Delta_{\text{in}} \geq \Delta_{\text{ex}}^{\max} \end{cases} \quad (11)$$

where x_e is the candidate solution to be evaluated, $\Delta_{\text{ex}}^{\max} = \max_{1 \leq i \leq k} \Delta_{\text{ex}}^i$ denotes the maximum external improvement, \tilde{x}_{\max} is the optimized solution of the source task with the maximum external improvement. When using the task adaptation, one only needs to replace the similarity in (10) by the adaptation-based similarity in (4) and adapt the source solutions, with the other procedures remained unaltered.

D. Implementation of SAS-CKT

The pseudo code of SAS-CKT is presented in Algorithm 1. The framed knowledge competition phase shown in lines 4–11 distinguishes SAS-CKT from the traditional SAS. The good portability of the proposed method enables it to boost the performance of various SAS engines. For an SAS-CKT algorithm equipped with a specific SAS engine, the knowledge competition can be uninterruptedly triggered by comparing the

Algorithm 1: SAS-CKT

Input : $f(\mathbf{x})$ (target task), $\mathcal{M} = \{[\mathbf{X}_i, \mathbf{y}_i], i = 1, \dots, k\}$
 (knowledge base), $\hat{f}_{1 \leq i \leq k}^s(\mathbf{x})$ (source surrogates)

Output : \mathbf{x}_b (the best solution)

- 1 $\mathcal{D} \leftarrow$ Initialize a database using a random sampling method;
- 2 **while** termination condition is not met **do**
- /* Internal Improvement */
- 3 $[\mathbf{x}_p, \Delta_{in}] \leftarrow$ SAS(\mathcal{D});
- 4 **if** CKT is allowed, i.e., $\text{mod}(|\mathcal{D}|, \delta) = 0$ **then**
- /* External Improvements */
- 5 $[\gamma^t, \gamma^{1 \leq i \leq k}] \leftarrow$ Fit the exponential decays;
- 6 **for** $i = 1 \rightarrow k$ **do**
- 7 $\Delta_{ex}^i \leftarrow$ The external improvement in Eq. (10);
- 8 /* Knowledge Competition */
- 9 **if** $\Delta_{in} < \max_{1 \leq i \leq k} \Delta_{ex}^i$ **then**
- 10 $\mathbf{x}_e \leftarrow \tilde{\mathbf{x}}_{\max}$;
- 11 **else**
- 12 $\mathbf{x}_e \leftarrow \mathbf{x}_p$;
- 13 /* Update of the Database */
- 14 $\mathcal{D} \leftarrow \mathcal{D} \cup [\mathbf{x}_e, f(\mathbf{x}_e)]$;
- 15 **return** the best solution $\mathbf{x}_b \in \mathcal{D}$;

internal and external improvements at each real evaluation. However, the landscape update of the target task is usually mild when only one newly evaluated solution is archived, as shown in line 14, making the external improvements bear insignificant changes. Therefore, a fixed internal δ in terms of real evaluation for triggering the knowledge competition can be used to reduce its frequent calculations, as shown in line 4. Lastly, it should be noted that the use of the task adaptation will not change the implementation of SAS-CKT. One only needs to replace the similarity by the adaptation-based similarity when estimating the external improvements in line 7 and adapt the source optimized solution with the learned mapping for knowledge transfer in line 9.

E. Computational Complexity

According to Algorithm 1, we find that the computational complexity of the proposed CKT method mainly comes from the estimation of the external improvements. This process involves calling the k source surrogate models for similarity quantification with (3), whose overall complexity is $O(kn \cdot c_p)$, where n represents the number of evaluated target solutions, c_p is the prediction cost of source surrogate. With respect to both k and n , CKT exhibits linear computational complexity. As for the prediction cost c_p in SAS-CKT, it depends on the source surrogate used. For most surrogate models, their prediction costs are far lower than the training costs. For instance, the prediction complexities of commonly used surrogates, such as RBF and GPR used in our experiments, are simply $O(dN)$, where d denotes the problem dimension, N is the number of training samples of the source surrogates.

IV. THEORETICAL ANALYSES

In this section, we conduct theoretical analyses of SAS-CKT. First, two prerequisites and three supporting lemmas

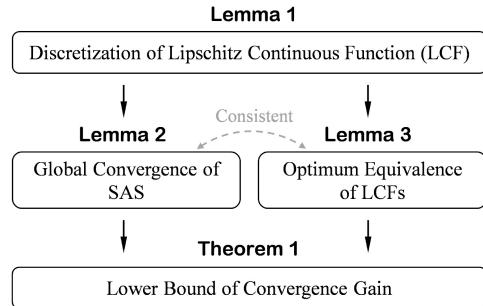


Fig. 3. Relations between the three lemmas and Theorem 1.

are briefly discussed. Then, the lower bound of the expected convergence gain brought by the knowledge competition is mathematically analyzed. Lastly, a brief discussion about the implications of the derived theoretical results is presented.

A. Supporting Lemmas

There are two prerequisites in SAS-CKT: 1) the credibility of optimization experience hidden in the source tasks and 2) the sufficiency of the similarity for the optimum equivalence.

- 1) The credibility of optimization experience ensures the improvement in (8) for analogizing the external improvement, which can be solidified further if the global convergence of SAS is proved. It also fulfills the optimality needed by the optimum equivalence.
- 2) The sufficiency of the similarity for the optimum equivalence is essential for the analogy: the entire improvement on the source task is used to infer the external improvement if the similarity value is one, as the optima of the source and target tasks are deemed identical in this case.

With the Lipschitz condition, the satisfaction of the two prerequisites can be analytically proved. First, the discretization of a Lipschitz continuous function is presented in Lemma 1, enabling us to deal with it in a finite search space. Then, the global convergence of SAS for the credibility of optimization experience and the sufficiency of the similarity for the optimum equivalence are confirmed by Lemmas 2 and 3, respectively. Lastly, the lower bound of the expected convergence gain brought by the knowledge competition is mathematically derived, with the analysis of positive convergence gain presented in Theorem 1. Fig. 3 sketches the relations between the three lemmas and the theorem. Due to the page limit, the three lemmas are provided in Section S-II of the supplementary document accompanying this article.

B. Analysis of Convergence Gain

It is assumed that the best objective values searched by SAS are subject to the exponential decay,³ which is given by

$$\gamma_t = \gamma_0 + \gamma_1 e^{-\lambda t}.$$

³The empirical evidence for supporting this postulation and the parameter estimation are provided in Section S-III of the supplementary document.

Suppose the exponential decay of the target task is given by $\gamma^t(\tau; \gamma_o^t, \gamma_i^t, \lambda^t)$, we can estimate the expected internal improvement from τ to $\tau + 1$ as follows:

$$\Delta_{\text{in}} = \gamma_\tau^t - \gamma_{\tau+1}^t.$$

For the i th source task characterized by $\gamma^i(\tau; \gamma_o^i, \gamma_i^i, \lambda^i)$, the expected external improvement brought by the optimized solution can be estimated as follows:

$$\Delta_{\text{ex}}^i = s_i^+ \left(\gamma_\tau^t - \gamma_{s_i(\tau+\Delta_\tau^*)}^t \right).$$

Given a knowledge base with k source tasks, the convergence gain brought by CKT can be estimated as follows:

$$\begin{aligned} \mathcal{C}(s_*) &= \max\{\Delta_{\text{in}}, \Delta_{\text{ex}}^{\max}\} - \Delta_{\text{in}} \\ &= \max\left\{0, s_*^+ \cdot \left[\gamma_\tau^t - \gamma_{s_*(\tau+\Delta_\tau^*)}^t \right] - [\gamma_\tau^t - \gamma_{\tau+1}^t]\right\} \\ &= \max\{0, \Psi(s_*, \gamma_o^t, \gamma_i^t, \lambda^t, \tau, \Delta_\tau^*)\} \end{aligned} \quad (12)$$

where s_* is the similarity of the source task with the maximum external improvement, which is termed *conditional similarity* for short throughout this article as it is conditioned on the improvement, Δ_τ^* denotes the estimated time interval. Suppose s_* of a knowledge base is a random variable, we define the expected convergence gain brought by CKT as follows.

Definition 1: For a target task and a knowledge base with the conditional similarity under a probability density function $f(s_*)$, the expected convergence gain brought by the knowledge competition to the target task at time τ is defined as

$$\mathbb{E}[\mathcal{C}(s_*)] = \int_{-1}^1 \mathcal{C}(s_*) f(s_*) ds_* = \int_0^1 \mathcal{C}^+(s_*) f(s_*) ds_*$$

where $\mathcal{C}^+ = \max\{0, \Psi(s_*, \gamma_o^t, \gamma_i^t, \lambda^t, \tau, \Delta_\tau^*)\}$ with $s_* \geq 0$.

It can be seen that the lower bound of the convergence gain is bounded by zero regardless of $f(s_*)$. Moreover, we show particular interest in conditions leading to positive convergence gain. A theorem in this respect is presented as follows.

Theorem 1: Given a target task and a knowledge base, the lower bound of the expected convergence gain brought by the knowledge competition is greater than zero if $f(s_*)$ of the knowledge base is bounded from zero within $[\tilde{s}_*, 1]$, where \tilde{s}_* is the root of the following equation:

$$\Psi(\tilde{s}_*, \gamma_o^t, \gamma_i^t, \lambda^t, \tau, \Delta_\tau^*) = 0.$$

Proof: With the non-negativity of \mathcal{C}^+ , we have

$$\mathbb{E}[\mathcal{C}(s_*)] = \sum_{i=1}^v \int_{\underline{s}_*^i}^{\bar{s}_*^i} \Psi(s_*, \gamma_o^t, \gamma_i^t, \lambda^t, \tau, \Delta_\tau^*) f(s_*) ds_*$$

where v is the total number of intervals in which Ψ is greater than 0, \underline{s}_*^i and \bar{s}_*^i denote the lower and upper bounds of the i th interval, respectively.

Given the known parameters $\zeta = \{\gamma_o^t, \gamma_i^t, \lambda^t, \tau, \Delta_\tau^*\}$, we denote Ψ with respect to s_* as $\psi(s_*; \zeta)$. Then, let us examine the monotonicity of ψ according to its first-order derivative

$$\frac{d\psi}{ds_*} = s_* \lambda^t (\tau + \Delta_\tau^*) \gamma_{s_*(\tau+\Delta_\tau^*)}^t + \left[\gamma_\tau^t - \gamma_{s_*(\tau+\Delta_\tau^*)}^t \right]. \quad (13)$$

According to (12), we find that the non-negativity of the convergence gain depends on the non-negativity of s_* and Δ_τ^* .

Then, with (13), we can deduce that the first derivative of ψ is non-negative if the following condition holds:

$$\tau < s_*(\tau + \Delta_\tau^*) \Rightarrow s_* > \frac{\tau}{\tau + \Delta_\tau^*}.$$

With the global convergence of the source tasks, we have

$$\lim_{\tau_{\max} \rightarrow \infty} \frac{\tau}{\tau + \Delta_\tau^*} = \lim_{\tau_{\max} \rightarrow \infty} \frac{\tau}{\tau + \tau_{\max} - \tau_v^*} = 0.$$

The non-negativity of the first-order derivative within $(0, 1]$ indicates that ψ is strictly increasing on $(0, 1]$. Thus, we have

$$\frac{d\psi}{ds_*} > 0 \quad \forall s_* \in (0, 1] \Rightarrow \psi(s_*; \zeta) > 0 \quad \forall s_* \in (\tilde{s}_*, 1].$$

Then, we can reformulate the expected gain as follows:

$$\begin{aligned} \mathbb{E}[\mathcal{C}(s_*)] &= \int_0^{\tilde{s}_*} \mathcal{C}^+(s_*) f(s_*) ds_* + \int_{\tilde{s}_*}^1 \mathcal{C}^+(s_*) f(s_*) ds_* \\ &= \int_{\tilde{s}_*}^1 \psi(s_*; \zeta) f(s_*) ds_*. \end{aligned}$$

Since $\psi(s_*; \zeta)$ is non-negative and $f(s_*)$ is bounded from zero within $[\tilde{s}_*, 1]$, we have

$$\mathbb{E}[\mathcal{C}(s_*)] = \int_{\tilde{s}_*}^1 \psi(s_*; \zeta) f(s_*) ds_* > 0.$$

■

C. Brief Discussion

In this part, we will briefly discuss the implications of Theorem 1. First, it should be noted that γ_o^t and γ_i^t do not influence the lower bound of the similarity for positive convergence gain. When analyzing the impacts of λ^t , τ and Δ_τ^* on \tilde{s}_* using the implicit term Ψ , we set τ to 1 for the sake of visualizing the joint impact of λ^t and Δ_τ^* . The impact of τ will be separately analyzed later. Fig. 4 illustrates the impacts of λ^t and Δ_τ^* on the lower bound of the conditional similarity for positive convergence gain, as presented in Fig. 4(a) and (c), respectively. The joint impact is demonstrated by a 3-D surface shown in Fig. 4(b), in which the lines with endpoints are used for indicating the similarity intervals for positive convergence gain (i.e., $[\tilde{s}_*, 1]$) and the black band for $\Delta_\tau^* < 1$ signifies the cancellation of knowledge transfer.

To separately examine the impacts of λ^t and Δ_τ^* , we investigate each of the two parameters by fixing the other one. Specifically, λ^t and Δ_τ^* are fixed to 0.16 and 50, respectively, as indicated by the dotted frames in Fig. 4(b). For the impact of λ^t shown in Fig. 4(a), we can observe the monotonically increasing trend of \tilde{s}_* with respect to λ^t . A question arises: what does the parameter λ^t represent? According to (7), we find that λ can reflect the decay rate, i.e., the bigger λ is, the faster the quantity vanishes. Since we model the best objective values found by SAS as the exponential decay, λ^t can reflect how difficult it is to optimize the target task, i.e., the bigger λ^t is, the easier the target task is to be optimized. Thus, the increasing trend in Fig. 4(a) makes intuitive sense: an easy task requires higher quality of knowledge (i.e., higher similarity) for positive convergence gain than a difficult task. By contrast, we can observe the monotonically

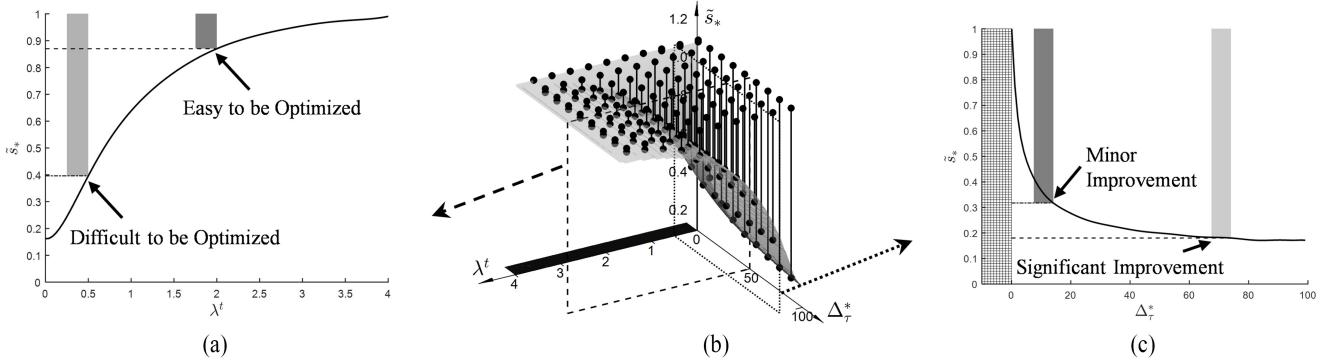


Fig. 4. Illustration of the impacts of λ^t and Δ_{τ}^* on the lower bound of the conditional similarity for positive convergence gain (i.e., \tilde{s}_*). (a) Lower bound of the similarity with respect to λ^t when $\Delta_{\tau}^* = 50$. (b) 3-D surface for demonstrating the joint impact of λ^t and Δ_{τ}^* on \tilde{s}_* . (c) Lower bound of the similarity with respect to Δ_{τ}^* when $\lambda^t = 0.16$.

decreasing trend of \tilde{s}_* with respect to Δ_{τ}^* in Fig. 4(c). The parameter Δ_{τ}^* indicates the potential improvement brought by the source optimized solution for the target search. This observation makes intuitive sense: the less significant the potential improvement of knowledge is, the higher \tilde{s}_* should be to achieve positive convergence gain. The impact of τ on \tilde{s}_* can be analyzed in the same way with the fixed λ^t and Δ_{τ}^* . The monotonically increasing trend of \tilde{s}_* with respect to τ indicates that the quality of knowledge should grow as the target search proceeds. This also makes intuitive sense: a solution that can speed up the target search at the early stage could be less promising or even harmful to the target search at the later stage, since the quality of target solutions gradually improves as the optimization proceeds.

Lastly, let us summarize the four important conditions for positive convergence gain, among which the first two conditions are target task-related while the last two conditions are knowledge base-related, as itemized as follows.

- 1) *Condition 1* ($\lambda^t < \infty$): This requires that the target task will not reach its global optimum immediately, leaving space for knowledge transfer to speed up the target search. A small λ^t caused by either difficult problems or weak optimizers will reduce \tilde{s}_* and make the knowledge transfer yield positive convergence gain more easily.⁴
- 2) *Condition 2* ($\tau < \infty$): This condition shows that the optimum of the target task has not been reached till time τ , indicating that there is still room for improvement through knowledge transfer. At the early optimization stage indicated by a small τ , the knowledge transfer can produce positive convergence gain more easily.
- 3) *Condition 3* ($\Delta_{\tau}^* > 1$): This requires that the potential improvement brought by the knowledge should be greater than the internal improvement brought by SAS, ensuring that the knowledge is qualified as the search experience from the source tasks to speed up the target search.
- 4) *Condition 4* ($s_* \geq \tilde{s}_*$): This condition requests the quality of knowledge to be transferred, enabling one to

analogize the potential improvements on the source tasks to the target task for positive convergence gain.

V. EXPERIMENTAL SETUP

This section presents the setup of our experiments, including the test problems, the backbone SAS solvers, and the general parameter settings.

A. Test Problems

In this study, the proposed methods are validated on a benchmark suite in [52], which contains 12 sequential transfer optimization problems (STOPs) with a broad spectrum of representation of the diverse similarity relationships between the optimal solutions of the source and target tasks in real-world problems. This broad representation enables a more comprehensive evaluation of various knowledge transfer methods [47], [53]. According to the similarity relationship, the 12 STOPs are divided into three categories: 1) high similarity (HS); 2) mixed similarity (MS); and 3) low similarity (LS). In this study, the number of source tasks is set to 50.

B. Backbone SAS Optimizers

In this study, we employ six backbone optimizers to validate the portability of CKT, including a standard Bayesian optimizer [48] (BO-LCB) and five state-of-the-art SAEAs listed in Table I, i.e., TLRBF [40], GL-SADE [41], DDEA-MESS [42], LSADE [43], and AutoSAEA [44]. An SAS optimizer suffixed with CKT (e.g., AutoSAEA-CKT) represents its knowledge competition-based version without the task adaptation, while a backbone solver suffixed with CKT-A denotes its use of the adaptation-based knowledge competition. For each STOP, its knowledge base is constructed by archiving the evaluated solutions obtained by BO-LCB on all the source tasks, i.e., $\mathcal{M} = \{\mathcal{D}_i^s : [X_i, y_i], i = 1, \dots, k\}$.

C. Parameter Settings

The general parameter settings are listed as follows.

- 1) The number of initial solutions provided by LHS: 50.
- 2) The population size of evolutionary solvers: 50.
- 3) The number of iterations of the evolutionary search: 50.

⁴The empirical evidence can be found in [47, Sec. VII].

TABLE II
WINS, TIES, AND LOSSES OF SAS-CKT AGAINST SAS WITH THE 6 BACKBONE OPTIMIZERS ON THE 12 TEST PROBLEMS

Optimizer	HS Problems	MS Problems	LS Problems	Summary
BO-LCB [48]	2/2/0	4/0/0	0/4/0	6/6/0
TLRBF [40]	3/1/0	4/0/0	1/3/0	8/4/0
GL-SADE [41]	0/4/0	1/3/0	0/4/0	1/11/0
DDEA-MESS [42]	2/2/0	2/2/0	0/4/0	4/8/0
LSADE [43]	3/1/0	3/1/0	1/3/0	7/5/0
AutoSAEA [44]	3/1/0	4/0/0	0/4/0	7/5/0
w/t/l (%)	54/46/0	75/25/0	8/92/0	46/54/0

- 4) The maximum number of true FEs: 500.
- 5) The transfer interval⁵ in terms of real evaluation (δ): 20.
- 6) The number of independent runs: 30.

More detailed parameter settings of the backbone SAS optimizers are provided in Section S-IV of the supplementary document.

VI. EXPERIMENTAL RESULTS AND ANALYSES

A. Portability of CKT

To demonstrate the portability of the proposed knowledge competition method, we summarize the wins, ties, and losses of SAS-CKT against SAS with the six backbone optimizers based on the significance test of the best objective values in Table II, whose detailed optimization results are provided in Section S-VI of the supplementary document due to the page limit. It can be seen that the knowledge competition yields a wide range of positive performance gain to the six backbone optimizers across the HS and MS problems while maintaining non-negative gain on all the LS problems, which validates its transfer learning capability to exhibit non-negative convergence gain when boosting different SAS engines. This indicates that the knowledge competition is able to boost the performance of contemporary SAS optimizers automatically in cases of HS or MS problems while exhibiting reasonably comparable performance to the backbone optimizers in the worst-case scenario of LS problems. Fig. 5 shows the averaged convergence curves of the six backbone solvers and their knowledge competition-based counterparts on four selected problems. It can be observed that the knowledge competition can achieve effective convergence speedups for the six SAS optimizes on the HS and MS problems while maintaining comparable convergence performance to the baseline solvers on the LS problems.

To investigate the underlying transfer mechanism of SAS-CKT, we show the averaged transfer rates of BO-LCB-CKT and AutoSAEA-CKT over 30 independent runs on a number of representative problems in Fig. 6. The transfer rate equals to one if $\Delta_{in} < \Delta_{ex}^{\max}$ while being zero otherwise. It can be seen that the knowledge competition exhibits different patterns of transfer rates on the three types of problems. Specifically, on the HS problems, we can observe the high

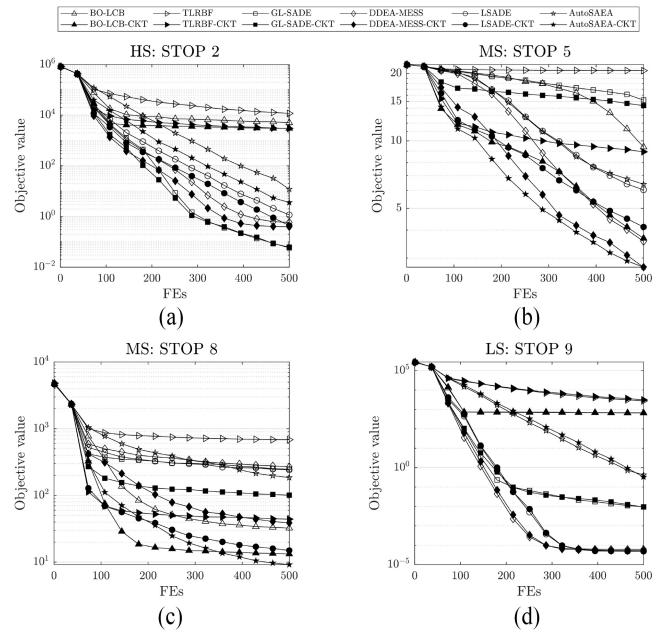


Fig. 5. Averaged convergence curves of SAS-CKT against SAS with the six backbone optimizers on four problems. (a) STOP 2. (b) STOP 5. (c) STOP 8. (d) STOP 9.

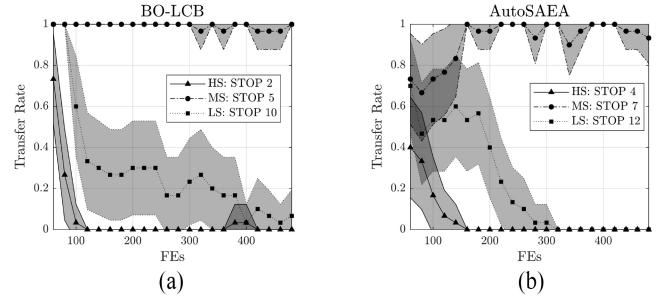


Fig. 6. Averaged transfer rates of BO-LCB and AutoSAEA based on the knowledge competition on a few representative problems. (a) BO-LCB. (b) AutoSAEA. The shaded area spans 1/2 standard deviation on either side of the mean.

transfer rates at the early stage, which are followed by their sharp declines. The early high transfer rates are largely due to the identification of the highly similar source tasks, while the sharp declines can be attributed to the automatic closure of knowledge transfer by (11) for avoiding the evaluation of similar optimized solutions with minor external improvements. On the MS problems, we can observe the stably high transfer rates along the evolutionary search, which can be attributed to the variety of the optimized solutions from the source tasks with the mixed similarities to the target task. As for the LS problems, the decreasing rates result from the adaptive suppression of transferring the optimized solutions from the source tasks with the low similarities to the target task. Such adaptive closure of knowledge transfer makes SAS-CKT degrade into its backbone SAS, which explains their comparable performance on the LS problems. In summary, the transfer mechanism of SAS-CKT fits the nature of STOPS with respect to the similarity relationship.

⁵The sensitivity analysis on this parameter is provided in the supplement.

TABLE III
WINS, TIES, AND LOSSES OF SSRC AGAINST THE FOUR SIMILARITY METRICS UNDER DIFFERENT FEs ON THE MS PROBLEMS

Metric	MS Problems				
	FEs=100	FEs=200	FEs=300	FEs=400	FEs=500
MMD [54]	1/3/0	2/2/0	1/3/0	0/4/0	0/4/0
WD [55]	4/0/0	3/1/0	3/1/0	3/0/1	2/2/0
MM [56]	4/0/0	3/1/0	3/1/0	2/2/0	2/2/0
RD [57]	2/1/1	2/2/0	2/2/0	2/1/1	2/1/1
w/t/l (%)	69/25/6	63/37/0	56/44/0	44/44/12	38/56/6

B. Reliability of CKT

To validate the reliability of CKT in similarity quantification, we first examine the superiority of the proposed SSRC over a number of similarity metrics commonly used in transfer optimization. Then, SAS-CKT is compared with a state-of-the-art transfer algorithm for EOPs.

1) *Comparison With Other Similarity Metrics*: We employ four similarity metrics, including the maximum mean discrepancy (MMD) [54], Wasserstein distance (WD) [55], mixture model (MM) [56], and representation-based distance (RD) [57], to compare with the proposed SSRC.

The optimization results of BO-LCB-CKT equipped with the five similarity metrics are provided in the supplementary document due to the page limit, from which we find that the metrics except for WD achieve comparable results to SSRC. This indicates that the final optimization results of SAS-CKT are not sensitive to the choice of similarity metric. However, the performance of the five similarity metrics varies when the computational budget is limited. To investigate, we compare SSRC with the other four metrics under different real FEs on the MS problems and summarize its wins, ties, and losses in Table III. It can be seen that SSRC outperforms all the other similarity metrics in cases of limited computational budget and its superiority decreases with the available FEs. This advantage of SSRC can be attributed to its more reliable similarity quantification as compared to the other similarity metrics in cases of limited target data, enabling it to identify the promising optimized solutions to speed up the target search effectively at the early stage. This merit is of particularly great significance for EOPs with limited real evaluations available.

As the number of available FEs increases, the other similarity metrics can also identify the promising optimized solutions due to the sufficient target data enabling more accurate similarity measurements. That explains the comparable results obtained by the five similarity metrics under 500 FEs. To demonstrate further, we compare the averaged convergence curves of BO-LCB-CKT based on the five similarity metrics in Fig. 7. It can be seen that SSRC shows the best convergence speedup among the five similarity metrics, especially in cases of limited FEs.

2) *Comparison With the State-of-the-Art*: To validate the superiority of SAS-CKT over the MPSs in [32], we compare them on two MS problems in Fig. 8. Specifically, STOP 5 involves the intrafamily transfer and its source tasks have the

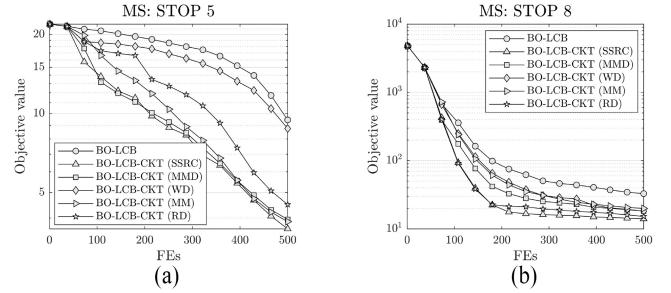


Fig. 7. Averaged convergence curves of BO-LCB and BO-LCB-CKT with the five similarity metrics on two problems. (a) STOP 5. (b) STOP 8.

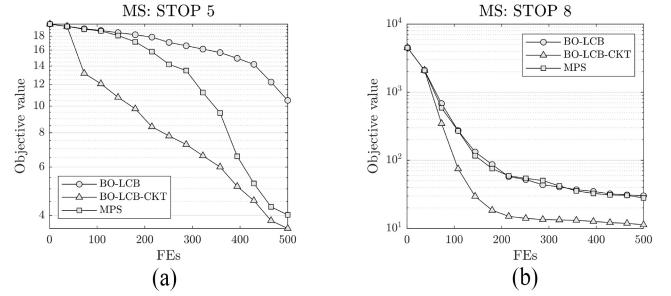


Fig. 8. Averaged convergence curves of BO-LCB, BO-LCB-CKT, and MPS on two problems. (a) STOP 5. (b) STOP 8.

same orders of magnitude to the target task. By contrast, the interfamily transfer in STOP 8 implies that the source and target tasks in this problem come from different task families and thus may bear HOM, which may pose a challenge to the similarity quantification for evaluating the source knowledge. To make fair comparisons, we employ BO-LCB as the backbone optimizer whose surrogate is consistent with MPS and equalize their surrogate-based evaluations in each acquisition cycle. The settings of the remaining parameters in MPS are consistent with its original paper [32].

Fig. 8 shows the averaged convergence curves of BO-LCB, BO-LCB-CKT, and MPS on STOP 5 and STOP 8, respectively. On STOP 5 with the intrafamily tasks, both BO-LCB-CKT and MPS achieve effective convergence speedups as compared to the backbone optimizer. Moreover, we can see that the proposed knowledge competition yields significantly higher convergence gain in comparison with MPS at the early stage of optimization, indicating its superior capability to identify the promising optimized solutions with limited target data. However, MPS fails to accelerate the target search on STOP 8 while the knowledge competition can still yield the positive convergence gain steadily, as shown in Fig. 8(b). This is because the similarity quantification in SAS-CKT is not sensitive to HOM while the performance of MPS could deteriorate a lot on problems characterized by HOM. For problems with HOM tasks, the objective value-based similarity quantification in MPS could make it underestimate the promising knowledge hidden in source tasks, e.g., STOP 8 in our experiments. In summary, the more reliable similarity quantification in SAS-CKT that uses ranks instead of exact objective values accounts for its superiority against MPS in assessing the knowledge from the source tasks.

TABLE IV
WINS, TIES, AND LOSSES OF SDA AGAINST THE FOUR DOMAIN ADAPTATION METHODS

Method	HS Problems	MS Problems	LS Problems	Summary
TCA [54]	1/3/0	4/0/0	1/3/0	6/6/0
AE [21]	2/2/0	3/1/0	1/2/1	6/5/1
AT [59]	2/2/0	4/0/0	1/2/1	7/4/1
SA [60]	1/3/0	3/1/0	1/3/0	5/7/0
w/l/t (%)	38/62/0	88/12/0	25/63/12	50/46/4

C. Adaptivity of CKT

To validate the adaptivity of CKT in adapting the optimized solutions from source tasks, we first examine the superiority of the proposed surrogate-based domain adaptation (SDA) over a few domain adaptation methods commonly used in transfer optimization [58]. Without loss of generality, we use the interior point algorithm to solve the problem in (5) based on the initial target data. Subsequently, we discuss the auxiliary role of SDA in SAS-CKT.

1) *Comparison With Other Domain Adaptation Methods*: We employ four domain adaptation methods, including the transfer component analysis (TCA) [54], autoencoder (AE) [21], affine transformation (AT) [59], and subspace alignment (SA) [60], to compare with SDA. The wins, ties, and losses of SDA against the four domain adaptation methods are reported in Table IV.⁶ We can observe that SDA outperforms all the other domain adaptation methods on both the HS and MS problems. This advantage of SDA is largely attributed to its mild yet appropriate task adaptation that preserves the high transferability of the promising optimized solutions in the HS or MS problems, enabling it to achieve the best final results among the five domain adaptation methods. On the LS problems, SDA, AE, and AT achieve the best optimization results on STOP 9, STOP 11 and STOP 12, respectively, indicating that these methods exhibit distinct advantages when adapting the optimized solutions of different LS problems.

2) *Auxiliary Role of SDA*: To investigate the effect of the task adaptation on the CKT, we compare BO-LCB with its adaptation-free and adaptation-based knowledge competition methods, as denoted by BO-LCB-CKT and BO-LCB-CKT-A, respectively. Fig. 9 demonstrates the comparison results based on the Wilcoxon rank-sum test of their final objective values. Overall, both CKT and CKT-A are effective in speeding up the target search of BO-LCB, which yield positive performance gain across the majority of the HS and MS problems while exhibiting comparable optimization results to BO-LCB on the remaining problems. Interestingly, there are two problems wherein CKT and CKT-A show distinguishably dominant convergence speedups, i.e., STOP 2 and STOP 9, as framed by the dotted boxes shown in Fig. 9. This tradeoff between the activation and cancellation of the task adaptation can be attributed to the inductive biases in the adaptation-free and adaptation-based methods. For the adaptation-free method, i.e., BO-LCB-CKT, its identity transformation (i.e., no adaptation)

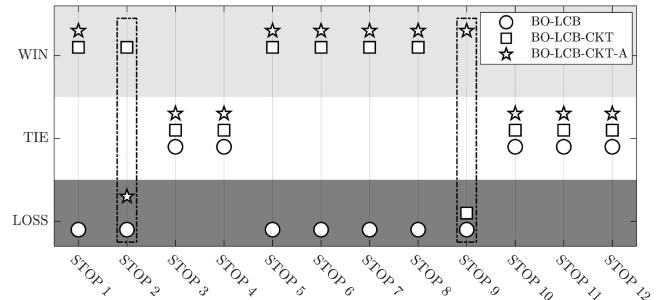


Fig. 9. Comparison results of BO-LCB, BO-LCB-CKT, and BO-LCB-CKT-A on the 12 test problems.

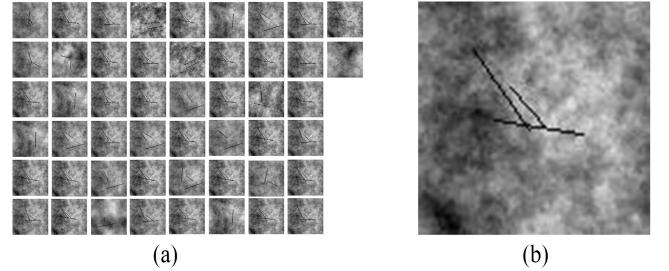


Fig. 10. Permeability fields of the source and target tasks. (a) 50 source tasks. (b) Target task.

enables it to focus on measuring the quality of knowledge without paying attention to adapting the knowledge, which fits the nature of HS or MS problems. By contrast, the learnable transformation in BO-LCB-CKT-A makes it put more emphasis on adapting the knowledge, which fits the nature of LS problems.⁷ In summary, the task adaptation can be seen as an auxiliary component of SAS-CKT, whose configuration is supposed to be driven by one's prior knowledge of the similarity relationship of STOPS at hand. In cases of unknown similarity relationship, both CKT and CKT-A are available for boosting SAS algorithms.

D. Practical Case Study

Production optimization in the petroleum industry is a representative EOP that involves high-fidelity numerical simulations, which also necessitates frequent reoptimization in view of newfound reservoirs.

As time goes by, an increasing number of developed reservoirs with available optimization experience accumulate in a database managed by the data processing center, providing an opportunity to the reservoir of interest for achieving better development effect through knowledge transfer. Thus, we use production optimization to evaluate the efficacy of the proposed method, whose ultimate goal is to search for the optimal production scheme that maximizes the development effect. The net present value (NPV) [61] is commonly used as the objective function for measuring the development effect, which is given by

$$\text{NPV}(v) = \sum_{t=1}^T Q_{po}^t r_{ro} + Q_{pw}^t r_{wd} + Q_{iw}^t r_{wi} \quad (14)$$

⁷More discussions in this respect can be found in [47, Sec. VI-B].

⁶The detailed optimization results are provided in the supplement.

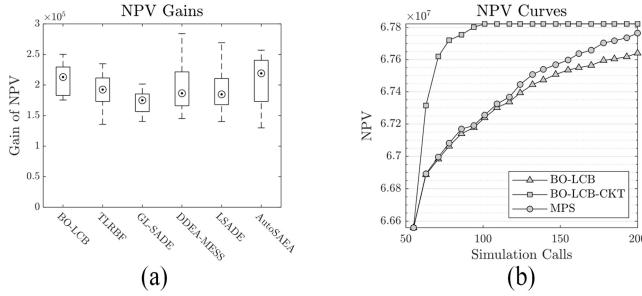


Fig. 11. NPV gains obtained by SAS-CKT based on the six SAS optimizers and averaged NPV curves of BO-LCB, BO-LCB-CKT, and MPS. (a) NPV gains. (b) NPV curves.

where v is the production scheme that consists of production and injection rates, t denotes the t th timestep, T is the total number of timesteps, Q_{po}^t and Q_{pw}^t represent the oil and water outputs of all the producers at the t th timestep, Q_{iw}^t denotes the water input of all the injectors at the t th timestep, r_o is the oil revenue, r_{wd} is the cost of disposing of produced water, and r_{wi} denotes the water-injection cost.

In this case study, we have 50 previously-solved production optimization tasks corresponding to 50 developed reservoirs acting as the source tasks and one reservoir model at hand serving as the target task, each of which has a 2-D grid pattern of $100 \times 100 \times 1$ and is developed by the nine-spot well pattern with five producers and four injectors. The permeability fields of the source and target tasks are shown in Fig. 10. For the sake of brevity, the remaining reservoir properties are not reported herein but can be found in [61]. The maximum number of simulation calls (i.e., expensive FEs) is set to 200 and all the algorithms are executed for 30 independent runs.

Fig. 11 shows the NPV gains achieved by SAS-CKT across the six backbone optimizers and the averaged convergence curves of BO-LCB, BO-LCB-CKT, and MPS over 30 independent runs. The NPV gain here is defined as the NPV improvement made by an SAS-CKT algorithm as compared with its backbone SAS optimizer. According to the results in Fig. 11(a), we can see that the NPV gains obtained by the knowledge competition for all the six backbone optimizers are significantly greater than zero, validating its efficacy in boosting different SAS algorithms on the production optimization. From the averaged NPV curves in Fig. 11(b), we can see that the knowledge competition shows better performance in terms of both convergence speed and solution quality as compared to the backbone BO-LCB and MPS, enabling more efficient decision-making for such computationally expensive production optimization problems.

VII. CONCLUSION

In this article, we have developed a novel CKT method for improving SAS algorithms, which is capable of leveraging the optimization experience hidden in previously-solved tasks to improve the optimization performance on a task at hand. The proposed CKT differs from the preceding methods due to its three merits itemized as follows.

- 1) *Portability*: The plug and play nature of CKT enables it to boost different SAS algorithms, making it coherent with the state-of-the-art in surrogate-assisted optimization.
- 2) *Reliability*: The rank-based similarity makes CKT reliable in assessing the knowledge from source tasks, especially in the case of source-target tasks with HOM.
- 3) *Adaptivity*: The task adaptation enables CKT to adapt the knowledge for higher transferability, making it applicable to problems with heterogeneous tasks.

The core of CKT is treating both the optimized solutions from the source tasks and the promising solutions acquired by the target surrogate as task-solving knowledge, enabling us to assess them from a consistent view. In this way, the most promising solution will be identified for true evaluation to maximize the search speed on the target task. Specifically, a CAE method has been proposed to enable the competition between the optimized solutions from the source tasks and the searched solutions in the target task. The efficacy of the proposed method has been validated empirically on a series of synthetic benchmarks and a practical application from the petroleum industry. In addition, we have conducted theoretical analyses on the convergence gain brought by the knowledge competition, including the lower bound of the convergence gain and the important conditions that could yield positive convergence gain. We anticipate that these analyses can strengthen the theoretical foundation of sequential transfer optimization and help researchers gain a deeper understanding of the underlying conditions contributing to effective knowledge transfer.

Despite the aforementioned merits of the proposed CKT, it exhibits certain limitations, e.g., the zero convergence gain on the LS problems. On such problems, some novel transferable cues that are not reliant on the similarity in terms of optimum could potentially yield positive convergence gain to the target search, which would be investigated in our future work. Moreover, we intend to generalize CKT to combat more problem complexities including, but not limited to, objective conflicts in multiobjective problems, high-dimensionality in large-scale problems, and constraint satisfaction in constrained problems. We also show particular interest in the generalization of CKT to other transfer optimization paradigms, e.g., multitask and multiform optimization.

REFERENCES

- [1] 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.
- [2] Y. Jin, H. Wang, T. Chugh, D. Guo, and K. Miettinen, "Data-driven evolutionary optimization: An overview and case studies," *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 442–458, Jun. 2019.
- [3] J.-Y. Li, Z.-H. Zhan, and J. Zhang, "Evolutionary computation for expensive optimization: A survey," *Mach. Intell. Res.*, vol. 19, no. 1, pp. 3–23, 2022.
- [4] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020.
- [5] M. Islam, A. Buijk, M. Rais-Rohani, and K. Motoyama, "Simulation-based numerical optimization of arc welding process for reduced distortion in welded structures," *Finite Elements Anal. Des.*, vol. 84, pp. 54–64, Jul. 2014.

- [6] H. C. Yoruklu, B. Ozkaya, and A. Demir, "Optimization of liquid fertilizer production from waste seaweed: A design of experiment based statistical approach," *Chemosphere*, vol. 286, Jan. 2022, Art. no. 131885.
- [7] S. Liu, Q. Lin, J. Li, and K. C. Tan, "A survey on learnable evolutionary algorithms for scalable multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 27, no. 6, pp. 1941–1961, Dec. 2023.
- [8] D. Lim, Y. Jin, Y.-S. Ong, and B. Sendhoff, "Generalizing surrogate-assisted evolutionary computation," *IEEE Trans. Evol. Comput.*, vol. 14, no. 3, pp. 329–355, Jun. 2010.
- [9] A. Źychowski, A. Gupta, J. Mañdziuk, and Y. S. Ong, "Addressing expensive multi-objective games with postponed preference articulation via memetic co-evolution," *Knowl.-Based Syst.*, vol. 154, pp. 17–31, Aug. 2018.
- [10] C. He, Y. Zhang, D. Gong, and X. Ji, "A review of surrogate-assisted evolutionary algorithms for expensive optimization problems," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119495.
- [11] G. E. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters*, vol. 664. New York, NY, USA: Wiley, 1978.
- [12] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2. Cambridge, MA, USA: MIT Press, 2006.
- [13] R. L. Hardy, "Multiquadric equations of topography and other irregular surfaces," *J. Geophys. Res.*, vol. 76, no. 8, pp. 1905–1915, 1971.
- [14] Q. Zhang, W. Liu, E. Tsang, and B. Virginas, "Expensive multiobjective optimization by MOEA/D with gaussian process model," *IEEE Trans. Evol. Comput.*, vol. 14, no. 3, pp. 456–474, Jun. 2009.
- [15] M. Sun, C. Sun, X. Li, G. Zhang, and F. Akhtar, "Surrogate ensemble assisted large-scale expensive optimization with random grouping," *Inf. Sci.*, vol. 615, pp. 226–237, Nov. 2022.
- [16] S. Wang, J. Liu, and Y. Jin, "Surrogate-assisted robust optimization of large-scale networks based on graph embedding," *IEEE Trans. Evol. Comput.*, vol. 24, no. 4, pp. 735–749, Aug. 2020.
- [17] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, "Recent advances in Bayesian optimization," *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–36, 2023.
- [18] H. Zhu, X. Wang, and Y. Jin, "Federated many-task Bayesian optimization," *IEEE Trans. Evol. Comput.*, vol. 28, no. 4, pp. 980–993, Aug. 2024.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] M. Shakeri, E. Miahia, A. Gupta, and Y.-S. Ong, "Scalable transfer evolutionary optimization: Coping with big task instances," *IEEE Trans. Cybern.*, vol. 53, no. 10, pp. 6160–6172, Oct. 2023.
- [23] J. Liu, A. Gupta, C. Ooi, and Y.-S. Ong, "ExTrEMO: Transfer evolutionary multiobjective optimization with proof of faster convergence," *IEEE Trans. Evol. Comput.*, early access, Jan. 2, 2024, doi: [10.1109/TEVC.2023.3349313](https://doi.org/10.1109/TEVC.2023.3349313).
- [24] 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.
- [25] S. Liu, Q. Lin, L. Feng, K.-C. Wong, and K. C. Tan, "Evolutionary multitasking for large-scale multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 27, no. 4, pp. 863–877, Aug. 2023.
- [26] Z. Wang, L. Cao, L. Feng, M. Jiang, and K. C. Tan, "Evolutionary multitask optimization with lower confidence bound-based solution selection strategy," *IEEE Trans. Evol. Comput.*, early access, Jan. 2, 2024, doi: [10.1109/TEVC.2023.3349250](https://doi.org/10.1109/TEVC.2023.3349250).
- [27] Z. Guo, H. Liu, Y.-S. Ong, X. Qu, Y. Zhang, and J. Zheng, "Generative multi-form Bayesian optimization," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4347–4360, Jul. 2023.
- [28] Y. Feng, L. Feng, S. Kwong, and K. C. Tan, "A multi-form evolutionary search paradigm for bi-level multi-objective optimization," *IEEE Trans. Evol. Comput.*, early access, Nov. 15, 2023, doi: [10.1109/TEVC.2023.3332676](https://doi.org/10.1109/TEVC.2023.3332676).
- [29] C. Yang, J. Ding, Y. Jin, and T. Chai, "Offline data-driven multiobjective optimization: Knowledge transfer between surrogates and generation of final solutions," *IEEE Trans. Evol. Comput.*, vol. 24, no. 3, pp. 409–423, Jun. 2020.
- [30] Y. Feng, L. Feng, S. Kwong, and K. C. Tan, "A multivariation multifactorial evolutionary algorithm for large-scale multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 26, no. 2, pp. 248–262, Apr. 2022.
- [31] X. Ji, Y. Zhang, D. Gong, X. Sun, and Y. Guo, "Multisurrogate-assisted multitasking particle swarm optimization for expensive multimodal problems," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2516–2530, Apr. 2023.
- [32] A. T. W. Min, Y.-S. Ong, A. Gupta, and C.-K. Goh, "Multiproblem surrogates: Transfer evolutionary multiobjective optimization of computationally expensive problems," *IEEE Trans. Evol. Comput.*, vol. 23, no. 1, pp. 15–28, Feb. 2019.
- [33] A. T. W. Min, A. Gupta, and Y.-S. Ong, "Generalizing transfer Bayesian optimization to source-target heterogeneity," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1754–1765, Oct. 2021.
- [34] X. Yan, J. Zhu, M. Kuang, and X. Wang, "Aerodynamic shape optimization using a novel optimizer based on machine learning techniques," *Aerosp. Sci. Technol.*, vol. 86, pp. 826–835, Mar. 2019.
- [35] J. Qi et al., "An evolutionary sequential transfer optimization algorithm for well placement optimization based on task characteristics," *SPE J.*, vol. 28, no. 3, pp. 1308–1328, 2023.
- [36] R. Li, I. Shikhov, and C. H. Arns, "Bayesian optimization with transfer learning: A study on spatial variability of rock properties using NMR relaxometry," *Water Resources Res.*, vol. 58, no. 9, 2022, Art. no. e2021WR031590.
- [37] G. Chen, J. J. Jiao, C. Jiang, and X. Luo, "Surrogate-assisted level-based learning evolutionary search for geothermal heat extraction optimization," *Renew. Sustain. Energy Rev.*, vol. 189, Jan. 2024, Art. no. 113860.
- [38] R. Bardenet, M. Brendel, B. Kégl, and M. Sebag, "Collaborative hyperparameter tuning," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 199–207.
- [39] D. Zhan and H. Xing, "A fast Kriging-assisted evolutionary algorithm based on incremental learning," *IEEE Trans. Evol. Comput.*, vol. 25, no. 5, pp. 941–955, Oct. 2021.
- [40] G. Li, Q. Zhang, Q. Lin, and W. Gao, "A three-level radial basis function method for expensive optimization," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 5720–5731, Jul. 2022.
- [41] W. Wang, H.-L. Liu, and K. C. Tan, "A surrogate-assisted differential evolution algorithm for high-dimensional expensive optimization problems," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2685–2697, Apr. 2023.
- [42] F. Yu, W. Gong, and H. Zhen, "A data-driven evolutionary algorithm with multi-evolutionary sampling strategy for expensive optimization," *Knowl.-Based Syst.*, vol. 242, Apr. 2022, Art. no. 108436.
- [43] J. Kudela and R. Matousek, "Combining Lipschitz and RBF surrogate models for high-dimensional computationally expensive problems," *Inf. Sci.*, vol. 619, pp. 457–477, Jan. 2023.
- [44] L. Xie, G. Li, Z. Wang, L. Cui, and M. Gong, "Surrogate-assisted evolutionary algorithm with model and infill criterion auto-configuration," *IEEE Trans. Evol. Comput.*, vol. 28, no. 4, pp. 1114–1126, Aug. 2024.
- [45] 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.
- [46] X. Xue et al., "Evolutionary sequential transfer optimization for objective-heterogeneous problems," *IEEE Trans. Evol. Comput.*, vol. 26, no. 5, pp. 1424–1438, Dec. 2022.
- [47] X. Xue, C. Yang, L. Feng, K. Zhang, L. Song, and K. C. Tan, "Solution transfer in evolutionary optimization: An empirical study on sequential transfer," *IEEE Trans. Evol. Comput.*, early access, Dec. 5, 2023, doi: [10.1109/TEVC.2023.3339506](https://doi.org/10.1109/TEVC.2023.3339506).
- [48] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016.
- [49] Z. Wang, B. Kim, and L. P. Kaelbling, "Regret bounds for meta Bayesian optimization with an unknown Gaussian process prior," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 10498–10509.
- [50] M. Wistuba, N. Schilling, and L. Schmidt-Thieme, "Two-stage transfer surrogate model for automatic hyperparameter optimization," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2016, pp. 199–214.
- [51] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian process meets big data: A review of scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4405–4423, Nov. 2020.

- [52] X. Xue, C. Yang, L. Feng, K. Zhang, L. Song, and K. C. Tan, "A scalable test problem generator for sequential transfer optimization," 2023, *arXiv:2304.08503*.
- [53] E. O. Scott and K. A. De Jong, "First complexity results for evolutionary knowledge transfer," in *Proc. 17th ACM/SIGEVO Conf. Found. Genet. Algorithms*, 2023, pp. 140–151.
- [54] M. Jiang, Z. Huang, L. Qiu, W. Huang, and G. G. Yen, "Transfer learning-based dynamic multiobjective optimization algorithms," *IEEE Trans. Evol. Comput.*, vol. 22, no. 4, pp. 501–514, Aug. 2018.
- [55] 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.
- [56] K. K. Bali, Y.-S. Ong, A. Gupta, and P. S. Tan, "Multifactorial evolutionary algorithm with online transfer parameter estimation: MFEA-II," *IEEE Trans. Evol. Comput.*, vol. 24, no. 1, pp. 69–83, Feb. 2020.
- [57] J. Yao et al., "Self-adaptive multifactorial evolutionary algorithm for multitasking production optimization," *J. Petrol. Sci. Eng.*, vol. 205, Oct. 2021, Art. no. 108900.
- [58] W. Lin, Q. Lin, L. Feng, and K. C. Tan, "Ensemble of domain adaptation-based knowledge transfer for evolutionary multitasking," *IEEE Trans. Evol. Comput.*, vol. 28, no. 2, pp. 388–402, Apr. 2024.
- [59] X. Xue et al., "Affine transformation-enhanced multifactorial optimization for heterogeneous problems," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6217–6231, Jul. 2022.
- [60] 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.
- [61] G. Chen et al., "Global and local surrogate-model-assisted differential evolution for waterflooding production optimization," *SPE J.*, vol. 25, no. 1, pp. 105–118, 2020.



Xiaoming Xue (Member, IEEE) received the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, SAR, China, in 2024.

His research interests include evolutionary transfer optimization, surrogate modeling, machine learning-aided optimization, and their practical applications.



Yao Hu (Member, IEEE) received the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, SAR, China, in 2024.

His research interests include domain adaptation, federated learning, evolutionary transfer optimization, and the applied deep learning for medical image analysis.



Liang Feng (Senior Member, IEEE) received the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore, in 2014.

He is currently a Professor with the College of Computer Science, Chongqing University, Chongqing, China. His research interests mainly include computational and artificial intelligence, big data optimization and learning, as well as transfer learning and optimization.

Prof. Feng has been honored with the 2019 IEEE TEVC Outstanding Paper Award, the 2023 IEEE TETCI Outstanding Paper Award, and the 2024 IEEE CIM Outstanding Paper Award.



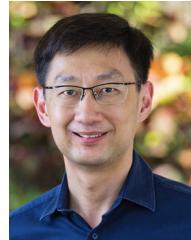
Kai Zhang (Member, IEEE) received the Ph.D. degree in oil-gas field development engineering from the China University of Petroleum (East China), Qingdao, China, in 2008.

From June 2007 to May 2008, he was with the University of Tulsa, Tulsa, OK, USA. In 2008, he joined as a Professor with the China University of Petroleum (East China). His research interests include reservoir simulation, production optimization, history matching, and development of nonconventional reservoir.



Linqi Song (Senior 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 from the University of California at Los Angeles, Los Angeles, CA, USA, in 2017.

He is an Associate 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. His research interests include information theory and machine learning.



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

He is currently a Chair Professor with the Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong. He currently serves as an Editorial Board member of 10+ journals.

Dr. Tan was an Editor-in-Chief of IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION from 2015 to 2020 and IEEE Computational Intelligence Magazine from 2010 to 2013. He is currently the Vice-President (Publications) of IEEE Computational Intelligence Society, an Honorary Professor with the University of Nottingham, U.K., and the Chief Co-Editor of Springer Book Series on Machine Learning: Foundations, Methodologies, and Applications.