IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 17, 2020, accepted October 2, 2020, date of publication October 12, 2020, date of current version October 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029968

Multi-Objective Sparse Reconstruction With Transfer Learning and Localized Regularization

BAI YAN[1,2], QI ZHAO[1,2], J. ANDREW ZHANG[3], (Senior Member, IEEE), AND ZHIHAI WANG[4]

[1]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[2]School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China
[3]Global Big Data Technologies Centre (GBDTC), University of Technology Sydney, Ultimo, NSW 2007, Australia
[4]Key Laboratory of Optoelectronics Technology, Ministry of Education, Beijing University of Technology, Beijing 100124, China

Corresponding author: Qi Zhao (zhaoq@sustech.edu.cn)

ABSTRACT Multi-objective sparse reconstruction methods have shown strong potential in sparse reconstruction. However, most methods are computationally expensive due to the requirement of excessive functional evaluations. Most of these methods adopt arbitrary regularization values for iterative thresholding-based local search, which hardly produces high-precision solutions stably. In this article, we propose a multi-objective sparse reconstruction scheme with novel techniques of transfer learning and localized regularization. Firstly, we design a knowledge transfer operator to reuse the search experience from previously solved homogeneous or heterogeneous sparse reconstruction problems, which can significantly accelerate the convergence and improve the reconstruction quality. Secondly, we develop a localized regularization strategy for iterative thresholding-based local search, which uses systematically designed independent regularization values according to decomposed subproblems. The strategy can lead to improved reconstruction accuracy. Therefore, our proposed scheme is more computationally efficient and accurate, compared to existing multi-objective sparse reconstruction methods. This is validated by extensive experiments on simulated signals and benchmark problems.

INDEX TERMS Sparse reconstruction, multi-objective evolutionary algorithm, transfer learning, regularization.

I. INTRODUCTION

The compressed sensing technology [1], [2] has been widely applied to many fields, such as medical imaging [3], [4] [5], face recognition, radar and sensor networks, and seismic data reconstruction [6], [7]. It reconstructs the sparse signal $\mathbf{x}$ from the measurements

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an unknown $k-$sparse signal, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the measurement matrix, $\mathbf{b} \in \mathbb{R}^m$ is the measurement vector, and $\mathbf{e} \in \mathbb{R}^m$ denotes the noise vector. Iterative thresholding algorithms [8], [9] are one dominant kind of sparse reconstruction (SR) methods. They formulates (1) to an unconstrained optimization problem

$$\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_p + \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2, \tag{2}$$

where $p \in (0, 1]$ is the norm, and λ is a predefined regularization parameter to balance the regularization term $\|\mathbf{x}\|_p$ and the measurement error $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$. The reconstruction performance is highly sensitive to λ [10], since it balances the feasibility and the sparsity of $\mathbf{x}$. Nowadays, several heuristic methods have been proposed for selecting λ, e.g., the projected generalized Stein unbiased risk estimator-based method [11], the method based on projection onto Epigraph Sets [12], Homotopy continuation methods [13], and the cross-validation approach [14].

To bypass the issue of selecting λ, problem (2) can be transferred to a bi-objective optimization problem (will be detailed in (3)), then being solved via the variants of multi-objective evolutionary algorithms (MOEAs), i.e., multi-objective SR (MOSR) approaches [15], [16] [17]. Compared with conventional SR methods, MOSR methods not only avoid setting the regularization parameters in the problem model but also is able to enforce the sparsity of $\mathbf{x}$ via the l_0-norm without relaxations, bringing enhanced reconstruction accuracy. The iterative search procedure of MOSR solvers generally involves

The associate editor coordinating the review of this manuscript and approving it for publication was Pietro Savazzi.

two parts, i.e., a population-based evolutionary search for global convergence, and a problem-specific local search for further refining the solutions. MOSR algorithms have been reported impressive performance in recent years. However, there are still some open challenges.

The first issue lies in the global search procedure. Evolutionary algorithms usually search from scratch, and MOSR optimizers are not exceptions. However, in practical reconstruction systems, the SR problems normally have similarities in terms of sparse pattern and evolutionary search process, etc. For example, in block-wise sparse reconstruction, an image block could easily find its similar structures in other blocks of the same image [18]. Considering this, the *knowledge* within the solutions to the previously solved SR problems can be used to facilitate the search of the current problem [19]. On the contrary, search from scratch may lead to unnecessary re-exploration in the space which is similar to previously solved problems.

The second challenge exists in local search operators. Conventional iterative thresholding methods [8], [9], [20], are generally integrated into MOSR solvers as local search operators [10], [15], [21]. Therefore, the local search process requires setting the regularization parameters. Generally, the regularization parameters are randomly selected within some ranges, ignoring the connection between the regularization term and measurement error. An inappropriately selected value can readily lead to bad recovery performance [15]. Therefore, it is desirable to develop more efficient and robust local search schemes to better deal with the SR problems.

In this article, we propose a novel MOSR scheme with innovations of transfer learning and localized regularization (referred to as MOSR-TLL hereafter), which is capable of solving the two aforementioned problems. The major contributions are summarized as follows.

- We introduce the idea of using a knowledge transfer operator to reuse the search experience from a previously solved MOSR problem. This operator can significantly accelerate the convergence of MOSR optimizers.
- We propose an efficient parallel search scheme, including a classic recombination operator and a novel knowledge transfer operator. The knowledge transfer operator includes three steps: (a) introducing the single-layer form of the deep nonlinear feature coding (DNFC) [22] to extract a feature mapping for the searching progress between the current and the past solved problems, (b) learning knowledge-induced solutions containing the structural search experience by the mapping and transferring them to the optimization procedure of the current problem, and (c) applying a sparse constraint strategy to ensure the sparse characteristics of the learned solutions. This scheme can accelerate the reconstruction speed of MOSR-TLL and improve the reconstruction quality.
- We propose a localized regularization strategy for the iterative thresholding-based local search process, which can greatly improve the reconstruction accuracy. In this strategy, the MOSR problem is divided into a number of subproblems by reference vectors and a carefully selected independent regularization value is applied according to each subproblem for local search.
- We conduct a comprehensive experimental study on various simulated signals and benchmark problems (through the Sparco toolbox) to verify the performance of the proposed MOSR-TLL. Extensive comparisons with state-of-the-art algorithms are also provided.

The rest of this article is organized as follows. Section II and III present related works and background knowledge. Section IV describes the proposed MOSR-TLL scheme. Experimental results are provided in Section V. Section VI further studies the impact of innovative technologies on system performance. Finally, Section VII concludes the paper.

II. RELATED WORKS

A. MULTI-OBJECTIVE SPARSE RECONSTRUCTION

The MOSR problem is formulated as

$$\mathbf{f}(\mathbf{x}) = \min_{\mathbf{x}}(\|\mathbf{x}\|_0, \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2), \tag{3}$$

where $\|\mathbf{x}\|_0$ and $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$ represent the sparsity of $\mathbf{x}$ and the measurement error, respectively. Since sparse reconstruction aims at finding the sparsest solution, it is natural to use the l_0-norm here. Nonetheless, the conventional SR methods solve the sparse reconstruction problems with relaxations (e.g., l_1, l_2 or $l_{1/2}$) to minimize the elements of x, but the relaxation approximation is not always equivalent.

A number of variants of MOEAs for sparse recovery have been developed. An early attempt is the soft-thresholding evolutionary multi-objective (StEMO) algorithm [15]. It integrates the soft thresholding method [8] into the NSGA-II framework [23]. A MOSR framework based on decomposition and a $l_{1/2}$ thresholding solver was suggested in [21]. A further improvement was proposed in [17], which searches the knee region with preference to reduce computational costs.

In [24], a hybrid approach with a modified linear Bregman method and differential evolution was provided. A two-phase evolutionary approach was proposed in [16], where the statistical feature of the non-dominated solutions was extracted, and non-zero entries were located. Later, an adaptive decomposition-based evolutionary approach (ADEA) [10] was provided. It not only searches the whole PF with the guidance of reference vectors, but also devotes additional search effort to the knee region. Different from these studies using genetic algorithms, the particle swarm optimization was also introduced for MOSR and achieved competitive performance [25], [26].

In the above works, iterative thresholding methods [8], [9] are generally employed for local search. The regularization parameters work as inputs of the local search, which are generally chosen randomly from some ranges or based on sparsity estimation. It is very likely to produce inferior solutions due to the ignorance of the connection

between the two objectives. Thus, it is expected to develop a more advanced local search operator with systematic devised regularization parameters to improve the SR performance.

B. KNOWLEDGE REUSE IN EVOLUTIONARY ALGORITHMS

Knowledge reuse strategies have been applied in various evolutionary optimization applications [19]. They utilize the knowledge from previously solved problems for enhancing the search efficiency of similar new problems. Early examples include direct reuse of previous solutions [27] and model-based knowledge transfer [28]. The former is unavailable for problems with different dimensions, while the latter cannot incorporate with model-free evolutionary algorithms. Recent strategies that reuse the structural information from a past problem are not subject to the above limitations and have gained promising performance [29]–[31].

In [30], a knowledge reuse strategy is incorporated with genetic programming for image classification tasks. The previously obtained tree-style code fragment was directly sent to the future problem-solving process. In [29], a single layer autoencoder [32] was adopted to extract knowledge from the past search experience. This method was further applied to multi-task problems [33]. These studies show efficacy in enhancing the evolutionary search speed, but the distribution difference between the solutions of problems is ignored.

In [34], new surrogate models are generated by stacking the previously obtained surrogates. This scheme saves the computational effort on building surrogates for computationally expensive problems. Similarly, literature [35] stacked the probabilistic density distribution of the solutions to past problems with that of the solutions to the current problem. The stacked distribution was used to guide the search for the current problem.

The knowledge transfer has also been conducted in the objective space. In [31], the latent objective space between the previous and the current problems were constructed by the transfer component analysis [36] method. Then the Pareto-optimal solutions of the past problem were transferred to the current problem through the latent space. In light of these advancements, we introduce the idea of transfer learning to MOSR for improving the recovery efficacy.

III. PRELIMINARIES

A. DECOMPOSITION-BASED MOEAs

Decomposition-based MOEAs transform a MOP into a number of subproblems with the help of reference vectors, then conducting a co-evolution of the subproblems [37]. Following the idea of decomposition, studies [38], [39] employ evenly distributed reference vectors to guide the evolutionary process, which can assure the convergence and diversity of solutions at a low computational cost. Some techniques in these algorithms that will be used in this article are briefly described below.

1) REFERENCE VECTORS

Reference vectors can be generated via the normal-boundary intersection method [40]. An example in a normalized 2-D objective space is shown in Fig. 1. N uniformly distributed reference points on a unit hyperball are generated [41]

$$\begin{cases} \mathbf{u}_j = (u_j^1, u_j^2), & j = \{1, 2, \dots N\} \\ u_j^h \in \{\dfrac{0}{H}, \dfrac{1}{H}, \dots, \dfrac{H}{H}\}, & \displaystyle\sum_{h=1}^{2} u_j^h = 1 \end{cases} \tag{4}$$

where $\mathbf{u}_j$ is the coordinate of the j-th reference point and represents the preferences for two objectives, N is the total number of reference points, $H = N - 1$ is a positive integer for the simplex-lattice design. Then, the unit vector from the coordinate origin to reference points $\{\mathbf{u}_j\}_{j=1}^N$ forms the reference vectors $\mathbf{V} = \{\mathbf{V}_0, \dots, \mathbf{V}_j, \dots \mathbf{V}_N\}$, where $\mathbf{V}_j = \mathbf{u}_j / \|\mathbf{u}_j\|$.

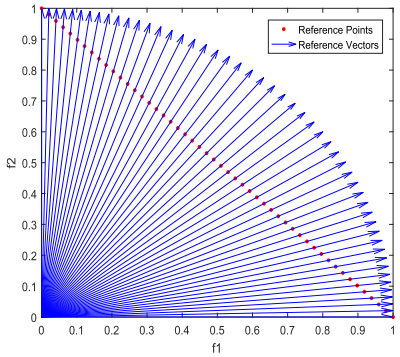


FIGURE 1. Reference vectors in a normalized 2-D objective space.

2) POPULATION NORMALIZATION

As the reference vectors exist in a $[0, 1]$ objective space, a normalization step [42] usually need to be applied to the objective function values of solutions. The M-th objective function value $f_M(\mathbf{x}_i)$ of a solution $\mathbf{x}_i$ is normalized to generate $\tilde{f}_M(\mathbf{x}_i)$ via

$$\tilde{f}_M(\mathbf{x}_i) = \frac{f_M(\mathbf{x}_i) - z_M^*}{z_M^{nad} - z_M^*} \tag{5}$$

where z_M^{nad} is the nadir point of the M-th dimension objective function in the current generation and can be obtained by the whole PF [43], and z_M^* is the minimum value of the M-th dimension objective function found so far.

3) SOLUTION ASSOCIATION

After normalization, the population is divided into a number of subpopulations in the following way: for a solution $\mathbf{x}_i \in \mathbf{P}$, the intersection angle θ between $\tilde{f}(\mathbf{x}_i)$ and each reference vector in $\mathbf{V}$ is obtained, then, $\mathbf{x}_i$ is associated with $\mathbf{V}_j$ if and only if their intersection angle is minimal. Therefore, $\mathbf{P}$ is partitioned into several subpopulations $\{\mathbf{P}_j\}_{j=1}^N$ by reference

Algorithm 1 The IHalfT-BB Algorithm

Require: $\mathbf{x}^l$, $\mathbf{x}^{l-1}$, λ, $\mathbf{A}$, $\mathbf{b}$

1: $\alpha^l = \frac{(\mathbf{s}^l)^T\mathbf{r}^l}{(\mathbf{s}^l)^T\mathbf{s}^l}$, where $\mathbf{s}^l = \mathbf{x}^l - \mathbf{x}^{l-1}$, $\mathbf{r}^l = \bigtriangledown\mathbf{F}(\mathbf{x}^l) - \bigtriangledown\mathbf{F}(\mathbf{x}^{l-1})$;

2: $\mu^l = 1/\alpha^l$;

3: $\mathbf{x}^{l+1} = \mathcal{H}_{\mu^l\lambda}(\mathbf{x}^l + \mu^l\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}^l))$;

vectors. This process can be mathematically described as

$$\begin{cases} \theta_{\mathbf{x}_i,\mathbf{v}_j} = \arccos(\dfrac{\tilde{f}(\mathbf{x}_i)\cdot(\mathbf{V}_j)^T}{\|\tilde{f}(\mathbf{x}_i)\|}), \quad j = \{1, 2, \ldots N\}, \\ \mathbf{P}_j = \{\mathbf{x}_i | j = \arg\min \theta_{\mathbf{x}_i,\mathbf{v}_j}\}. \end{cases} \tag{6}$$

B. ITERATIVE HALF THRESHOLDING ALGORITHMS

The iterative half thresholding (IHalfT) algorithm [9] solves the $L_{1/2}$-norm problem (2) in the form of

$$\mathbf{x}^{l+1} = \mathcal{H}_{\mu^l\lambda}(\mathbf{x}^l + \mu^l\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}^l)) \tag{7}$$

with an operator

$$\mathcal{H}_{\mu^l\lambda}(v_i^{l+1}) = \begin{cases} \chi(v_i^l), & |v_i^l| > \dfrac{\sqrt[3]{54}}{4}(\mu^l\lambda)^{\frac{2}{3}} \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where v_i^l is the i-th element of $\mathbf{v}$ at the l-th iteration, μ^l is the step size, $\chi(v_i^l) = (2/3)v_i^l(1 + \cos(2\pi/3 - (2/3)\varphi(v_i^l)))$ with $\varphi(v_i^l) = \arccos((\mu^l\lambda/8)(|v_i^l|/3)^{(-3/2)})$. The regularization value λ can be either selected as a constant value or determined by the prior estimation of signal sparsity. μ^l can also be either a constant or one related to the Lipschitz constant of the negative gradient $\bigtriangledown\mathbf{F} = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$.

In this article, we adopt an IHalfT algorithm based on the intelligent and accelerated Barizilai-Borwein (BB) method [44] because of its promising convergence performance [9]. The procedure is presented in Algorithm 1, where the step size μ^l mimics the behavior of the Hessian.

IV. THE PROPOSED MOSR-TLL SCHEME

We first design a parallel search, including a classic recombination operator and a novel knowledge transfer operator. In the knowledge transfer process, a single-layer NFC technique [22] is used to transfer the search experience of a previously solved MOSR problem to the evolution of the current problem. Based on the decomposition idea, we then propose a localized regularization strategy to improve the local search efficiency, i.e., different regularization values are restricted to work within the local search of the solutions to different decomposed subproblems.

A. FRAMEWORK OF MOSR-TLL

The workflow of the proposed MOSR-TLL is depicted in Figure 2, with the corresponding pseudo-code given in Algorithm 2. For convenience, we denote the current MOSR problem as the *target problem* and name the previously solved

Algorithm 2 Framework of MOSR-TLL

Require: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{PS}$, N

1: /*Initialization*/
2: $\mathbf{P}^0 \leftarrow$ Population_Initialization;
3: $\mathbf{V} \leftarrow$ Reference_Vector_Initialization;
4: **while** "the stopping criterion is not met" **do**
5: $\quad\mathbf{Q}^t$ =Recombination($\mathbf{P}^t$);
6: $\quad$**if** "the criterion of knowledge transfer is met" **then**
7: $\quad\quad\mathbf{T}^t$ =Knowledge_Transfer($\mathbf{P}^t$, $\mathbf{PS}^t$, $\mathbf{PS}^{t_{\max}}$);
8: $\quad\quad\mathbf{C}^t$ =Crossover($\mathbf{T}^t$, $\mathbf{P}^t$);
9: $\quad\quad\mathbf{P}^t = \mathbf{P}^t \cup \mathbf{C}^t$;
10: $\quad$**end if**
11: $\quad\mathbf{P}^t = \mathbf{P}^t \cup \mathbf{Q}^t$;
12: $\quad\mathbf{L}^t$ =Localized_Regularization_Local_Search($\mathbf{P}^t$, $\mathbf{V}$);
13: $\quad\mathbf{P}^{t+1}$ =Selection($\mathbf{P}^t \cup \mathbf{L}^t$);
14: $\quad t = t + 1$;
15: **end while**
16: $\mathbf{x} \leftarrow$ Final_Solution_Identification($\mathbf{P}^t$);

problem which we intend to transfer knowledge from as the *source problem*. $\mathbf{P}^t$ and $\mathbf{PS}^t$ are the sets of solutions to the target and source problems at generation t, respectively, and $\mathbf{PS}^{t_{\max}}$ is the set of the optimized solutions to the source problem. Let n_s and n_t denote the *population size*, i.e., the number of solutions of $\mathbf{PS}^t$ and $\mathbf{P}^t$, respectively. The procedure of MOSR-TLL is described below.

1) *Initialization*: An initial population $\mathbf{P}^0$ of size N is randomly generated, and N reference vectors uniformly distributed in the [0, 1] normalized objective space are produced (refer to Section III-A1).
2) *Parallel Search*: On one hand, the recombination operators, including the simulated binary crossover [45] and polynomial mutation [46], are employed to create the offspring (line 5 of Algorithm 2). On the other hand, the proposed knowledge transfer operator is implemented to obtain a set of sparsified knowledge-induced solutions $\mathbf{T}^t$, and the solutions from $\mathbf{T}^t$ and $\mathbf{P}^t$ are used to integrate the learned search experience from the source problem into the evolution of the target problem via the simulated binary crossover (line 6-10 of Algorithm 2). This scheme can efficiently reduce the harm of over-fitting caused by knowledge transfer, thus guaranteeing the robustness of this algorithm.
3) *Local Search And Selection*: A localized regularization-based local search operator is then executed to boost the search efficiency. All these obtained solutions undergo the selection step to select a number of N solutions for the next generation.
4) *Termination*: When the iteration terminates, the Pareto knee solution is identified as the final solution by the kink method [47].

Next we will elaborate the major innovations (i.e., the knowledge transfer and localized regularization-based local search operators) and the selection operator.

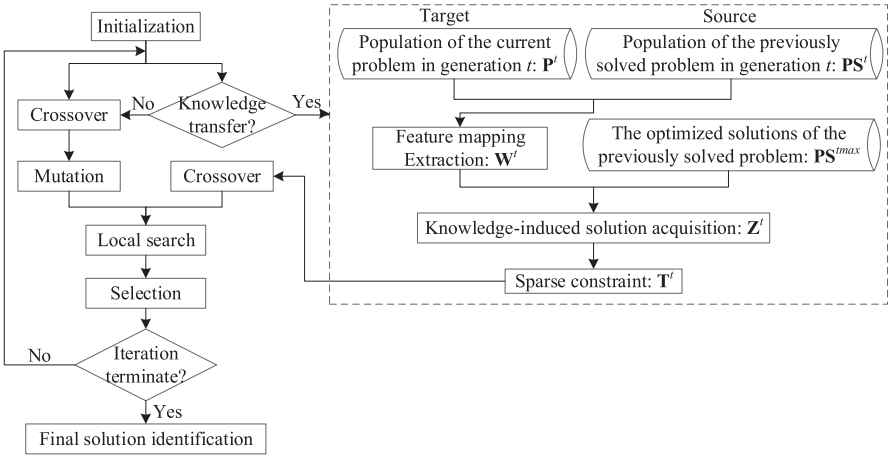


FIGURE 2. The workflow of the MOSR-TLL.

B. KNOWLEDGE TRANSFER

We present the detailed process of the knowledge transfer operator in this subsection. The knowledge transfer firstly constructs a mapping as the connection between the source problem and the target problem, then the optimized solutions of the source problem are mapped into the common latent space with this mapping, lastly the learned solutions are updated by inheriting the sparse structure of the current population. The obtained solutions are combined into the current problem-solving to bias the search process, which can speed up the convergence of the current problem and improve the reconstruction accuracy with great probability. Otherwise, the bad solutions yielded by negative knowledge transfer are filtered by the subsequent selection operator. As will be seen next, this operator provides a closed-form solution.

The first thing for knowledge transfer is how to obtain the source problem. One exemplified process is shown in Fig. 3, where q is the sequence number of the target problem. For the first target problem/task, a MOSR-TLL version without knowledge transfer (denoted as MOSR-L) is employed. The obtained sparse search experiences ($\{\mathbf{P}^1, \mathbf{P}^2, \ldots \mathbf{P}^{t_{max}}\}$) of this problem are saved to a source pool. Since the second target problem, MOSR-TLL can be employed with a source problem randomly selected from the source pool, and the new search experiences are also put into the source pool. Look back and forth, a source pool is enriched.

The source and target problems can be either homogeneous or heterogeneous. Here, these two problems are homogeneous (or heterogeneous) if the dimension of their sparse signals $\mathbf{x}$ is the same (or different). If $\mathbf{P}^t$ and $\mathbf{PS}^t$ have different dimensions, we pad $\mathbf{P}^t$ or $\mathbf{PS}^t$ with zeros to make them of equal dimensions. The procedure of the knowledge transfer operator is shown in Algorithm 3, with its three main steps detailed next.

Algorithm 3 Sparse-Constraint Knowledge Transfer

Require: $\mathbf{P}^t, \mathbf{PS}^t, \mathbf{PS}^{t_{\max}}, \theta$

/*Feature mapping extraction*/

$\mathbf{W}^t = \mathrm{E}[\mathbf{R}1](\mathrm{E}[\mathbf{R}2] + \theta\mathrm{E}[\mathbf{R}3])^{-1}$; // (12)

/*Knowledge-induced solutions acquisition*/

$\mathbf{Z}^t = \mathbf{W}_k^t \mathbf{K}(\mathbf{PS}^t \cup \mathbf{P}^t, \mathbf{PS}^{t_{\max}})$;

/*Solution sparsification*/

for $i = 1 : N$ **do**

$\mathcal{Z} \triangleq \{z | [\mathbf{P}^t]_{i,z} = 0\}$;

$$[\mathbf{T}^t]_{i,z} = \begin{cases} [\mathbf{Z}^t]_{i,z}, & z \notin \mathcal{Z}, \\ 0, & z \in \mathcal{Z}. \end{cases}$$

end for

1) EXTRACTION OF FEATURE MAPPING

This operation aims to evaluate a mapping $\mathbf{W}^t$ which makes the distribution of $\mathbf{P}^t$ and $\mathbf{PS}^t$ as similar as possible in the latent space. This mapping builds a connection for the searching progress between the source and target problems, which can be used for transferring knowledge across problems.

We employ the NFC technique [22] for acquiring the mapping $\mathbf{W}^t$. The principle of feature mapping extraction is provided in Fig. 4(a). The maximum mean discrepancy (MMD) [48] and nonlinear coding by kernelization are incorporated into the marginalized denoising autoencoder (mDA) [49] method, in which the MMD and kernelization can ensure the extracted features to have a small distribution discrepancy and the nonlinearity in the sparse data to be well exploited.

Let $\mathbf{X} \triangleq \mathbf{PS}^t \cup \mathbf{P}^t$, and $\overline{\mathbf{X}} = [\mathbf{X}, \mathbf{X}, \ldots \mathbf{X}]$ represent the union of the r-times copies. Let $\widetilde{\mathbf{X}^r}$ be the r-th corrupted version of $\mathbf{X}$ by random feature removal (i.e., each feature is set to 0 with probability g), $\overline{\widetilde{\mathbf{X}}} = [\widetilde{\mathbf{X}^1}, \widetilde{\mathbf{X}^2}, \ldots \widetilde{\mathbf{X}^r}]$ represent the union of r-times corrupted versions of $\mathbf{X}$, and $\Phi(\mathbf{X})$ is the mapped $\mathbf{X}$ in the Reproducing Kernel Hilbert Space $\mathcal{H}$. Let $\mathbf{W}^t = \mathbf{W}_k^t \Phi(\mathbf{X})^T$ and $\mathbf{K} = \Phi(\mathbf{X})^T \Phi(\mathbf{X})$ be the kernel matrix, we can formulate the objective function for extracting the mapping $\mathbf{W}^t$ as

$$\Gamma(\mathbf{W}^t) = \mathrm{Tr}\left((\overline{\mathbf{X}} - \mathbf{W}^t \widetilde{\Phi(\mathbf{X})})^T (\overline{\mathbf{X}} - \mathbf{W}^t \widetilde{\Phi(\mathbf{X})})\right)$$

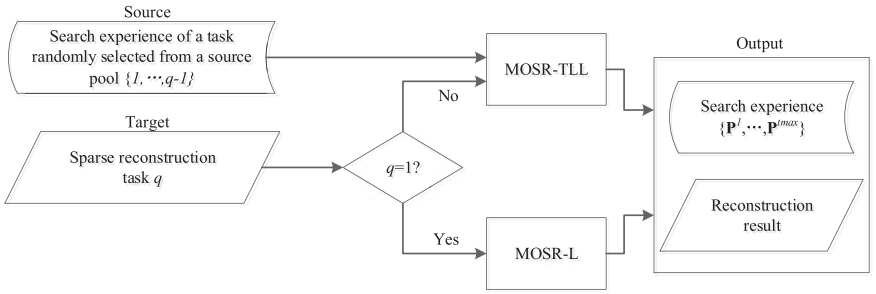


FIGURE 3. Illustration of how to obtain a source problem.

$$
\begin{aligned}
&+\|\frac{1}{n_s}\sum_{i=1}^{n_s}\mathbf{W}^t\widetilde{\Phi(\mathbf{X}_i)}-\frac{1}{n_t}\sum_{i=n_s+1}^{n_s+n_t}\mathbf{W}^t\widetilde{\Phi(\mathbf{X}_i)})\|^2\\
&=\underbrace{\mathrm{Tr}\left((\overline{\mathbf{X}}-\mathbf{W}_k^t\widetilde{\mathbf{K}})^T(\overline{\mathbf{X}}-\mathbf{W}_k^t\widetilde{\mathbf{K}})\right)}_{\mathrm{mDA}}\\
&+\theta\cdot\underbrace{\mathrm{Tr}(\mathbf{W}_k\widetilde{\mathbf{K}}\widetilde{\mathbf{G}}\widetilde{\mathbf{K}}^T\mathbf{W}_k^T)}_{\mathrm{MMD}}
\end{aligned}
\tag{9}
$$

where the last two terms indicate the mDA and MMD function, $\widetilde{\Phi(\mathbf{X})}$ and $\widetilde{\mathbf{K}}$ stand for the corrupted data with r different corruptions of $\Phi(\mathbf{X})$ and $\mathbf{K}$ respectively, $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, $\mathbf{G}=[\mathbf{G}_{i,j}]_{(n_s+n_t)\times(n_s+n_t)}$ with $G_{i,j}=1/n_s^2$ if $\mathbf{X}_{i,j}\in\mathbf{PS}^t$, $G_{i,j}=1/n_t^2$ if $\mathbf{X}_{i,j}\in\mathbf{P}^t$, and $G_{i,j}=-1/(n_sn_t)$ otherwise; and θ is a balancing parameter.

(9) has a closed-form solution:

$$
\mathbf{W}_k^t=\mathbf{R}_1(\mathbf{R}_2+\theta\mathbf{R}_3)^{-1} \tag{10}
$$

with $\mathbf{R}_1=\overline{\mathbf{X}}\widetilde{\mathbf{K}}^T$, $\mathbf{R}_2=\widetilde{\mathbf{K}}\widetilde{\mathbf{K}}^T$ and $\mathbf{R}_3=\widetilde{\mathbf{K}}\widetilde{\mathbf{G}}\widetilde{\mathbf{K}}^T$. Applying the weak law of large numbers and computing the expectations when the number of corruption $r\to\infty$, a robust and closed-form solution for $\mathbf{W}_k^t$ can be formulated as

$$
\mathbf{W}_k^t=\mathrm{E}[\mathbf{R}_1](\mathrm{E}[\mathbf{R}_2]+\theta\mathrm{E}[\mathbf{R}_3])^{-1} \tag{11}
$$

with

$$
\begin{cases}
\mathrm{E}[\mathbf{R}_1]=(1-g)\mathbf{X}\mathbf{K}^T\\
\mathrm{E}[\mathbf{R}_2]_{i,j}=\begin{cases}(1-g)^2\mathbf{K}\mathbf{K}^T, & i\neq j\\ (1-g)\mathbf{K}\mathbf{K}^T, & i=j\end{cases}\\
\mathrm{E}[\mathbf{R}_3]_{i,j}=\begin{cases}(1-g)^2\mathbf{K}\mathbf{G}\mathbf{K}^T, & i\neq j\\ (1-g)^2\mathbf{K}\mathbf{G}\mathbf{K}^T+g(1-g)\mathbf{K}\mathbf{F}\mathbf{K}^T, & i=j\end{cases}
\end{cases}
\tag{12}
$$

where $\mathbf{F}$ is a diagonal matrix having the same diagonal elements with $\mathbf{G}$. Due to page limitations, please refer to [22] for more details about the proof.

2) ACQUISITION OF KNOWLEDGE-INDUCED SOLUTIONS

$\mathbf{PS}^{t_{\max}}$ contains most of the search experience from the source problem. We can apply the experience to the target problem by mapping $\mathbf{PS}^{t_{\max}}$ into the latent space constructed by $\mathbf{W}_k^t$. As $\mathbf{W}_k^t$ is a connective mapping between $\mathbf{PS}^t$ and $\mathbf{P}^t$, we have

$$
\begin{aligned}
\mathbf{Z}^t&=\mathbf{W}^t\Phi(\mathbf{PS}^{t_{\max}})\\
&=\mathbf{W}_k^t\Phi(\mathbf{X})^T\Phi(\mathbf{PS}^{t_{\max}})\\
&=\mathbf{W}_k^t\mathbf{K}(\mathbf{X},\mathbf{PS}^{t_{\max}})
\end{aligned}
\tag{13}
$$

where $\mathbf{Z}^t$ contains a set of knowledge-induced solutions which incorporate the search experience from the source problem and also evaluates the similarity for the search progress between the source and target problems. This procedure is also illustrated in Fig. 4(b).

3) SPARSE CONSTRAINT

To ensure the sparsity characteristics of solutions in $\mathbf{Z}^t$, we combine the sparse structure of $\mathbf{P}^t$ into $\mathbf{Z}^t$ to obtain the sparse version of $\mathbf{Z}^t$, $\mathbf{T}^t$. As a result, $\mathbf{T}^t$ will not only possess the valuable knowledge extracted from the search experience for the target problem, but also inherit the sparse structure of $\mathbf{P}^t$.

In detail, firstly, we identify the positions of zero elements in $\mathbf{P}^t$, and form a set $\mathcal{Z}\triangleq\{z|[\mathbf{P}^t]_{i,z}=0\}$, where $[\cdot]_{i,z}$ represents the (i,z)-th element in a matrix. We then set the elements in $\mathbf{Z}^t$ with indexes belonging to $\mathcal{Z}$ to be zeros and generate $\mathbf{T}$ as

$$
[\mathbf{T}^t]_{i,z}=\begin{cases}[\mathbf{Z}^t]_{i,z}, & \text{if } z\notin\mathcal{Z}\\ 0, & \text{if } z\in\mathcal{Z}\end{cases}\quad i=\{1,2,\ldots N\}. \tag{14}
$$

An example of sparse constraint is depicted in Fig. 4(c).

C. LOCAL SEARCH BASED ON LOCALIZED REGULARIZATION

Using the two-stage iterative soft-thresholding (IST) based local search strategy in ADEA [10] as a baseline, we elaborate the proposed localized regularization-based local search, as shown in Algorithm 4. The population firstly undergoes normalization and solution association (refer to Section III-A) as a preparatory process. Then, the fitness function in [10] is employed to evaluate the fitness of each solution

$$
\mathrm{fitness}(\mathbf{x}_i)=c_{\mathbf{x}_i}+5\times d_{\mathbf{x}_i} \tag{15}
$$

where $c_{\mathbf{x}_i}=\|\tilde{f}(\mathbf{x}_i)\|\cos(\theta_{\mathbf{x}_i,\mathbf{v}_j})$, $\tilde{f}(\mathbf{x}_i)$ is the normalized bi-objective function value of solution $\mathbf{x}_i$ via (5), and $d_{\mathbf{x}_i}=$

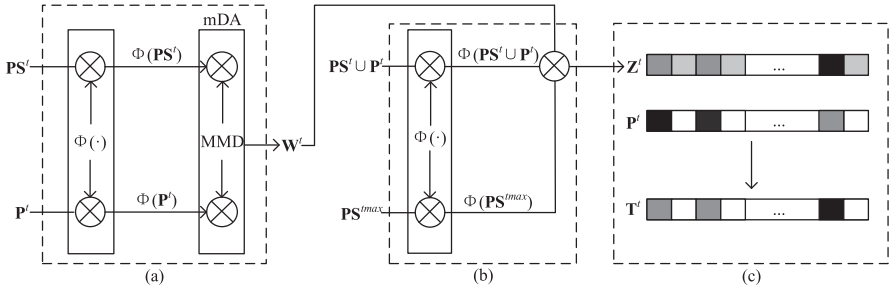


FIGURE 4. Illustration of the sparse-constraint transfer learning operator. (a) Extraction of feature mapping, (b) acquisition of knowledge-induced solutions, and (c) sparse constraint. Each lattice denotes an entry of the signal, and the grey and white grids represent the nonzero and zero entries, respectively.

Algorithm 4 Localized-Regularization-Based Local Search Operator

Require: $\mathbf{P}^t$, $\mathbf{V}^t$, N

1: $\mathbf{Q}$ =Population_Normalization($\mathbf{P}^t$);
2: $\mathbf{P}_j^t$ =Solution_Association($\mathbf{P}^t$,$\mathbf{Q}$, $\mathbf{V}^t$);
3: $\mathbf{B} = \{\mathbf{x}_{i,j} | i = \arg\min c_{x_i} + 5 \times d_{x_i}, \mathbf{x}_i \in \mathbf{P}_j^t, j = 1, 2, \ldots N\}$;
4: $\mathbf{S} = \mathbf{P}^t \setminus \mathbf{B}$;
5: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \ldots \lambda_N) \leftarrow$ Localized_Regularization;
6: /*Two-stage local search*/
7: **for** $i = 1 : |\mathbf{B}|$ **do**
8: $\quad$ **L1** =IhalfT-BB($\mathbf{B}_{i,j}$, $\mathbf{S}_i$, λ_j); // Algorithm 1
9: **end for**
10: **for** $i = 1 : \lfloor |\mathbf{B}|/2 \rfloor$ **do**
11: $\quad$ **L2** =IhalfT-BB($\mathbf{B}_{i,j}$, $\mathbf{B}_{|\mathbf{B}|-i}$, λ_j); // Algorithm 1
12: **end for**
13: $\mathbf{L}^t = \mathbf{L1} \cup \mathbf{L2}$;

$\|\tilde{f}(\mathbf{x}_i)\| \sin(\theta_{\mathbf{x}_i,\mathbf{v}_j})$ measure the convergence and diversity of $\mathbf{x}_i$ with $\mathbf{v}_j$ stands for the j-th reference vector. With this criterion, the solutions can be divided into two sets (line 3-4 of Algorithm 4): B (the set of the solution with the best fitness in each subpopulation) and S (the remaining solutions). It basically satisfies $B \preccurlyeq S$. After that, a localized regularization strategy (line 5) is proposed to define a set of the regularization parameters (denoted as $\boldsymbol{\lambda}$) for the subsequent two-stage local search, which will be detailed in the following paragraphs. Finally, the two-stage local search is executed via the IHalf-BB algorithm (i.e., Algorithm 1), in which the first stage (line 7-9) is for accelerating the convergence performance, and the second stage (line 10-12) for spreading the solutions over the PF.

Please note that there are two differences between our proposed local search operator and the IST-based local search in [10]. First, the IST-based local search utilizes the IST algorithm, while we integrate the IHalfT algorithm, as IHalfT is able to achieve higher convergence accuracy [9]. Second, both local search operators require the regularization values as an input, and the correctness of $\boldsymbol{\lambda}$ severely influences the quality of the produced solutions. ADEA assigns a random value to each solution for its local search, which is hardly to produce high-accuracy solutions stably. By contrast, we propose a localized regularization strategy, in which an independent regularization value is assigned to each subpopulation, and this value can be determined by using the preference of the corresponding subproblem over the two objectives as a *priori*. This strategy can boost the reconstruction accuracy of MOSR-TLL.

The proposed localized regularization strategy is motivated by the decomposition idea. By use of uniformly distributed reference vectors, the MOSR problem is divided into several subproblems, and the whole population is divided into some subpopulations. For the j-th reference vector, the corresponding subproblem is cast as

$$\tilde{\mathbf{f}}_j(\mathbf{x}_i) = \arg\min_{\mathbf{x}} \; u_j^1 \cdot \tilde{f}_1(\mathbf{x}_i) + u_j^2 \cdot \tilde{f}_2(\mathbf{x}_i) \tag{16}$$

where (u_j^1, u_j^2) is the coordinate of the j-th reference point and it represents the preference over the two objectives, a symbol with the tilde denotes its value "normalized" through (5), $\tilde{f}_1(\mathbf{x}_i)$ and $\tilde{f}_2(\mathbf{x}_i)$ represent the normalized $\|\mathbf{x}\|_0$ and $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$, respectively. (16) can be rewritten as

$$\tilde{\mathbf{f}}_j(\mathbf{x}) = \arg\min_{\mathbf{x}} \; \frac{u_j^1}{2u_j^2} \cdot \tilde{f}_1(\mathbf{x}_i) + \frac{1}{2} \cdot \tilde{f}_2(\mathbf{x}_i) \tag{17}$$

It is self-evident that both the $u_j^1/(2u_j^2)$ in (17) and the λ in equation (2) stand for the preference over two objectives. Therefore, it is natural to determine λ_j by using $u_j^1/(2u_j^2)$ as a priori, where λ_j is the regularization parameter for the j-th subproblem, all the regularization parameters make up $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots)$. To obtain $\boldsymbol{\lambda}$, we define a vector based on $u_j^1/2u_j^2$ and equation (4)

$$\frac{1}{2} \cdot (\frac{0}{H}, \frac{1}{H-1}, \ldots, \frac{j-1}{H-j+1}, \ldots, \frac{H-1}{1}, \frac{H}{0}) \tag{18}$$

"$\frac{H}{0}$" is meaningless, we will modify this value as follows. Referring to [50], the upper bound for λ_j is suggested to be $\|A^T b\|_\infty$ to avoid useless zero solutions. Therefore, we keep $\boldsymbol{\lambda}$ below the upper bound by multiplying (18) by $2\|A^T b\|_\infty/H$, and substitute $H = N - 1$ into (18)

$$\lambda_j = \max(\frac{j-1}{(N-j)(N-1)} \|A^T b\|_\infty, 1e-3), \tag{19}$$

where $j \in \{1, \ldots, N-1\}, 0 \leq \frac{j-1}{(N-j)(N-2)} \leq 1, 1e-3$ ensures λ_j not too small.

D. SELECTION

The selection operator from ADEA [10] is adopted in $\mathbf{P}^t \cup \mathbf{L}^t$ for selecting N promising solutions for the next generation. Firstly, solutions in $\mathbf{P}^t \cup \mathbf{L}^t$ are normalized (refer to Section III-A2) and associated with their closest reference vectors (refer to Section III-A3). In each subpopulation, the solutions are sorted in an ascending order according to their fitness evaluated by (15) and marked with level $\{1, 2, \ldots\}$ sequentially. Then, all the solutions are selected successively from low to high levels. If some of the solutions deduced by transfer learning survive in the selection procedure, the knowledge transfer is beneficial for the optimization of the current MOSR problem; otherwise, the learned solutions will not be sent to the next generation, avoiding negative transfer.

E. COMPUTATIONAL COMPLEXITY ANALYSIS

In Algorithm 2, the computational complexity of MOSR-TLL is dominated by two operators, namely, knowledge transfer and local search. At each iteration, the complexity of knowledge transfer is mainly determined by the inversion operation in (11), which is $O((n_s + n_t)^3)$ (n_s and n_t are the population size of the source problem and the target problem). The complexity of local search is $O(n^2)$, where n is the length of the signal of interest. Thus, the total computational cost is $O(n^2) + O((n_s + n_t)^3)$.

V. EXPERIMENTS AND DISCUSSIONS

In this section, experiments on simulated signals and benchmark problems (from Sparco toolbox [51]) are implemented to investigate the reconstruction performance of the proposed MOSR-TLL, compared with some state-of-the-art algorithms, including three single-objective sparse reconstruction algorithms (i.e., OMP [52], FISTA [53] and ISTC [13]) and two MOSR algorithms (i.e., StEMO [15] and ADEA [10]).

A. EXPERIMENTAL SETTING

For MOSR-TLL, parameters for the crossover and mutation operators are set with reference to [10]: crossover probability $p_c = 1.0$, distribution index of crossover $\eta_c = 20$, mutation probability $p_m = 1/n$ and the distribution index of mutation $\eta_m = 20$. The corruption probability g is set between 0.5 and 0.9 with a step of 0.1 via cross validation on the population of a past problem in the first generation. The balancing parameter θ and the kernel function are set to be 10^3 and 'rbf' respectively, with reference to [22]. The method of selecting a source problem is as same as Fig. 3.

For FISTA, λ is set to 0.02 according to [15]. For a fair comparison, the population sizes are set to 50 for StEMO and MOSR-TLL, and 40 for ADEA with 10 adaptive reference vectors.

The termination criterion can be either of the following two: (1) the variation of measurement errors in ten consecutive iterations are consistently smaller than a threshold, e.g., $< 10^{-6}$, or (2) the maximum generations reach 200. Each algorithm runs 30 times in a test case. The Wilcoxon Sign Test with 95% confidence level is conducted on the obtained results to show the statistical significance of the experimental comparison. The reconstruction error (RE), which measures the difference between the estimated signal $\mathbf{x}$ and the true one $\mathbf{x}_{true}$, is used as the performance metric. It is calculated as

$$\mathrm{RE} = \frac{\|\mathbf{x} - \mathbf{x}_{true}\|_2}{\|\mathbf{x}_{true}\|_2}. \tag{20}$$

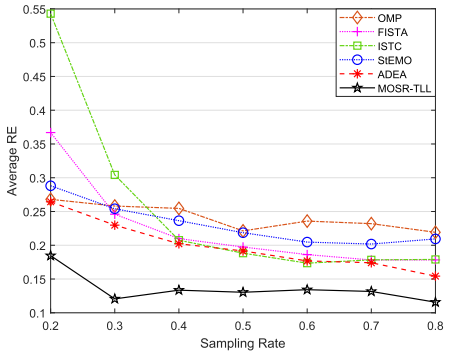
All algorithms were implemented in MATLAB R2016a on a PC with Intel Xeon E3 processor at 3.3 GHz, 8 GB memory, and the Window 7 operating system.

B. EXPERIMENTAL RESULTS ON SIMULATED SIGNALS

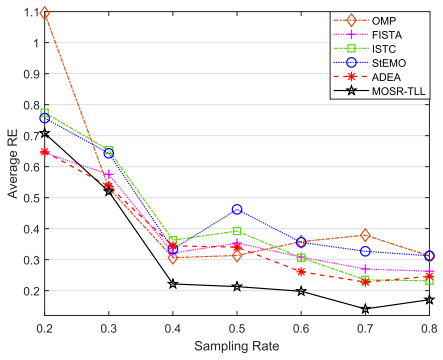
Test problems on simulated signals are randomly generated to verify the performance of MOSR-TLL. The process of constructing a test problem is as follows: a) A k-sparse signal $\mathbf{x}$ is produced, in which the nonzero elements are sampled from a Gaussian distribution $\mathcal{N}(0, 1)$. b) A Gaussian random matrix $\mathbf{A}$ is created and the measurement vector $\mathbf{b}$ is obtained by $\mathbf{b} = \mathbf{A}\mathbf{x}$. c) The measurement $\mathbf{b}$ is corrupted by additive white Gaussian noise with elements from the normal distribution $N(0, 0.01)$. The dimension n of the true signal is fixed to 1000. We investigate the average reconstruction performance of all the algorithms in terms of sampling ratios (m/n) and sparsity levels (k/n). The method of selecting the source problem is the same as Fig. 3. For the first test problem, we solve it by MOSR-L and save the search experiences into the source pool. Thereafter, for a subsequent test problem to be solved, we randomly choose a past solved problem from the source pool as the source, and collect the new search experiences to the source pool.

Fig. 5 shows the average REs for six algorithms under various sampling ratios from 0.2 to 0.8. The sparsity ratios in the two sub-graphs are set to 0.05 and 0.1 respectively. In Fig. 5(a), when the sparsity ratio equals to 0.05, MOSR-TLL is significantly better than other algorithms under all the sampling ratios. Similar results can be seen from Fig. 5(b). When the sparsity ratio is 0.10, the reconstruction performance of MOSR-TLL shows advantages over other other algorithms with the sampling ratio in [0.3,0.8].

Fig. 6 compares the performance for six algorithms when the sparsity ratio varies from 0.05 to 0.30. In Fig. 6(a), when the sampling rate equals to 0.4, MOSR-TLL achieves the best performance within the sparsity ratios [0.05, 0.20], except for FISTA in the sparsity ratio range of [0.25, 0.3]. Note that FISTA obtains better results here because $\lambda = 0.02$ happens to be the optimal choice for this case, but might perform badly in other cases. In Fig. 6(b), where the sampling ratio is increased to 0.6, MOSR-TLL achieves the best performance over all sparsity ratios.

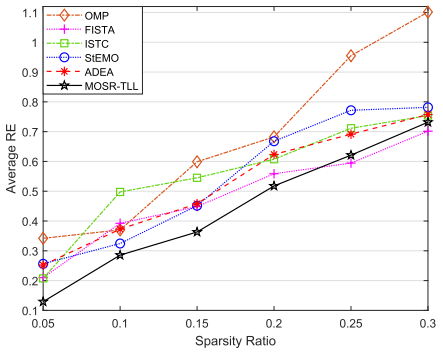


(a)

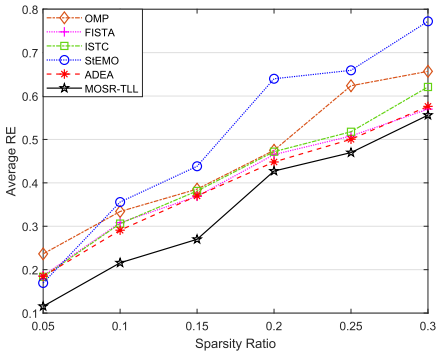


(b)

FIGURE 5. Performance comparisons for simulated signals under different sampling ratios: 0.2 ∼ 0.8. (a) $k/n = 0.05$; (b) $k/n = 0.10$.



(a)



(b)

FIGURE 6. Performance comparison for simulated signals under different sparsity levels: 0.05 ∼ 0.30, (a) $m/n = 0.4$; (b) $m/n = 0.6$.

C. EXPERIMENTAL RESULTS ON BENCHMARK PROBLEMS

This subsection presents the reconstruction performance for the six algorithms on Benchmark problems, i.e., Sparco toolbox [51]. The features of signals in this toolbox are similar to those in practical applications.

Five signals, including sgnspike, gausspike, cosspike, gcosspike and jitter, are generated using the Sparco toolbox. The sgnspike, gausspike and cosspike signals are explicitly sparse, which can be directly reconstructed from undersampling measurements. The gcosspike and jitter signals are implicitly sparse, and they have a sparse representation $\mathbf{x} = \Psi\mathbf{w}$ with respect to the basis matrix Ψ. For all signals, the measurements are corrupted by additive Gaussian noise following the distribution $N(0, 0.01)$. For MOSR-TLL, the source problems are randomly chosen from the source pool generated in Section V-B

The bar graph in Fig. 7 illustrates the average reconstruction performance for the six algorithms. For all signals, the proposed MOSR-TLL achieves the best or comparable reconstruction performance. For sgnspike, gcosspike and jitter, MOSR-TLL achieves the lowest RE. For gausspike, MOSR-TLL and ISTC show comparable results, but perform much better than other algorithms. For cosspike, OMP provides the best reconstruction perofrmance, together with ADEA and MOSR-TLL, while StEMO fails to realize the reconstruction.

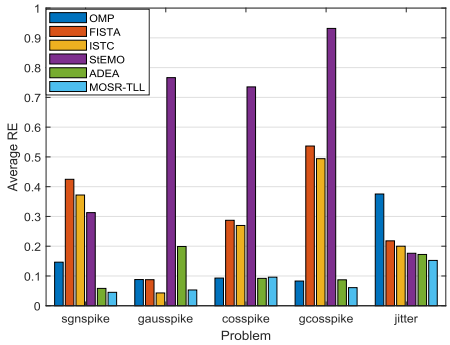


FIGURE 7. Performance comparison for six algorithms on benchmark problems.

D. EXPERIMENTAL RESULTS ON CAMERAMAN IMAGE

This section further evaluates the quality of MOSR-TLL on image reconstruction. The test image is Cameraman with 256×256 pixels. This image is sparse on a Haar wavelet basis and 2-D Haar wavelets are used as the basis matrix Ψ. In this regard, the CS sampling can be rewritten to $\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{A}\Psi\mathbf{w}$, where $\mathbf{w}$ denotes the wavelet coefficients. We can reconstruct $\mathbf{x} = \Psi\mathbf{w}$ by firstly evaluating $\mathbf{w}$ with different solvers. The test image is sampled using a Gaussian random measuring matrix with the sampling ratio 0.7, and the measurement is corrupted by white Gaussian noise with elements following

the normal distribution $N(0, 0.01)$. The mean square error MSE is defined as

$$\text{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - J(i,j)]^2 \tag{21}$$

where m and n are the sizes of rows and columns of the original image respectively, and I and J are the original and reconstructed images. The peak-signal-to-noise ratio (PSNR) is employed to measure the reconstruction performance [54] via MSE: $\text{PSNR(dB)} = 10 \times \log_{10}(255^2/\text{MSE})$. Higher PSNR value indicates better reconstruction quality.

Fig. 8 presents the typical reconstructed images and average PSNR results of the six solvers. It can be observed that MOSR-TLL achieves the highest PSNR and the best visual quality. Images recovered by other algorithms are contaminated severely by noise, while images recovered by MOSR-TLL are much cleaner.

E. EXPERIMENTAL RESULTS ON AVERAGE RUNNING TIME

Although the complexity of the knowledge transfer operator is higher than other operators (refer to Section IV-E), the MOSR-TLL is still computational efficient due to few executions of knowledge transfer. To better illustrate this, we compare the complexity of MOSR-TLL with other algorithms using the running time, using the same settings with those in Fig. 5(a), i.e., $n = 1000$, $m \in [200, 600]$ with the interval of 100, $k = 50$ and $\delta = 0.01$. The stopping criterion is set the same as that in Section V-A.

The average running time for all six algorithms is presented in Table 1. The running time of OMP, FISTA, ISTC and StEMO increases as the sampling rate grows, while ADEA and MOSR-TLL roughly remain unchanged within the range [0.3,0.6]. Among the three single-objective algorithms, FISTA has the fastest running speed. The three MOSR algorithms, which achieve significantly better reconstruction performance than the single-objective algorithms (as shown in Fig. 5 and Fig. 6), run relatively slower due to multiple solution paths. Among the MOSR algorithms, MOSR-TLL is the fastest due to the introduction of transfer learning, with less than ten seconds.

Conventional single-objective sparse reconstruction methods cannot be speeded up via parallel implementations, since only a single solution is searched. For MOSR methods, they search multiple solutions in the evolutionary process, in which the operations of the initialization, crossover, mutation, knowledge transfer (if any), and local search on each solution is able to be accomplished independently without mutual interference. Hence, these operations can be accelerated by parallel implementations [55].

VI. IMPACT STUDY FOR THE INNOVATIONS IN MOSR-TLL

The great performance of the proposed MOSR-TLL scheme, as we have seen in Section V-B, is mainly credited to its innovative operators of knowledge transfer and localized regularization in local search. To demonstrate this clearer, we specifically investigate the impact of these two operators on the convergence performance of MOSR-TLL. The Wilcoxon Sign Test with 95% confidence level is conducted on the obtained results to show the statistical significance of the experimental comparison. We will see that the major performance improvement by knowledge transfer is on improving the reconstruction speed and reconstruction quality, and the localized regularization strategy can improve the reconstruction accuracy notably.

TABLE 1. Average running time (in seconds) of all algorithms on different problems.

method \ m/n	0.2	0.3	0.4	0.5	0.6
OMP	0.2251	0.2303	0.2808	0.3887	0.3468
FISTA	0.1294	0.1394	0.1722	0.2215	0.2837
ISTC	0.2932	0.3204	0.3750	0.4987	0.4623
StEMO	50.40	105.4	162.7	282.2	101.6
ADEA	16.17	16.07	12.12	10.64	8.71
MOSR-TLL	9.142	7.466	7.572	3.658	3.802

A. IMPACT OF SPARSE-CONSTRAINT KNOWLEDGE TRANSFER

The simulated signals with key parameters (n, m, k, δ) are employed to study the impact of knowledge transfer and the gap of signal dimension between the source problem and target problem on the convergence performance of MOSR-TLL. Four complicated test problems, denoted as P1 to P4, are randomly generated, with parameters as specified in Table 2. These problems are heterogeneous, as the dimensions of their $\mathbf{x}$ are different. We compare the performance of MOSR-L and four MOSR-TLL versions with the knowledge transferred from {P1, P2, P3, P4}, in which MOSR-L does not involve knowledge transfer. To exclude the influence of the proposed localized regularization strategy, the regularization values are all fixed at 0.001. Note that even though the source problem and the target one have the same (n, m, k, δ), they are independently generated, that is, the elements of their signals are completely different.

For a fair comparison, all versions stop running when the maximum function evaluations reach 5000 times. To evaluate the convergence performance, RE and hypervolume (HV) [56], which measures the quality of a solution set, are used as the performance metrics. For the HV calculation, the objective function values of solutions are firstly normalized to [0, 1] and (1.1, 1.1) is set as the reference point [57]. Larger HV and lower RE values indicate better convergence performance.

TABLE 2. A list of the test problems.

Problem	(n, m, k, δ)	Problem	(n, m, k, δ)
P1	(500, 100, 25, 0.01)	P3	(2000, 400, 100, 0.01)
P2	(1000, 200, 50, 0.01)	P4	(5000, 1000, 250, 0.01)

(a) Original image (b) Partial enlarged view (c) OMP: 28.3475dB (d) FISTA: 28.6416dB

(e) ISTC: 27.6892dB (f) StEMO: 29.4210dB (g) ADEA: 30.0746dB (h) MOSR-TLL: 31.5224dB

FIGURE 8. Comparison on image reconstruction of Cameraman image.

TABLE 3. Average REs and IQR values for MOSR-TLLs with 5000 function evaluations.

Target \ Source	null	P1	P2	P3	P4
P1	0.2021 (4.28E-3)	0.1791 (3.87E-3) +	0.1773 (3.99E-3) +	0.1428 (3.11E-3) +	0.1535 (2.08E-3) +
P2	0.2364 (4.92E-3)	0.1588 (6.10E-3) +	0.1878 (3.09E-3) +	0.1818 (4.05E-3) +	0.1645 (2.56E-3) +
P3	0.3688 (5.69E-2)	0.3996 (3.28E-2) −	0.3241 (2.58E-2) +	0.3407 (2.03E-2) +	0.3502 (3.64E-2) +
P4	0.4494 (2.03E-2)	0.4658 (1.48E-2) −	0.4150 (8.12E-3) +	0.3729 (8.73E-3) +	0.4078 (6.53E-3) +
Mean	0.3142 (2.16E-2)	0.3008 (1.44E-2) +	0.2761 (8.12E-3) +	0.2600 (9.05E-3) +	0.2690 (1.19E-2) +

The average REs and IQR values of MOSR-TLLs for solving P1 to P4 with 5000 evaluations are presented in Table 3. In the table, symbols "+" and "−" represent that the reconstruction quality of the solver with knowledge transfer is better and worse than the corresponding "null" version, respectively. As can be observed from Table 3, the MOSR-TLL with knowledge transfer achieve lower RE with respect to MOSR-L in 14 out of 16 scenarios. The two failure cases may be due to the too large dimension gap between the source and target problem. The mean RE values of MOSR-TLLs with the knowledge transferred from different problems are also smaller than that without knowldege transfer.

Figure 9 provides the corresponding mean HV values for solving P1 to P4 by MOSR-L, MOSR-TLL+P1, MOSR-TLL+P2, MOSR-TLL+P3, and MOSR-TLL+P4, where the term after "+" stands for the source problem. Except for solving P1, the MOSR-TLL versions with knowledge transfer show obvious faster reconstruction speed compared to the "null" version. Specifically, for P2, the four MOSR-TLLs with knowledge transfer save at least 1000 function evaluations, in which the "P1→P2" version converges slightly slower than other knowledge-transfer versions. For P3 and P4, the four knowledge-transfer versions of MOSR-TLL achieves remarkable HV improvements. For P1, although the MOSR-TLLs with knowledge transfer have very similar reconstruction speed compared to that without knowledge transfer, they achieve better reconstruction quality (as shown in Table 3).

To summarize, no matter whether the source problem is homogeneous or heterogeneous with respect to the target problem, the MOSR-TLL scheme can obtain better reconstruction speed and quality with great probability.

B. IMPACT OF LOCALIZED REGULARIZATION

We investigate the impact of the localized regularization strategy on reconstruction performance via comparing local search with fixed regularization values (i.e., 0.0001, 0.001, 0.01, 0.1), random values, and continuation-based parameters. Note that the sparse-constraint knowledge transfer is removed for clarity. For random inputs, as suggested in [10], the elements in $\boldsymbol{\lambda}$ are randomly chosen between 0 and $\|A^T b\|_\infty$, in which the maximum value to be $\|A^T b\|_\infty$ can

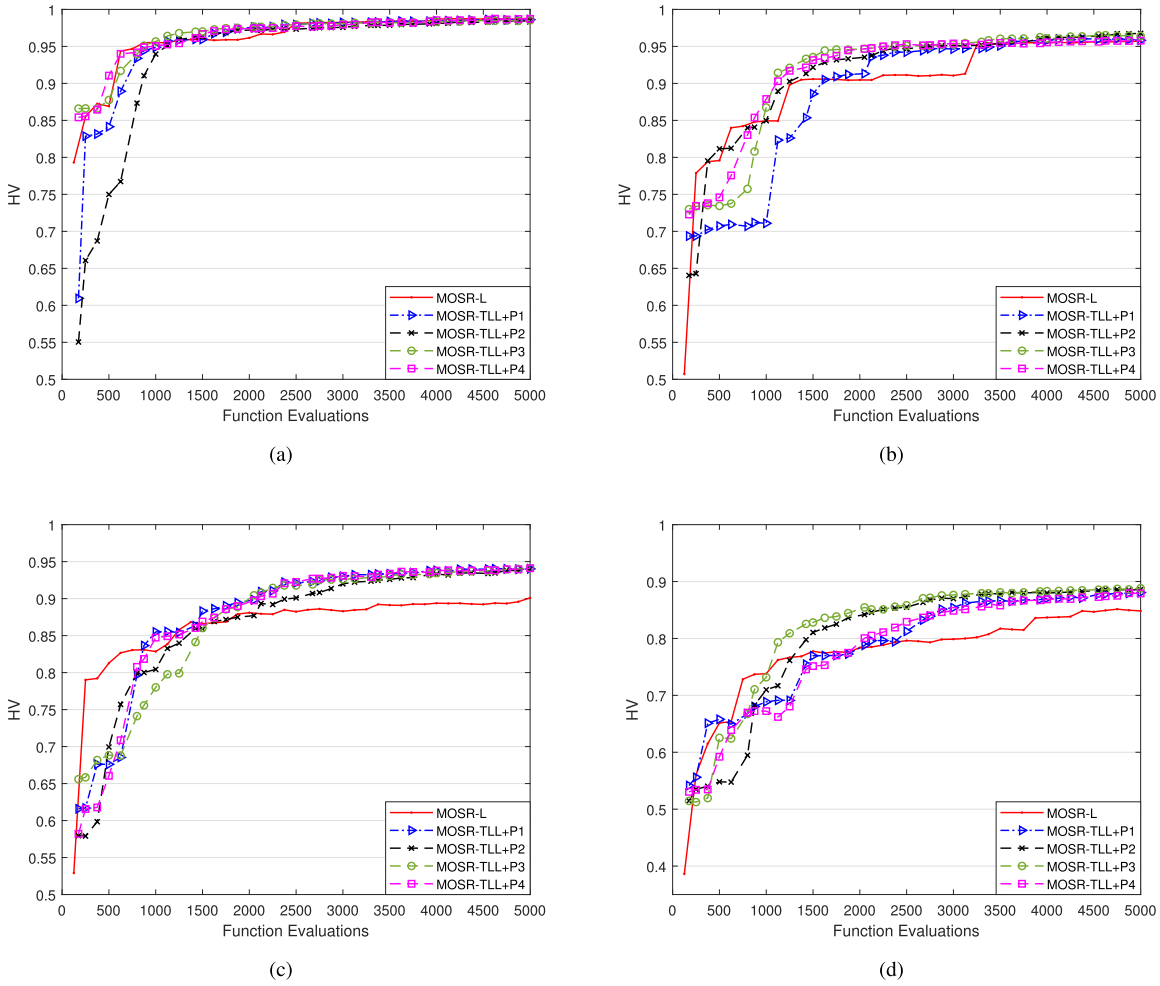


FIGURE 9. The median HV values by MOSR-L and MOSR-TLLs with 5000 function evaluations, where the term after "+" stands for the source problem. (a)∼(d) are the target problems to be solved: (a) P1. (b) P2. (c) P3. (d) P4.

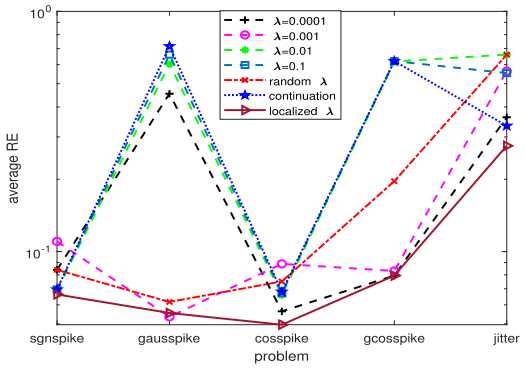


FIGURE 10. Average REs for five benchmarks under different regularization parameters.

avoid yielding zero solutions. The continuation-based input employs [58] as the reference, where the maximum value is 0.1, and the decaying rate is 0.03. All the methods stop when the number of iterations reaches 100.

Fig.10 provides the comparison results over different regularization values, for which the test problems are the aforementioned five Benchmark problems with Gaussian noise satisfying the distribution $N(0, 0.02)$. It can be found that our localized regularization achieves the best or comparable recovery performance for all the problems. Nonetheless, there is a huge gap among the RE results of different fixed $\boldsymbol{\lambda}$. Therefore, it is clear that the localized regularization strategy makes important contributions to performance improvement.

VII. CONCLUSION

We presented a MOSR scheme with transfer learning and localized regularization (MOSR-TLL) that can accelerate the reconstruction speed and improve the reconstruction accuracy. We developed a knowledge transfer operator that reuses the searching experience from a previously solved problem. This operator is shown to significantly accelerate the reconstruction speed, with improvements in reconstruction accuracy with great probability. By dividing a problem to many subproblems, we also developed a localization regularization strategy for local search to produce good solutions stably. By assigning carefully designed independent regularization value to each subproblem, the reconstruction accuracy is

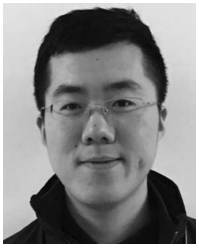
shown to be notably improved. The two techniques used in our MOSR-TLL scheme are very promising for various practical applications, such as hyperspectral image unmixing. Our future work will focus on more computational-efficient knowledge transfer strategies with smaller memory for multi-objective sparse reconstruction.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
[4] T. Çukur, M. Lustig, and D. G. Nishimura, "Improving non-contrast-enhanced steady-state free precession angiography with compressed sensing," *Magn. Reson. Med.*, vol. 61, no. 5, pp. 1122–1131, May 2009.
[5] T. Cukur, M. Lustig, E. U. Saritas, and D. G. Nishimura, "Signal compensation and compressed sensing for magnetization-prepared MR angiography," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1017–1027, May 2011.
[6] D. Zhang, Y. Zhou, H. Chen, W. Chen, and Y. Chen, "Hybrid rank-sparsity constraint model for simultaneous reconstruction and denoising of 3D seismic data," *Geophysics*, vol. 82, no. 5, pp. 351–367, 2017.
[7] Y. Chen, M. Bai, and Y. Chen, "Obtaining free USArray data by multi-dimensional seismic reconstruction," *Nature Commun.*, vol. 10, no. 1, pp. 1–13, Dec. 2019.
[8] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Jan. 2005.
[9] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.
[10] B. Yan, Q. Zhao, Z. Wang, and J. A. Zhang, "Adaptive decomposition-based evolutionary approach for multiobjective sparse reconstruction," *Inf. Sci.*, vol. 462, pp. 141–159, Sep. 2018.
[11] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.
[12] M. Shahdloo, E. Ilicak, M. Tofighi, E. U. Saritas, A. E. Cetin, and T. Cukur, "Projection onto epigraph sets for rapid self-tuning compressed sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1677–1689, Jul. 2019.
[13] Y. Jiao, B. Jin, and X. Lu, "Iterative soft/hard thresholding with homotopy continuation for sparse recovery," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 784–788, Jun. 2017.
[14] C. G. Sentelle, G. C. Anagnostopoulos, and M. Georgiopoulos, "A simple method for solving the SVM regularization path for semidefinite kernels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 709–722, Apr. 2016.
[15] L. Li, X. Yao, R. Stolkin, M. Gong, and S. He, "An evolutionary multiobjective approach to sparse reconstruction," *IEEE Trans. Evol. Comput.*, vol. 18, no. 6, pp. 827–845, Dec. 2014.
[16] Y. Zhou, S. Kwong, H. Guo, X. Zhang, and Q. Zhang, "A two-phase evolutionary approach for compressive sensing reconstruction," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2651–2663, Sep. 2017.
[17] H. Li, Q. Zhang, J. Deng, and Z.-B. Xu, "A preference-based multiobjective evolutionary approach for sparse optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1716–1731, May 2018.
[18] F. Liu, L. Lin, L. Jiao, L. Li, S. Yang, B. Hou, H. Ma, L. Yang, and J. Xu, "Nonconvex compressed sensing by nature-inspired optimization algorithms," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1042–1053, May 2015.
[19] 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.
[20] T. Bulmensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmon. A.*, vol. 27, no. 3, pp. 265–274, 2009.
[21] H. Li, Y. Fan, Q. Zhang, Z. Xu, and J. Deng, "A multi-phase multiobjective approach based on decomposition for sparse reconstruction," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 601–608.
[22] P. Wei, Y. Ke, and C. K. Goh, "Deep nonlinear feature coding for unsupervised domain adaptation," in *Proc. IJCAI*, 2016, pp. 2189–2195.
[23] 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.
[24] B. Yan, Q. Zhao, Z. Wang, and X. Zhao, "A hybrid evolutionary algorithm for multiobjective sparse reconstruction," *Signal Image Video Process.*, vol. 11, no. 6, pp. 993–1000, Sep. 2017.
[25] C. Yue, J. Liang, B. Qu, H. Song, G. Li, and Y. Han, "A knee point driven particle swarm optimization algorithm for sparse reconstruction," in *Proc. Asia–Pacific Conf. Simulated Evol. Learn.* Cham, Switzerland: Springer, 2017, pp. 911–919.
[26] C. Yue, J. Liang, B. Qu, Y. Han, Y. Zhu, and O. D. Crisalle, "A novel multiobjective optimization algorithm for sparse signal reconstruction," *Signal Process.*, vol. 167, Feb. 2020, Art. no. 107292.
[27] 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.
[28] 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.
[29] 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.
[30] M. Iqbal, B. Xue, H. Al-Sahaf, and M. Zhang, "Cross-domain reuse of extracted knowledge in genetic programming for image classification," *IEEE Trans. Evol. Comput.*, vol. 21, no. 4, pp. 569–587, Aug. 2017.
[31] 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.
[32] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
[33] L. Feng, L. Zhou, J. Zhong, A. Gupta, Y.-S. Ong, K.-C. Tan, and A. K. Qin, "Evolutionary multitasking via explicit autoencoding," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3457–3470, Sep. 2019.
[34] 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.
[35] 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.
[36] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[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] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A reference vector guided evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 773–791, Oct. 2016.
[39] C. Liu, Q. Zhao, B. Yan, S. Elsayed, T. Ray, and R. Sarker, "Adaptive sorting-based evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 23, no. 2, pp. 247–257, Apr. 2019.
[40] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM J. Optim.*, vol. 8, no. 3, pp. 631–657, Aug. 1998.
[41] J. A. Cornell, *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, vol. 403. Hoboken, NJ, USA: Wiley, 2011.
[42] 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.
[43] K. Deb and K. Miettinen, "A review of nadir point estimation procedures using evolutionary approaches: A tale of dimensionality reduction," in *Proc. MCDM*. Cham, Switzerland: Springer, 2009, pp. 1–14.
[44] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.
[45] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 115–148, 1995.
[46] K. Deb and M. Goyal, "A combined genetic adaptive search (GeneAS) for engineering design," *Comput. Sci. Inf.*, vol. 26, pp. 30–45, Aug. 1996.
[47] I. Mierswa and M. Wurst, "Information preserving multi-objective feature selection for unsupervised learning," in *Proc. 8th Annu. Conf. Genetic Evol. Comput. (GECCO)*, 2006, pp. 1545–1552.

[48] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Proc. Conf. Neural Inf. Process. Syst.*, 2007, pp. 513–520.
[49] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," in *Proc. 29th Int. Conf. Mach. Learn.* Edinburgh, Scotland, Jun. 2012, pp. 1–8.
[50] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
[51] E. van den Berg, M. P. Friedlander, G. Hennenfent, F. J. Herrmann, R. Saab, and Ö. Yilmaz, "Algorithm 890: Sparco: A testing framework for sparse reconstruction," *ACM Trans. Math. Softw.*, vol. 35, no. 4, pp. 1–16, Feb. 2009.
[52] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
[53] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
[54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[55] R. Xie and X. Jia, "Transmission-efficient clustering method for wireless sensor networks using compressive sensing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 3, pp. 806–815, Mar. 2014.
[56] J. Bader and E. Zitzler, "HypE: An algorithm for fast hypervolume-based many-objective optimization," *Evol. Comput.*, vol. 19, no. 1, pp. 45–76, Mar. 2011.
[57] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, "Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 169–190, Apr. 2017.
[58] L. Wu, Z. Sun, and D.-H. Li, "A Barzilai–Borwein-like iterative half thresholding algorithm for the $L_{1/2}$ regularized problem," *J. Sci. Comput.*, vol. 67, no. 2, pp. 581–601, May 2016.

BAI YAN received the Ph.D. degree from the Beijing University of Technology, Beijing, China, in 2019. From 2017 to 2018, she was a Visiting Scholar with the School of Computing and Communications, University of Technology Sydney, Ultimo, NSW, Australia. She is currently a Postdoctoral Fellow with the Department of Computer Science and Engineering, Southern University of Science and Technology, Guangdong, China. Her research interests include evolutionary computation, multi-objective optimization, and their applications.



QI ZHAO received the Ph.D. degree in management science and engineering from the Beijing University of Technology, China, in 2019. From 2017 to 2018, he was a Visiting Scholar with the School of Engineering and Information Technology, University of New South Wales, Australia. Since 2019, he has been a Postdoctoral Fellow with the Department of Computer Science and Engineering, Southern University of Science and Technology, China. His research interests include computational intelligence, evolutionary optimization, and computational finance.

J. ANDREW ZHANG (Senior Member, IEEE) received the B.Sc. degree from Xi'an Jiaotong University, Xi'an, China, in 1996, the M.Sc. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1999, and the Ph.D. degree from Australian National University, Canberra, ACT, Australia, in 2004. He was a Researcher with ZTE Corporation, Nanjing, from 1999 to 2001, the Networked Systems, NICTA, Australia, from 2004 to 2010, and the Data61, CSIRO, Australia, from 2010 to 2016. He is currently an Associate Professor with the School of Computing and Communications, University of Technology Sydney, Ultimo, NSW, Australia. His research interests include signal processing for wireless communications and sensing, and autonomous vehicular networks. He has authored more than 120 papers in leading international journals and conference proceedings, and has won four Best Paper Awards for his work. He was a recipient of the CSIRO Chairman Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multigigabit wireless communications.

ZHIHAI WANG received the B.Sc. and M.Sc. degrees from Beijing University, Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree from Princeton University, NJ, USA, in 2007. His is currently a Professor with the Key Laboratory of Optoelectronics Technology, Beijing University of Technology, Beijing. His research interests include 3D printing, optofluidics, compressive sensing, and its applications.