

Elsevier required licence: © 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Adaptive Decomposition-based Evolutionary Approach for Multiobjective Sparse Reconstruction

Bai Yan^{a,*}, Qi Zhao^{b,*}, Zhihai Wang^c, J. Andrew Zhang^d

^a*Institute of Laser Engineering, Beijing University of Technology, Beijing 100124, China*

^b*College of Economics and Management, Beijing University of Technology, Beijing 100124, China*

^c*Key Laboratory of Optoelectronics Technology, Ministry of Education, Beijing University of Technology, Beijing 100124, China*

^d*Global Big Data Technologies Centre (GBDTC), University of Technology Sydney, NSW 2007, Australia*

Abstract

This paper aims at solving the sparse reconstruction (SR) problem via a multiobjective evolutionary algorithm. Existing multiobjective evolutionary algorithms for the SR problem have high computational complexity, especially in **scenarios of** high-dimensional reconstruction. Furthermore, these algorithms focus on estimating the whole Pareto front rather than the knee region, thus leading to **solutions with** limited diversity in **the** knee region and **causing a** waste of computational effort. To tackle these issues, this paper proposes an adaptive decomposition-based evolutionary approach (ADEA) for the SR problem. Firstly, we employ the decomposition-based evolutionary paradigm to guarantee a high computational efficiency and **the** diversity of solutions in the whole objective space. Then, we propose a two-stage iterative soft-thresholding (IST)-based local search operator **for improving** the convergence. Finally, we develop an adaptive decomposition-based environmental selection strategy, by which the decomposition in the knee region can be adjusted dynamically. This strategy **makes it possible** to focus the selection effort on the knee region, **hence involving** low computational complexity. Experimental results on **images, and simulated and** benchmark signals demonstrate the superiority of ADEA in terms of reconstruction accuracy and computational efficiency, compared to five state-of-the-art algorithms.

Keywords: sparse reconstruction, multiobjective evolutionary algorithm, adaptive decomposition, reference vector

*Corresponding author

Email addresses: yanbai@emails.bjut.edu.cn (Bai Yan), qzhao@emails.bjut.edu.cn (Qi Zhao), wangzhihai@bjut.edu.cn (Zhihai Wang), Andrew.Zhang@uts.edu.au (J. Andrew Zhang)

1. Introduction

The sparse reconstruction (SR) problem widely exists in the under-determined system of linear equations [19, 8], especially in the field of signal/image processing. There have been many successful applications in this field, such as action recognition [50], image super-resolution [21] and image classification [11], ~~human detection and background subtraction~~. In the SR problem, we want to recover the unknown sparse signal \mathbf{x} from the measurement \mathbf{b} in the following under-determined system

$$\mathbf{b} = \mathbf{A}\mathbf{x} \quad (1)$$

where \mathbf{x} is an unknown sparse vector with k nonzero elements ($\mathbf{x} \in \mathbb{R}^n, k \ll n$), \mathbf{A} is a full-rank sensing matrix ($\mathbf{A} \in \mathbb{R}^{m \times n}, m < n$) and should satisfy the restricted isometry property (RIP) [8], and \mathbf{b} is the measurement vector. Sometimes the signal \mathbf{x} is not sparse but has a sparse representation \mathbf{w} with respect to some bases Ψ , such as the Fourier bases and the wavelet bases. In this case, equation (1) can be transformed to $\mathbf{b} = \mathbf{A}\Psi\mathbf{w}$, and we only need to reconstruct \mathbf{w} before solving \mathbf{x} by $\mathbf{x} = \Psi\mathbf{w}$. In this paper, matrices and vectors are denoted by boldface upper-case and lower-case letters respectively.

For simplicity, here we focus on the case when the signal \mathbf{x} itself is sparse. The compressed sensing theory [9] can be employed to reconstruct \mathbf{x} by solving the following SR problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \quad (2)$$

where $\|\mathbf{x}\|_0$ represents the number of nonzero elements in \mathbf{x} , and $\|\cdot\|$ denotes the standard Euclidean norm for a vector. If noise is included in the measurements, this problem is updated to

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ s.t. } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \leq \sigma \quad (3)$$

where $\sigma > 0$ is a given constant related to the noise.

The problem in (2) is known to be NP-hard [34]. Quite a few algorithms have been developed to solve this l_0 -norm SR problem, such as orthogonal matching pursuit (OMP) [37, 7], compressive sampling matching pursuit [35] and iterative hard thresholding methods [41, 5]. They perform well only when the measurement samples are much more than nonzero elements in \mathbf{x} . The l_0 -norm problem can also be transformed to a convex optimization problem [10], or relaxed by the l_1 -norm [18, 42, 4, 39] or $l_p(0 < p < 1)$ -norm [31, 48, 43, 45]. The l_1 -norm algorithms are more robust to noise and can recover signals with better reconstruction quality. But in some cases (e.g., the matrix \mathbf{A} does not satisfy the low coherence conditions), they are unable to guarantee the equivalence between l_1 -norm and l_0 -norm. The $l_p(0 < p < 1)$ -norm is nonconvex and nonsmooth. Its convergence is yet to be proven theoretically, and it is very challenging to derive fast and efficient solutions for $l_p(0 < p < 1)$ -norm problems. For these relaxation algorithms, the problem in (2) is commonly transformed to the following

continuous optimization problem

$$\mathbf{x} = F(\mathbf{x}) = \arg \min_{\mathbf{x}} \lambda \|\mathbf{x}\|_p + \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \quad (4)$$

where $p \in [0, 1]$ and λ denotes the regularization parameter. There exists the problem of determining λ to balance the objective term and the penalty function term $\|\mathbf{x}\|_p + \|\mathbf{x}\|_p$, since λ is closely related to the reconstruction quality. However, there is no exact method for finding the optimal λ value in practical applications.

All of the algorithms mentioned above are single objective, and they solve the combined objective function in an independent way, where the solution path is fixed. To exploit joint optimization and provide adaptability to the solution path, a new approach is proposed to transform Problem (4) to a multiobjective sparse reconstruction (MOSR) problem:

$$f(\mathbf{x}) = \min_{\mathbf{x}} (\|\mathbf{x}\|_0, \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2) \quad (5)$$

15 where $\|\mathbf{x}\|_0$ and $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ represent the sparsity and the measurement errors, respectively.

Multiobjective evolutionary algorithms (MOEAs), such as NSGA-II [16], differential evolution [36] and MOEA/D [49], are widely used to tackle optimization problems with two or more objectives. MOEAs optimize all of the objectives simultaneously, and are capable of providing a variety of trade-off solutions (termed as Pareto front, PF) among the objectives. Recently, MOEAs, such as the soft-thresholding evolutionary multiobjective (StEMO) algorithm [30], are applied to solve the MOSR problem. StEMO is based on the NSGA-II framework and is combined with the soft-thresholding algorithm [24] for local search to further improve the convergence performance. The knee region on the final PF is proved to provide the best trade-off, because it has the largest marginal rates of return [30]. It is selected as the final solution and can be identified by the angle-based method [6]. Another two algorithms are able to find a local part of Pareto front near the knee region, accounting for preference. The first is an improved MOEA/D with L1/2 solver (MOEA/D-L1/2) [28], and the second is sparse preference based local search (SPLS) [29]. Both algorithms integrate iterative threshold algorithms into MOEAs. They first use a single starting solution from chain search results and weakly Pareto front respectively for local search, and then execute multiple truncations to update the solution set and increase the diversity of population. In [44], the LBEA algorithm is developed by embedding a linear Bregman-based [25] local search operator into the differential evolution paradigm. An adaptive strategy is designed for the linear Bregman-based local search, where the number of individuals and iterations for local search is set to be adaptive for accelerating convergence.

40 These algorithms demonstrate the advantages of MOEA in solving the MOSR problem. However, they cannot provide fast reconstruction due to the use of the Pareto nondominance principle. This can be verified by a simple experiment. Consider a sparse signal with length ~~$N = 1000$~~ $n = 1000$ and sparsity

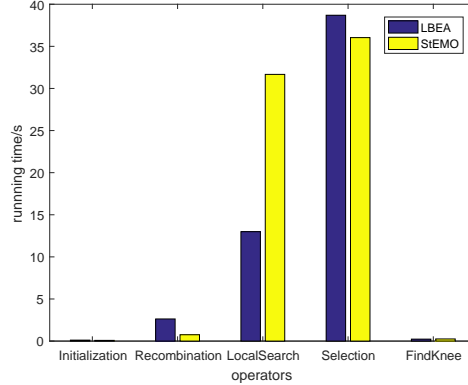


Figure 1: Comparison for the running time of each component in StEMO and LBEA

ratio $\frac{k}{n} = 0.05$. The nonzero elements are randomly chosen from the Gaussian distribution. The sensing matrix is a Gaussian random matrix with the dimension 400×1000 . The population size and the maximum number of generation are both set to 100. Use StEMO and LBEA as examples. The average running time is shown in Fig.1. We can see that the Pareto dominance-based selection operators have larger running time in both algorithms.

On the other hand, the reconstruction quality of these algorithms is limited. For StEMO and LBEA, they put search effort uniformly over the whole PF. However, the knee region has the solutions with the maximal marginal rates of return, hence requiring more search effort. Even if in cases when the knee region does not provide the best approximation for the ground-truth data, the solutions in this region are still Pareto optimal [30]. Therefore, the computational effort of StEMO and LBEA is likely to be wasted in estimating the area far from the knee region, and the diversity of final solutions in the knee region will be poor due to the lack of sufficient solutions there. We implement an experiment with the same setting as that in Fig. 1 to illustrate this issue, and the results are presented in Fig. 2. From Fig. 2, the distribution of solutions around the knee region is not uniform, and the curve generated from those solutions is not sufficiently smooth, potentially leading to the wrong estimation of knee point. For MOEA/D-L1/2 and SPLS, they implement local search on one selected solution, treating the solution obtained from local search as the potential knee point, and maintaining the diversity around this knee point by multiple truncations. However, they only select one solution from the local search, and it is possible that this single solution may not be located in the true knee region. This can misguide the evolutionary process. Furthermore, it is difficult to find the optimal inner parameters of MOEA/D-L1/2 and SPLS for various problems.

To address the above problems, this paper proposes an adaptive decomposition based evolutionary approach (ADEA). In ADEA, we employ predefined reference vectors to partition the whole population of the objective space into

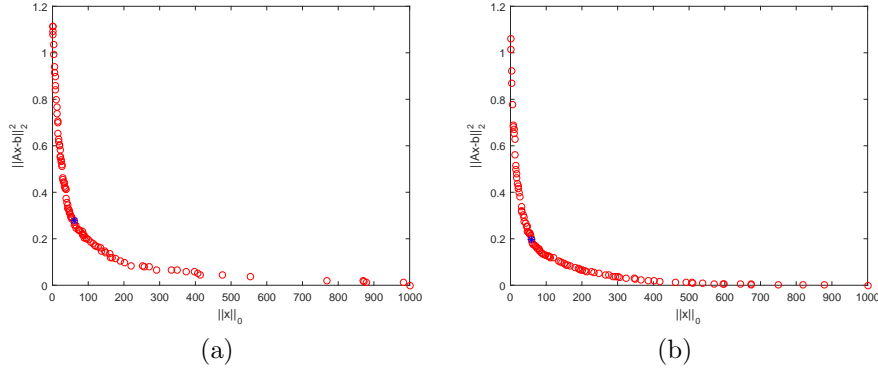


Figure 2: Final solutions obtained by (a) StEMO and (b) LBEA. The points in blue are the actual knee points.

a number of subpopulations, and these subpopulations will converge along the reference vectors toward the true PF. Compared to dominance-based MOEAs, this approach has demonstrated its efficiency in reducing computational cost and improving solution quality on multiobjective optimization problems, e.g., [12, 47]. Then we introduce an improved two-stage IST-based strategy for local search. This strategy significantly improves the convergence and diversity of the solutions. We further propose an environmental selection strategy based on adaptive decomposition. This strategy devotes additional search effort to the potential knee region in each generation to improve the diversity of the knee region. At the same time, the search effort on the whole PF and hence the overall diversity of solutions in the whole objective space are still maintained. Our major contributions in this paper are summarized as follows:

- We introduce a decomposition-based evolutionary paradigm for the MOSR problem. In this scheme, the population is decomposed into a number of subpopulations by uniformly distributed reference vectors in the objective space. This evolutionary scheme can maintain a good diversity of solutions and direct search results towards the true PF. At the same time, the computational efficiency is substantially better compared to existing MOEAs for the MOSR problem.
- We propose an improved two-stage IST-based local search operator. The population is divided into two sets, one containing the best solutions for each subpopulation and the other containing the rest. The first stage is convergence-oriented, i.e., selecting two solutions from both sets for local search to push the solutions to converge to the true PF. The second stage is diversity-oriented, updating solutions by the first set to spread them along the PF. In addition, an extra constraint is added to the λ vector to exclude zero solutions. This local search strategy improves the exploitation ability of ADEA.

- We propose an adaptive decomposition-based environmental selection strategy. In each generation, this strategy identifies the potential knee region, then adds additional reference vectors in this region. Following this procedure, more promising solutions in the potential knee region will be selected, which improves the diversity of solutions in the knee region and finally enhances the quality of the overall solution.
- We provide extensive experimental results for the proposed algorithm. We compare the proposed algorithm with six state-of-the-art SR algorithms using simulated signals, benchmark signals and images. The results demonstrate the superiority of our proposed ADEA scheme.

The rest of the paper is organized as follows. Section II presents the background technology. Section III introduces the framework and detailed components of the proposed ADEA scheme. Experimental results are provided in Section IV. Finally, Section V concludes the paper.

2. Background and Basis

In this section, we briefly introduce the ~~basis~~ basics of the proposed algorithm, including the decomposition, population normalization, individual association and the iterative soft-thresholding (IST) algorithm.

2.1. Decomposition

In the decomposition-based evolutionary paradigm, predefined reference vectors partition the population into several subpopulations. The reference vectors assist in selecting the elitism solutions from each subpopulation, thus guaranteeing the convergence and diversity simultaneously. To generate the reference vectors uniformly in the two-dimensional objective space, the normal-boundary intersection method [13] is adopted. Firstly, H uniformly distributed reference points on the hyperball are generated, so that the hyperball will be divided into $(H - 1)$ parts. Then, the unit vectors from the coordinate origin to each reference point are used as reference vectors.

For the MOSR problem, an example with 100 reference vectors in the two-dimensional objective space is shown in Fig. 3. The reference points and reference vectors are marked in blue and red respectively. We can see that the hyperball is partitioned by 100 evenly distributed reference points, and then the 100 unit vectors from the origin to each reference point form reference vectors in the objective space.

2.2. Population Normalization

Since the two objective functions are disparately scaled and all the reference vectors always start from the coordinate origin, the adaptive normalization method [15] [46] is used. Let the M -th ($M = 1, 2$) objective function of the

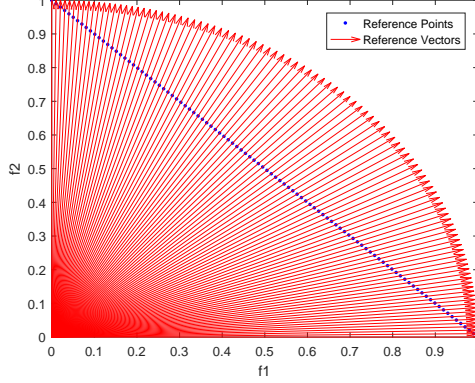


Figure 3: An example with 100 reference vectors in the two-dimensional objective space.

individual \mathbf{x}_i be $\mathbf{f}_M(\mathbf{x}_i)$. It can be normalized to $\tilde{\mathbf{f}}_M(\mathbf{x}_i)$:

$$\tilde{\mathbf{f}}_M(\mathbf{x}_i) = \frac{\mathbf{f}_M(\mathbf{x}_i) - \mathbf{z}_M^*}{\mathbf{z}_M^{\text{nad}} - \mathbf{z}_M^*} \quad (6)$$

where \mathbf{z}_M^* is the M -th dimension of the ideal point (the ideal point is determined by the minimum value of each objective function) and $\mathbf{z}_M^{\text{nad}}$ is the M -th dimension of the nadir point obtained from the current generation. The approach of evaluating $\mathbf{z}_M^{\text{nad}}$ is given in [15]. Firstly, the extreme points of the two objective axes are calculated. Then a hyper ball is determined based on the extreme points. Its intercepts on each objective axis are calculated. Finally, $\mathbf{z}_M^{\text{nad}}$ is updated based on the intercepts.

2.3. Individual Association

After population normalization, each solution will be associated with its closest reference vector, so that the population can be partitioned into several subpopulations. Specifically, for a solution $\mathbf{x}_i \in \mathbf{P}$, the acute angles between $\mathbf{f}(\mathbf{x}_i)$ and each reference vector $\mathbf{v}_j (j = 1, 2, \dots, N)$ are first calculated:

$$\theta_{\mathbf{x}_i, \mathbf{v}_j} = \arccos\left(\frac{\mathbf{f}(\mathbf{x}_i) \cdot \mathbf{v}_j}{\|\mathbf{f}(\mathbf{x}_i)\|}\right). \quad (7)$$

Then, \mathbf{x}_i is associated with \mathbf{v}_j if and only if the angle between them is minimal. As a result, the solutions associated with \mathbf{v}_j constitute the subpopulation \mathbf{P}_j :

$$\mathbf{P}_j = \{\mathbf{x}_i | j = \arg \min \theta_{\mathbf{x}_i, \mathbf{v}_j}\}. \quad (8)$$

An example is given in Fig. 4, where \mathbf{v}_1 and \mathbf{v}_2 are two reference vectors, $\tilde{\mathbf{f}}(\mathbf{x})$ is the normalized objective function of individual \mathbf{x} , and \mathbf{x} is associated with \mathbf{v}_1 as $\theta_1 < \theta_2$.

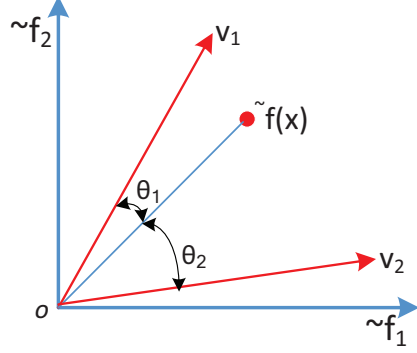


Figure 4: An example of individual association

2.4. Iterative Soft-thresholding Algorithm

Iterative soft-thresholding (IST) algorithm [24] is a SR algorithm which solves Problem (4). It updates solutions based on the approximation

$$\begin{aligned} \mathbf{x}^{(l+1)} &\approx \arg \min_{\mathbf{x}} (\mathbf{x} - \mathbf{x}^{(l)})^T \nabla \mathbf{F}(\mathbf{x}^{(l)}) + \frac{\alpha^{(l)}}{2} \|\mathbf{x} - \mathbf{x}^{(l)}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ &= \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{u}^{(l)}\|_2^2 + \frac{\lambda}{\alpha^{(l)}} \|\mathbf{x}\|_1, \end{aligned} \quad (9)$$

where $\mathbf{u}^{(l)} = \mathbf{x}^{(l)} - \frac{1}{\alpha^{(l)}} \nabla \mathbf{F}(\mathbf{x}^{(l)})$, and $\nabla \mathbf{F}(\mathbf{x}^{(l)}) = \mathbf{A}^T (\mathbf{A} \mathbf{x}^{(l)} - \mathbf{b})$ is the negative gradient vector. The IST algorithm includes three major steps: adaptive step-size, gradient descent, and thresholding truncation. The procedure is given in Algorithm 1. There are two main methods for determining λ in the IST algorithm. The first method uses a fixed value from experience. The other is known as the ‘‘Continuation method’’ [22], where λ decreases with increasing iterations. The stepsize parameter $\alpha^{(l)}$ is determined by the Barzilai-Borwein method, [3], and is used to accelerate convergence. The term αI is approximated by the hessian $\nabla^2 \mathbf{F}(\mathbf{x})$. Let $\mathbf{s}^{(l)} = \mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}$ and $\mathbf{r}^{(l)} = \nabla \mathbf{F}(\mathbf{x}^{(l)}) - \nabla \mathbf{F}(\mathbf{x}^{(l-1)})$. Then $\alpha^{(l)}$ can be obtained in the least-squares sense, i.e.,

$$\alpha^{(l)} = \arg \min_{\alpha} \|\alpha \mathbf{s}^{(l)} - \mathbf{r}^{(l)}\|_2^2 = \frac{(\mathbf{s}^{(l)})^T \mathbf{r}^{(l)}}{(\mathbf{s}^{(l)})^T \mathbf{s}^{(l)}} \quad (10)$$

Algorithm 1: The IST Algorithm

Input: $\mathbf{x}^{(l)}, \mathbf{x}^{(l-1)}, \lambda, \mathbf{A}, \mathbf{b}$ **Output:** $\mathbf{x}^{(l+1)}$

```
1 /*Adaptive Stepsize*/
2  $\mathbf{s}^{(l)} = \mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}$ ;
3  $\mathbf{r}^{(l)} = \nabla \mathbf{F}(\mathbf{x}^{(l)}) - \nabla \mathbf{F}(\mathbf{x}^{(l-1)})$ ;
4  $\alpha^{(l)} = \frac{(\mathbf{s}^{(l)})^T \mathbf{r}^{(l)}}{(\mathbf{s}^{(l)})^T \mathbf{s}^{(l)}}$ ;
5 /*Gradient Descent*/
6  $\mathbf{u}^{(l)} = \mathbf{x}^{(l)} - \frac{1}{\alpha^{(l)}} \nabla \mathbf{F}(\mathbf{x}^{(l)})$ ;
7 /*Thresholding Truncation*/
8
```

$$\mathbf{x}^{(l+1)} = \begin{cases} \text{sign}(\mathbf{u}^{(l)}) \left(|\mathbf{u}^{(l)}| - \frac{\lambda}{\alpha^{(l)}} \right) & \text{if } \mathbf{u}^{(l)} > \frac{\lambda}{\alpha^{(l)}} \\ 0 & \text{otherwise.} \end{cases}$$

3. The Proposed ADEA Scheme

3.1. Motivations and New Features

150 Our proposed ADEA scheme is mainly motivated by the following observations and facts, and addresses these associated problems.

155 Firstly, existing MOSR algorithms employ the principle of Pareto nondominance, which has a limited computational speed. To address this issue, we apply a decomposition-based evolutionary paradigm in our proposed ADEA scheme, which can significantly reduce the computational cost of the selection process. Furthermore, as the population is decomposed evenly into subpopulations by reference vectors, faster convergence and higher diversity can be achieved in ADEA.

160 Secondly, it is proved that nearly optimal solutions exist in the knee region of the objective space [30]. However, StEMO and LBEA uniformly allocate selection effort over the whole objective space and only consider the knee region in the last generation. As a result, only a limited number of solutions can be obtained from the knee region, which leads to solutions with insufficient diversity there. MOEA/D-L1/2 and SPLS focus on maintaining the diversity 165 in the knee region by truncating a single solution generated by the local search. This may have low search efficiency at early iterations because of misguidance of knee region. To address this problem, we propose an adaptive decomposition-based environmental selection strategy. It searches the whole Pareto front (PF), and meanwhile carries on additional selection of solutions in the knee region by adding more and denser reference vectors in the knee-approximation region at 170 each generation. Hence more solutions can be obtained in this region and the diversity of solutions can be improved.

Thirdly, the local search ability is closely related to the convergence performance and computational time. In light of this, we propose a two-stage local search strategy based on the IST algorithm in ADEA. The first stage provides a high-efficiency local search method for speeding up the convergence, while the second stage helps disperse solutions over the Pareto front PF. Compared with StEMO, which executes the local search over the whole population, ADEA ~~constraints~~ **constrains** the search among subpopulations, which can reduce the reconstruction error while maintaining the diversity. Furthermore, ADEA applies an extra constraint to the regularization parameters, which avoids the generation of useless zero solutions.

3.2. Framework of ADEA

The procedure of our proposed ADEA scheme is summarized in Algorithm 2. In the algorithm, Δr represents the number of added reference vectors in the knee-approximation region, and is used for adjusting the diversity of solutions in the knee region; t denotes the index of the generation; ~~\mathbf{P}^0 and \mathbf{V}^0 represent the initial population and reference vectors respectively.~~

The algorithms start with initialization, where the initial population \mathbf{P}^0 is generated with N randomized individuals, and N uniformly distributed reference vectors are yielded (refer to Section 2.1), denoted as \mathbf{V}^0 . Iterative generations follow the initialization. In each generation, the recombination operator, including the simulated binary crossover[1] and the polynomial mutation [14], are employed to generate offspring solutions \mathbf{Q}^t without any explicit mating selection strategies (Line 4). **In Line 5, \mathbf{Q}^t is combined into \mathbf{P}^t . Then,** the IST-based local search operator is implemented to obtain \mathbf{L}^t (line 6). The population \mathbf{P}^t is updated again, and then the adaptive decomposition-guided environmental selection operator is employed to select the elitism solutions for the next generation and to automatically add reference vectors into the knee-approximation region (line 8), where rl is the lower bound for interpolation section of the reference vector index. When the stopping criterion is met, the Pareto knee point is **obtained** by the kink method [32] and treated as the final solution. The IST-based local search operator and the adaptive decomposition-guided environmental selection operator are detailed in the following subsections.

3.3. Improved IST-based Local Search

Since the existing MOEA shows a slow convergence speed in solving the SR problems, we integrate the ~~improved~~ IST algorithm into ADEA as a local search operator for accelerating the convergence. According to Section 2.4, $\mathbf{x}^{(l)}$, $\mathbf{x}^{(l-1)}$ and λ should be determined first. Next, we discuss how to choose these variables for our improved IST-based local search operator.

3.3.1. Selection of $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$

According to the theorem in [30], in the solution sequence, $\{\mathbf{x}^{(l)}\}$ is generated by the IST algorithms, and $\mathbf{x}^{(l+1)} \preceq \mathbf{x}^{(l)} \preceq \mathbf{x}^{(l-1)}$, **hence the newly generated solution is always as good, or even better than the last one.** ~~From~~ Therefore,

Algorithm 2: Framework of ADEA

Input: \mathbf{A} , \mathbf{b} , Δr **Output:** \mathbf{x}

```
1  $\mathbf{P}^0 \leftarrow$  PopulationInitialization;  
2  $\mathbf{V}^0 \leftarrow$  ReferenceVectorInitialization;  
3 while "the stopping criterion is not met" do  
4    $\mathbf{Q}^t =$ Recombination( $\mathbf{P}^t$ );  
5    $\mathbf{P}^t = \mathbf{P}^t \cup \mathbf{Q}^t$ ;  
6    $\mathbf{L}^t =$ IST-basedLocalSearch( $\mathbf{P}^t, \mathbf{V}^t$ );  
7    $\mathbf{P}^t = \mathbf{P}^t \cup \mathbf{L}^t$ ;  
8   ( $\mathbf{P}^{t+1}, \mathbf{V}^{t+1}, rl$ )= $\text{Selection}(\mathbf{P}^t, \mathbf{V}^t, \mathbf{V}^0, \Delta r, rl)$ ;  
9    $t = t + 1$ ;  
10 end  
11  $\mathbf{x} \leftarrow$  FinalSolutionIdentification( $\mathbf{P}^t$ );
```

215 the selection of $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ in our improved IST-based local search operator should also follow this theorem.

In addition, the local search in ADEA is expected to be executed with the guidance of reference vectors. This can lead the search to converge to the coordinate origin and the diversity of solutions can also be achieved at the same time. More specifically, the current population \mathbf{P}^t should first be normalized (refer to Section 2.2), then in the normalized objective space, each solution in \mathbf{P}^t associates with its closest reference vector (refer to Section 2.3), and finally \mathbf{P}^t is partitioned into several subpopulations \mathbf{P}_j^t by these reference vectors, where $j, j = 1, 2, \dots, N$ are the indexes of the reference vectors.

We introduce the metric fitness evaluation function in [46] as the selection criterion. The fitness of a solution \mathbf{x}_i can be represented by

$$D_{x_i} = d1_{x_i} + 5 \times d2_{x_i} \quad (11)$$

225 where $d1_{x_i} = \|\mathbf{f}(\mathbf{x}_i)\| \times \cos(\theta_{\mathbf{x}_i, \mathbf{v}_j})$ and $d2_{x_i} = \|\mathbf{f}(\mathbf{x}_i)\| \times \sin(\theta_{\mathbf{x}_i, \mathbf{v}_j})$ measure the convergence and diversity of \mathbf{x}_i respectively. The physical meaning of $d1_{x_i}$ and $d2_{x_i}$ can be seen from Fig. 4. We can see that a smaller $d1_{x_i}$ leads to better convergence, and a smaller $d2_{x_i}$ leads to better diversity. Based on the fitness of solutions, we divide \mathbf{P}^t into two sets \mathbf{B} and \mathbf{S} , where \mathbf{B} is made up of the solution with the best fitness in each subpopulation (i.e. $\mathbf{B} = \{\mathbf{x}_i | i = \arg \min D_{x_i}, \mathbf{x}_i \in \mathbf{P}_j^t, j = 1, 2, \dots, N\}$), and the remaining solutions in \mathbf{P}^t constitute \mathbf{S} (namely $\mathbf{S} = \mathbf{P}^t \setminus \mathbf{B}$). It can be found that most solutions in \mathbf{B} are better than those in \mathbf{S} .

230 Now we discuss two cases when selecting $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ from \mathbf{B} and \mathbf{S} , as illustrated in Fig. 5. In Fig. 5(a), $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ are selected from the sets \mathbf{B} and \mathbf{S} respectively. We call it as *convergence-oriented local search*, since the newly generated solution $\mathbf{x}^{(l+1)}$ is more likely to converge to the true PF thanks to $\mathbf{x}^{(l)} \prec \mathbf{x}^{(l-1)}$. In Fig. 5(b), both $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ are chosen from the set \mathbf{B} .

240 We call it as *diversity-oriented local search*. Since $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ are prone to be nondominant with respect to each other, this method disperses the updating solutions over the true PF with high probability.

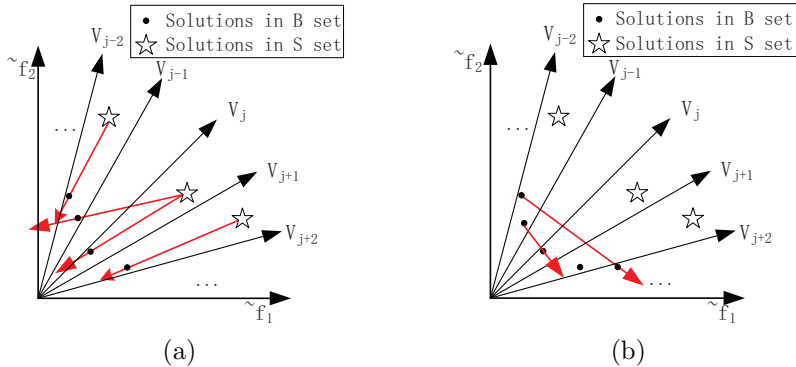


Figure 5: (a)Convergence-oriented local search, $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ are selected from the set \mathbf{B} and \mathbf{S} respectively, which pushes the solutions to converge to the true Pareto front. (b)Diversity-oriented local search. $\mathbf{x}^{(l)}$ and $\mathbf{x}^{(l-1)}$ are both randomly selected from the set \mathbf{B} without repetition, which disperses the solutions over the Pareto front with high probability.

3.3.2. Determining λ

The IST algorithms employ the steady-state strategy, namely only a λ scalar is utilized in each generation. This strategy provides a rather fixed search path and may fall into suboptimal solutions. Existing MOSR algorithms avoid selecting λ values, and they execute more effective parallel searching by using a vector $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \dots, \lambda_K]$ in each generation, where every element is randomly selected between 0 and 1. However, in (4), it can be found that if the chosen λ_i is too large, for example, $\lambda_i \geq \|\mathbf{A}^T \mathbf{b}\|_\infty$, the zero vector would be the unique solution [20][26]. Therefore, jointly considering this phenomenon and the advantage of the parallel search strategy, we propose a more accurate local parallel search strategy: in each generation, we use a vector $\mathbf{\Lambda}$ whose elements follow a uniform distribution:

$$\|\mathbf{A}^T \mathbf{b}\|_\infty \cdot \{\text{rand}(0, 1)\}_{1 \times |\mathbf{B}|}. \quad (12)$$

Equation (12) guarantees that no zero solutions will be attempted, thus improving the local search efficiency.

245 3.3.3. The Complete IST-based Local Search Strategy

Based on the above discussions, the pseudocode of IST-based local search strategy is given in Algorithm 3. First, the population \mathbf{P}^t is normalized. Then, in the normalized objective space, each individual is associated with its nearest reference vector, hence \mathbf{P}^t is divided into a number of subpopulations \mathbf{P}_j^t ($j = 1, 2, \dots, N$). The solution with the best fitness in each subpopulation is moved to the set \mathbf{B} , and the rest are placed into \mathbf{S} . The $\mathbf{\Lambda}$ vector is generated by

250

Algorithm 3: IST-based Local Search Operator

Input: $\mathbf{P}^t, \mathbf{V}^t$
Output: \mathbf{L}^t
1 $\tilde{\mathbf{f}}^t = \text{PopulationNormalization}(\mathbf{P}^t)$;
2 $\mathbf{P}_j^t = \text{IndividualAssociation}(\mathbf{P}^t, \tilde{\mathbf{f}}^t, \mathbf{V}^t)$;
3 $\mathbf{B} = \{\mathbf{x}_i | i = \arg \min D_{x_i}, \mathbf{x}_i \in \mathbf{P}_j^t, j = 1, 2, \dots, N\}$;
4 $\mathbf{S} = \mathbf{P}^t \setminus \mathbf{B}$;
5 $\mathbf{\Lambda}^{1 \times |\mathbf{B}|} \leftarrow \|\mathbf{A}^T \mathbf{b}\|_{\infty} \cdot \{\text{rand}(0, 1)\}_{1 \times |\mathbf{B}|}$;
6 /*Two-stage manner*/
7 **for** $i = 1 : |\mathbf{B}|$ **do**
8 | $\mathbf{L1} = \text{IST}(\mathbf{B}_i, \mathbf{S}_i, \mathbf{\Lambda}^{1 \times |\mathbf{B}|})$;
9 **end**
10 **for** $i = 1 : \lfloor |\mathbf{B}|/2 \rfloor$ **do**
11 | $\mathbf{L2} = \text{IST}(\mathbf{B}_i, \mathbf{B}_{|\mathbf{B}|-i}, \mathbf{\Lambda}^{1 \times \lfloor |\mathbf{B}|/2 \rfloor})$;
12 **end**
13 $\mathbf{L}^t = \mathbf{L1} \cup \mathbf{L2}$;

equation (12). After that, the two-stage manner is executed to obtain the new population \mathbf{L}^t : the convergence-oriented local search is conducted to accelerate the convergence in the first stage (line 8), and the diversity-oriented local search is implemented to spread the solutions over the PF in the second stage (line 11).
255

3.4. Adaptive Decomposition-based Environmental Selection

The pseudo-code of the adaptive decomposition-based environmental selection operator is given in Algorithm 4. In order to use reference vectors to guide the evolutionary process, we first apply population normalization and individual association, as introduced in Section 2.2 and 2.3 respectively. The fitness of each solution is measured by equation (11) (line 6). We then apply the strategy of adaptive decomposition-based environmental selection that includes two main steps: *Elitism selection* and *Reference vector adaptation*. In elitism selection, the reference vectors guide the subpopulations to converge when maintaining the diversity of solutions. In reference vector adaptation, uniformly distributed reference vectors can be dynamically added into the knee-approximation region, thus assisting in selecting more promising solutions in the knee-approximation region in the next generation.
260
265

3.4.1. Elitism Selection

The elitism selection step (line 10~22 of Algorithm 4) is proposed to select elitism individuals for the next generation. Firstly, each subpopulation \mathbf{P}_j^t is updated by solutions sorted through their fitness \mathbf{D} in ascending order (line 11). Then, the solution \mathbf{x}_i with fitness ranked the u -th in \mathbf{P}_j^t is collected to the set \mathbf{E}_u (line 13). Thereafter, several levels $\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_{|\mathbf{V}^t|}\}$ are obtained. Finally, $|\mathbf{V}^t|$ solutions are selected from the highest level \mathbf{E}_1 to low levels and placed
270
275

Algorithm 4: Adaptive Decomposition-based Environmental Selection Operator

Input: $\mathbf{P}^t, \mathbf{V}^t, \mathbf{V}^0, \Delta r, rl, \mathbf{P}^{t+1} = \phi$
Output: $\mathbf{P}^{t+1}, \mathbf{V}^{t+1}, rl$

- 1 $\tilde{\mathbf{f}}^t = \text{PopulationNormalization}(\mathbf{P}^t)$;
- 2 $\mathbf{P}_j^t = \text{IndividualAssociation}(\mathbf{P}^t, \tilde{\mathbf{f}}^t, \mathbf{V}^t)$;
- 3 /*fitness Measurement*/
- 4 **for** $j = 1 : |\mathbf{V}^t|$ **do**
- 5 **for** $i = 1 : |\mathbf{P}_j^t|$ **do**
- 6 $D_{x_i} = d1_{x_i} + 5 \times d2_{x_i}$;
- 7 **end**
- 8 **end**
- 9 /*Elitism Selection*/
- 10 $\mathbf{E} = \{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_{|\mathbf{V}^t|}\} = \{\phi, \phi, \dots, \phi\}$;
- 11 $\mathbf{P}_j^t \leftarrow$ sort \mathbf{P}_j^t based on \mathbf{D} in ascending order;
- 12 **for** $u = 1 : \max |\mathbf{P}_j^t|$ **do**
- 13 $\mathbf{E}_u \leftarrow$ the u -th solution in $\mathbf{P}_j^t, j \in 1, 2, \dots, |\mathbf{V}^t|$;
- 14 **end**
- 15 $u = 1$;
- 16 **while** $\sum |\mathbf{E}_u| < |\mathbf{V}^t|$ **do**
- 17 $\mathbf{P}^{t+1} = \mathbf{P}^{t+1} \cup \mathbf{E}_u$;
- 18 $u = u + 1$;
- 19 **end**
- 20 **if** $|\mathbf{P}^{t+1}| > |\mathbf{V}^t|$ **then**
- 21 $\mathbf{P}^{t+1} = \mathbf{P}^{t+1} \setminus \text{RandomSelect}(\mathbf{E}_{u-1})$;
- 22 **end**
- 23 /*Reference Vector Adaptation*/
- 24 Compute the fitness of solutions: $F_{\mathbf{x}_i, \mathbf{v}_j} = \lambda_0 \cdot \|\mathbf{x}_i\|_1 + \frac{1}{2} \|\mathbf{A}\mathbf{x}_i - \mathbf{b}\|_2^2$,
where $\mathbf{x}_i \in \mathbf{P}^{t+1}, j \in 1, 2, \dots, |\mathbf{V}^t|$, and $\lambda_0 = \text{rand}(0, 1) \times \|\mathbf{A}^T \mathbf{b}\|_\infty$;
- 25 $ro \leftarrow \{j | \mathbf{x}_i = \arg \min F_{\mathbf{x}_i, \mathbf{v}_j}\}$;
- 26 **if** $ro > N$ **then**
- 27 $ro = ro - N + rl$;
- 28 **end**
- 29 $rh = \min(N, ro + \Delta r)$;
- 30 $rl = \max(1, ro - \Delta r)$;
- 31 $\mathbf{V}^{t+1} \leftarrow \mathbf{V}^0$;
- 32 $\mathbf{V}^{t+1} \leftarrow$ add $(rh - rl)$ new reference vectors between rl and rh and
place them after the last row of \mathbf{V}^{t+1} ;

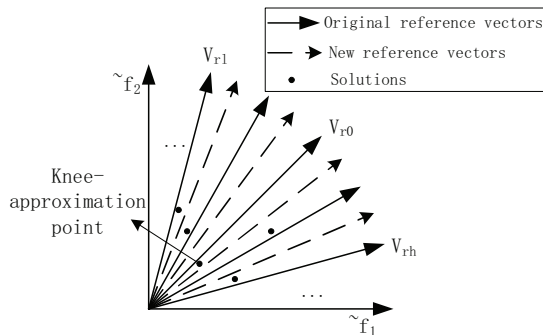


Figure 6: An example of reference vector adaptation.

into \mathbf{P}^{t+1} (Line 16-19). Once the number of selected solutions exceeds $|\mathbf{V}^t|$, the extra selected solutions from the last accepted level will be randomly removed (Line 20-22).

3.4.2. Reference Vector Adaptation

280 It has been proved that a nearly optimal solution exists in the knee region [30], so the selection should be more focused on the knee-approximation region in each generation. Therefore, the reference vector adaptation is introduced (line 23~32 of Algorithm 4) to add extra reference vectors into the knee-approximation region. By the proposed reference vector adaptation strategy, more solutions in the knee-approximation region will be selected in the next generation. [There is the problem of estimating the knee-approximation region in each generation here.](#) To solve this problem, equation (4) is used to evaluate the quality of each solution, and the solution with the smallest value can be regarded as the knee-approximation point in the current generation. For each generation, we employ (12) to determine the λ scalar in (4).
290

The pseudo-code of reference vector adaptation is shown in line 23~32 of Algorithm 4. Firstly, the fitness of each solution is computed according to (4). Then, the solution with the best fitness is [treated](#) as the knee-approximation point, and [the index of its associating reference vector](#) is assigned to ro . As the reference vector adaptation strategy adds new reference vectors [to the bases](#) of \mathbf{V}^0 , the ro -th reference vector should belong to \mathbf{V}^0 . If not, ro will be updated by $ro = ro - N + rl$. Next, the reference vector interpolation section $[rl, rh]$ in V^0 can be obtained by line 29~30. [Finally](#), Δr new reference vectors are inserted at both sides of \mathbf{V}_{ro}^0 to [generate](#) \mathbf{V}^{t+1} . An example of reference vector adaptation is given in Fig. 6, where \mathbf{V}_{rl}^0 and \mathbf{V}_{rh}^0 determine the reference interpolation section [via](#) the position of \mathbf{V}_{ro}^0 , and $2 \times \Delta r$ new reference vectors are inserted in this section.
300

3.5. Final Solution Identification

285 The knee region provides a [good](#) trade-off between the two objectives, and it has been proved that a nearly optimal solution exists in this region [30]. A

number of methods have been proposed to find the knee solution [6, 33]. The process of **identifying the final solution** is shown in Algorithm 5. The Pareto nondominant solutions from the final solution set are selected and normalized in the interval [0, 1]. Then, the normalized solutions are fitted by B-splines and sampled evenly. Finally, the kink method [32] is used to locate the knee point as the final solution.

Algorithm 5: Final Solution Identification

Input: \mathbf{P}^t

Output: \mathbf{x}

- 1 $\mathbf{PF} \leftarrow$ Find the nondominant solutions from \mathbf{P}^t ;
 - 2 Normalize \mathbf{PF} ;
 - 3 Employ B-splines to yield the \mathbf{PF} fitting curve $\bar{\mathbf{PF}}$;
 - 4 Sample evenly on $\bar{\mathbf{PF}}$;
 - 5 $\mathbf{x} \leftarrow$ find knee point on $\bar{\mathbf{PF}}$ using the kink method;
-

In the kink method, the solution with the largest slope variance is determined as the nearly optimal solution. To be specific, for a solution \mathbf{x}_i and its objective values $(f_1(\mathbf{x}_i), f_2(\mathbf{x}_i))$, its slope variance of $\Delta\beta_i$ is

$$\begin{cases} \beta_i = \arctan\left(\frac{f_1(\mathbf{x}_i) - f_1(\mathbf{x}_{i+1})}{f_2(\mathbf{x}_i) - f_2(\mathbf{x}_{i+1})}\right), \\ \Delta\beta_i = \frac{\beta_i}{\beta_{i-1}}, \end{cases} \quad (13)$$

where $\beta_0 = 2/\pi$. Thereafter, the solution with the largest $\Delta\beta$ value is defined as the nearly optimal solution.

3.6. Computational Complexity Analysis

As shown in Algorithms 2 to 5, in each generation, the computational complexity of ADEA is dominated by the improved IST-based local search and the adaptive decomposition-based environment selection procedures. The complexity of the proposed IST-based local search method is upper bounded by $o(\frac{3}{2}(N + 2\Delta r)n^2)$. In the proposed selection procedure, the complexity of individual association, elitism selection and reference vector adaptation is $o(2 \times (N + 2\Delta r)^2)$, $o((N + 2\Delta r)^2)$ and $o(2 \times \Delta r)$ in the worst case, respectively. Since $N \ll n$, the overall computational complexity of our proposed ADEA scheme is $o(\frac{3}{2}(N + 2\Delta r)n^2)$ in the worst case.

Comparatively, the computational complexity of StEMO and LBEA is $o(2Nn^2)$ and $(15Nn^2)$ in the worst case respectively. Hence ADEA achieves reduced computational complexity, in addition to the improved performance as will be seen from the experimental results in next Section.

4. Experimental Results

4.1. Experiment Settings

330 In this section, numerical experiments on simulated signals, benchmark problems and images are provided to compare the reconstruction quality between the proposed ADEA and other state-of-the-art algorithms, i.e. basis pursuit (BP) [10], orthogonal matching pursuit (OMP) [7], Homotopy [27], fast iterative shrinkage-thresholding algorithm (FISTA) [42], StEMO [30] and LBEA 335 [44]. Among them, BP, OMP, Homotopy and FISTA belong to single-objective SR algorithms, and StEMO, LBEA and ADEA are MOSR algorithms. The reconstruction error (RE) is used to evaluate the reconstruction quality, which is defined by $RE = \|\mathbf{x} - \mathbf{x}_{true}\|/\|\mathbf{x}_{true}\|$, where \mathbf{x}_{true} and \mathbf{x} are the real and estimated signal respectively. All experiments are implemented in MATLAB 340 R2014a on the PC with 3.3GHz Intel Xeon E3 processor and 8G memory on Window 7 system. The λ value for BP, OMP, Homotopy and FISTA is set to 0.02. The major parameters of ADEA include crossover probability, distribution index of crossover, mutation probability and distribution index of mutation, which are set to 1, 20, $1/n$ and 20 respectively. For a fair comparison, 345 the population size for StEMO, LBEA and ADEA are set to 120, 120 and 100 respectively, because of the reference adaptation strategy in ADEA.

4.2. Comparison Using Simulated Signals

In the following experiments, a sparse signal \mathbf{x}_{true} is generated from k Gaussian random nonzero elements ($k \ll n$). The measurement vector b is obtained 350 by the product of \mathbf{x}_{true} and the Gaussian matrix \mathbf{A} . The measurement vector is corrupted by white noise whose standard deviation is δ . The sampling rate and sparsity ratio are $\frac{m}{n}$ and $\frac{k}{n}$ respectively. Each algorithm runs 20 times in each test case. We first investigate the reconstruction quality of the MOSR algorithms, and then compare all seven algorithms.

4.2.1. Comparisons of the MOSR algorithms

355 The reconstruction performance and running time of the MOSR algorithms are compared under different sampling rates, sparsity ratios, noise levels and length of signals as shown in Figs. 7~10. The stopping criterion is that all the algorithms stop running when the number of iterations reach reaches the maximum generation t_{max} . 360

In Fig. 7, the average RE results and running time of StEMO, LBEA and ADEA are compared under different sampling rates from 0.25 to 0.65 under noiseless condition. The signal parameters are $n = 1000$ and $k = 50$. The maximum iterations t_{max} is 100. As can be observed from Fig. 7, when the sampling 365 rates increase, the reconstruction quality for all three algorithms improves. The running time for LBEA and ADEA roughly remains unchanged, but for StEMO, it rapidly increases. ADEA outperforms StEMO and LBEA in reconstruction quality under all the sampling rates, even when the sampling rate is relatively small. ADEA always runs much faster than StEMO and LBEA.

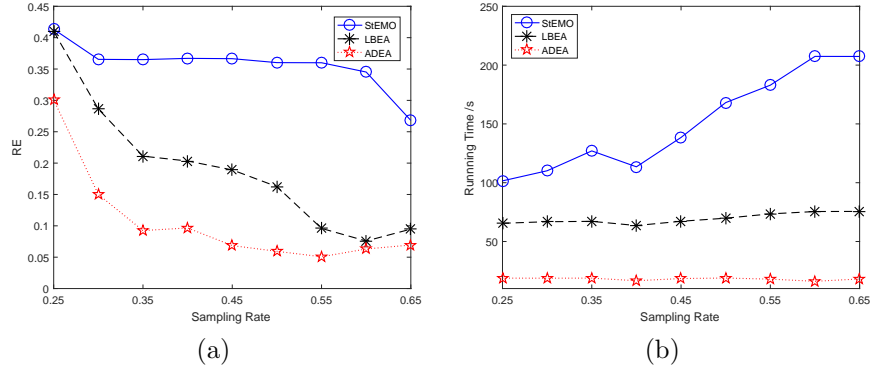


Figure 7: Comparisons of the MOSR algorithms under different sampling rates from 0.25 to 0.65 under noiseless condition, $n = 1000$, $k = 50$, (a) RE. (b) Running Time.

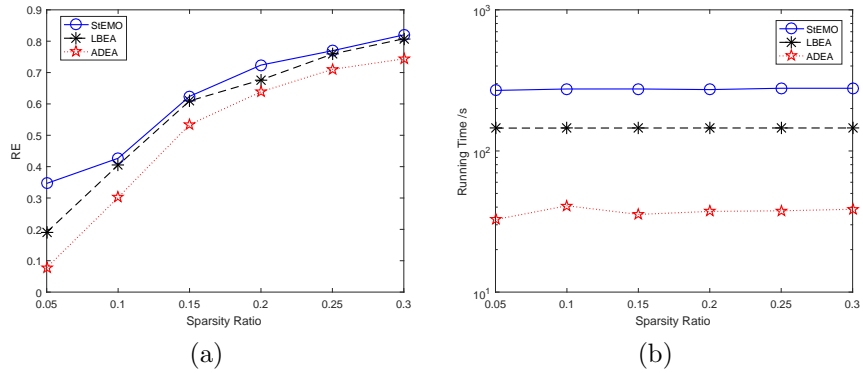


Figure 8: Comparisons of the MOSR algorithms under different sparsity ratios from 0.05 to 0.30 under noiseless condition, $n = 1000$, $m = 400$, (a) average RE. (b) average running time.

370 In Fig. 8, the average RE results and running time of StEMO, LBEA and
ADEA are compared under different sparsity ratios from 0.05 to 0.30 under
noiseless condition. The signal parameters are $n = 1000$ and $m = 400$. As it
becomes harder to recover less-sparse signals, the maximum iterations t_{max} is
increased to 200. Fig. 8(a) shows that the reconstruction quality of all ~~there~~
375 ~~three~~ algorithms deteriorate with higher sparsity ratio, but ADEA is always
superior to StEMO and LBEA. Fig. 8(b) illustrates that the average running
time of all the algorithms remains unchanged with the increasing sparsity ratio,
and ADEA is always the fastest among the three.

380 In Fig. 9, the average RE results and running time for StEMO, LBEA and
ADEA are compared under different noise levels from 0 to 0.018. The signal
parameters are $n = 1000$, $m = 400$ and $k = 50$. Fig. 9(a) indicates that
the performance of ADEA is slightly better than that of StEMO and LBEA
when the noise level δ surpasses 0.009. While in other cases, ADEA achieves

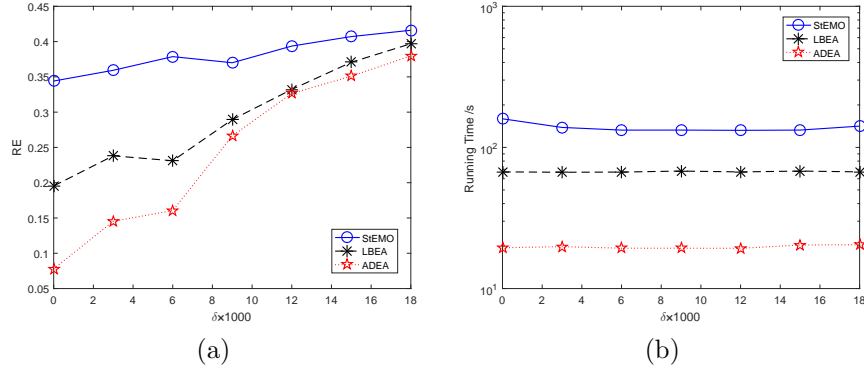


Figure 9: Comparisons of the MOSR algorithms under different noise levels from 0 to 0.018, $n = 1000$, $m = 400$, $k = 50$, (a) average RE. (b) average running time.

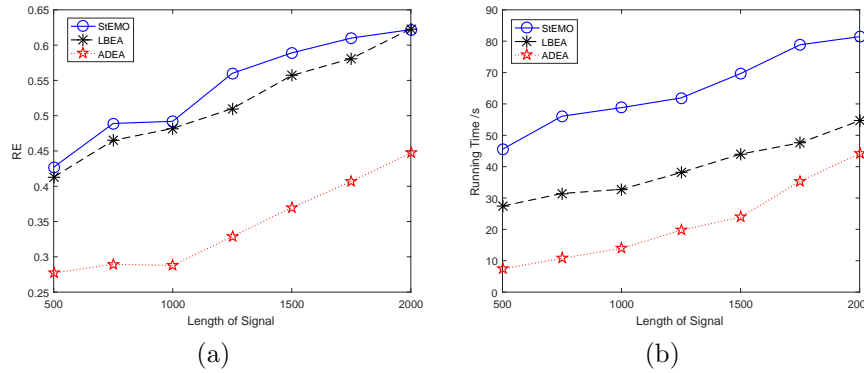


Figure 10: Comparisons of the MOSR algorithms under different lengths of signals from 500 to 2000, $m = 400$, $k = 50$ and $\delta = 0$. (a) average RE. (b) average running time.

the smallest reconstruction errors. In Fig. 9(b), it is shown that the average running time is independent of noise levels, and ADEA has faster reconstruction speed than StEMO and LBEA.

The average RE results and running time of StEMO, LBEA and ADEA are compared under different lengths of signals from 500 to 2000. The signal parameters are $m = 400$, $k = 50$, $\delta = 0$ and $t_{max} = 100$. As can be seen from Fig. 10(a), all algorithms get worse reconstruction quality with larger n , but the performance of ADEA still surpasses that of StEMO and LBEA. In Fig. 10(b), the running time of three algorithms increases as n grows, and ADEA has the fastest reconstruction speed under all conditions.

To summarize, ADEA outperforms StEMO and LBEA in reconstruction quality, regardless of the changes in measurement times, sparsity, noise levels, or lengths of signals. Furthermore, due to the low computational complexity of the decomposition-based paradigm, ADEA requires much less running time in

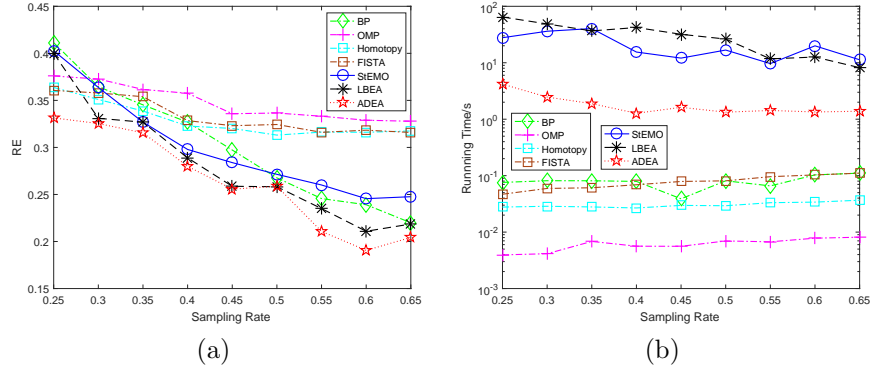


Figure 11: Comparisons of all seven algorithms under different sampling ratios from 0.25 to 0.65 under noiseless condition, $n = 1000$, $k = 50$. (a) Average RE, and (b) Average running time.

all test scenarios.

4.2.2. Comparison of MOSR and single-objective SR algorithms

400 This part compares the reconstruction quality and runtime for all the seven algorithms, i.e., BP, OMP, Homotopy, FISTA, StEMO, LBEA and ADEA with increased times of measurement or sparsity levels. The parameter of stopping criterion is set to be 0.1. First, the parameters of the simulated signal are set to $n = 1000$ and $k = 50$ under noiseless measurement condition.

405 Fig. 11 shows the average RE results and running time under different sampling rates (0.25 \sim 0.65). Fig. 11(a) demonstrates that our method ADEA achieves the best reconstruction performance compared with other five algorithms. It is worth noting that ADEA performs quite well even with limited measurements, which is desired for real applications. Fig. 11(b) shows that single-objective SR algorithms (i.e. BP, OMP, Homotopy and FISTA) run faster than MOSR algorithms. However, it is worth noting that ADEA achieves the fastest reconstruction among the evolutionary optimizers.

410 Fig. 12 shows the average RE results and running time under different sparsity levels and noiseless conditions. From Fig. 12(a), it can be observed that ADEA outperforms other algorithms in terms of reconstruction quality, except when the sparsity ratio is equal to or greater than 0.2. From Fig. 12(b), we can see that the average running time of the MOSR algorithms is larger than that of traditional CS algorithms, but ADEA runs much faster than StEMO and LBEA.

420 Therefore, ADEA achieves the best reconstruction quality in all test cases, particularly when the measurement times are small. Evolutionary optimizers typically run more slowly than traditional methods, mainly because they search multiple solutions in a single run. In spite of this, ADEA achieves enormous improvement over existing MOSR algorithms on running time.

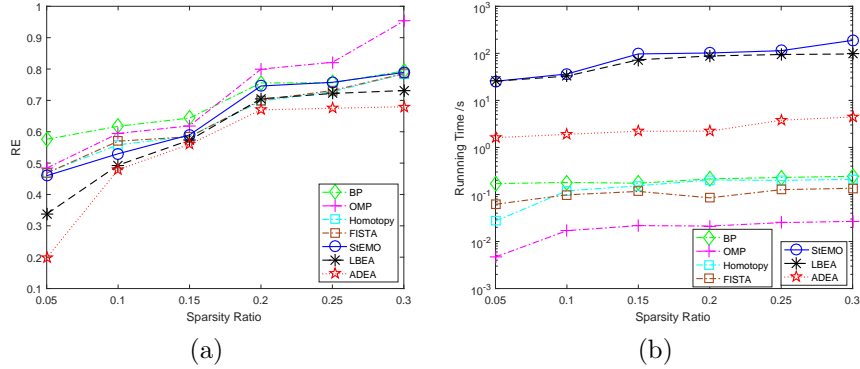


Figure 12: Comparisons of all seven algorithms under different sparsity levels from 0.05 to 0.3 under noiseless condition, $n = 1000$, $m = 400$. (a) Average RE, and (b) Average running time.

4.3. Comparison on Benchmark Problems

In this section, experimental results on practical benchmark problems are presented to compare the reconstruction performance of the seven algorithms. The Sparco toolbox [38] is used to provide a variety of benchmark problems, in which five signals, including sngspike, gausspike, cospike, gcospike and jitter, are adopted to compare the reconstruction quality of the seven algorithms. The sngspike and gausspike signals are sparse themselves, which can be sampled and directly reconstructed. The cospike, gcospike and jitter signals \mathbf{x} are not sparse themselves, but they have a sparse representation \mathbf{w} regarding to some basis vectors Ψ . We need to reconstruct \mathbf{w} and then obtain \mathbf{x} by $\mathbf{x} = \Psi\mathbf{w}$. All the measurements are corrupted by additive Gaussian noise, with mean 0 and standard deviation 0.01.

Our ADEA scheme performs very well in reconstructing these benchmark signals, as can be seen from Fig.13, which demonstrates the average reconstruction errors for these seven algorithms. All the seven algorithms show good performance for reconstructing sngspike, and ADEA achieves the smallest reconstructing error. For the gausspike problem, BP, FISTA and ADEA perform best, while StEMO is the worst. Except for Homotopy, other algorithms can roughly realize exact reconstruction on the cospike problem, and ADEA provides superior reconstruction performance, together with OMP and FISTA. In regard to gausspike, Homotopy and StEMO fail to reconstruct this signal, and FISTA ranks first, and both BP and ADEA rank second, in terms of reconstruction performance. Finally, ADEA is placed in the leading position in regards to achieving excellent performance for reconstructing the jitter signal.

4.4. Comparison on Images

This section further evaluates the performance of ADEA on image reconstruction. The test images include Fruits, Lena, Cameraman and Peppers

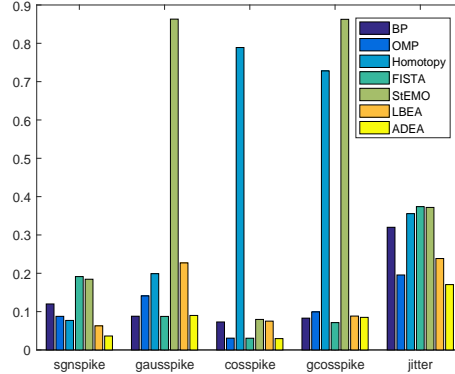


Figure 13: Average reconstruction errors of seven algorithms on the benchmark problems



Figure 14: Original images of 256×256 pixels. (a) Fruits. (b) Lena. (c) Cameraman. (d) Peppers.

images with the size of 256×256 , as shown in Fig. 14. Many state-of-the-art algorithms can robustly recover k -sparse and compressible signals from $m \geq O(k \log(n/k))$ noisy measurements [2]. Sparser \mathbf{x} may lead to better reconstruction quality with the same number of measurements. Use the Lena image as an example. Fig. 15 (a) and (b) shows its histogram in the image and wavelet domain, respectively. We can see that the histogram in the image domain is not sparse, but in the wavelet domain, it has a large peak near zero indicating the sparsity of wavelet coefficients. Therefore, Haar wavelets are employed as the sparse basis matrix Ψ in order to achieve better reconstruction results. In this experiment, we sample the wavelet coefficients of those images with a Gaussian random matrix. The sampling rate is fixed at 0.5, and the measurement data is corrupted by white noise ($\delta = 0.01$). The parameter σ for the stopping criterion is set to 0.1. The reconstruction quality is measured with respect to PSNR and SSIM [40].

Table 1 lists the reconstruction performance for the seven algorithms, and the best results are highlighted in bold. We can see that ADEA achieves at least 1 dB better PSNR values than other algorithms for the first three images. For peppers image, ADEA has the similar reconstruction performance to

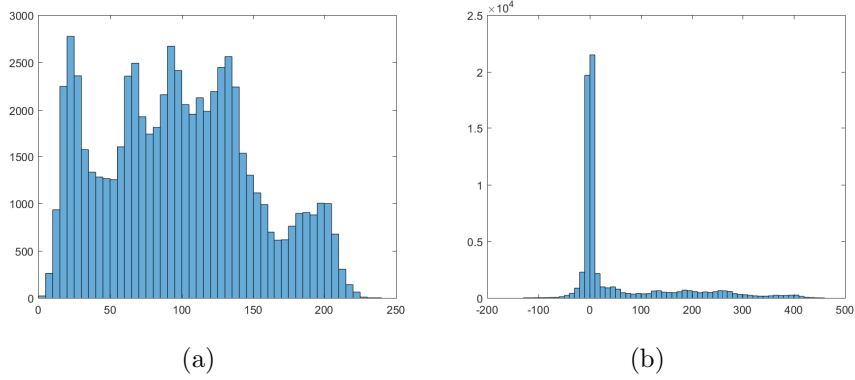


Figure 15: Histograms of Lena image in (a) image domain. (b) wavelet domain.

Table 1: Comparisons of PSNR and SSIM values for seven algorithms on image reconstruction

Algorithms	Fruits	Lena	Cameraman	Peppers
BP	25.7499(0.9655)	28.3154(0.9708)	26.7079(0.9819)	25.7079(0.9718)
OMP	24.2819(0.9527)	26.2946(0.9713)	23.3382(0.9623)	24.8741(0.9674)
Homotopy	26.0502(0.9702)	28.3491(0.9799)	25.9777(0.9791)	26.9523(0.9779)
FISTA	26.9525(0.9711)	28.5824(0.9815)	26.0053(0.9792)	26.9563(0.9779)
StEMO	28.3582(0.9843)	29.8972(0.9880)	26.2489(0.9804)	27.3173(0.9805)
LBEA	28.3299(0.9810)	30.5323(0.9897)	27.1543(0.9849)	27.4767(0.9812)
ADEA	29.9721(.9893)	31.5485(0.9928)	28.4509(0.9923)	27.4582(0.9824)

470 StEMO and LBEA, and is substantially better than the single-objective algo-
rithms. Fig. 16 also displays the visual quality of the reconstructed Lena image,
corresponding to the quantitative results as shown in Table 1. In the figure, the
odd rows show the original or reconstructed images and the even rows show the
corresponding enlarged local parts. ADEA demonstrates the best visual quality
475 among the seven algorithms, and OMP is the worst due to speckles.



Figure 16: Original and reconstructed Lena images by seven algorithms.

5. Further Study of ADEA

In this section, the advantages of ADEA operators, diversity of solutions in knee region, and robustness of final solutions, are further investigated to demonstrate the superior performance of [the proposed ADEA scheme](#).

480 5.1. Study of ADEA Operators

Each operator in ADEA plays a [specific role in reconstruction](#). The decomposition-based evolutionary paradigm partitions the current population into several subpopulations and carries out the [collaborative evolution](#) on those subpopulations. It therefore searches the potential solutions [globally and improves the diversity of solutions](#) across the PF. The IST-based local search operator [improves](#)

485

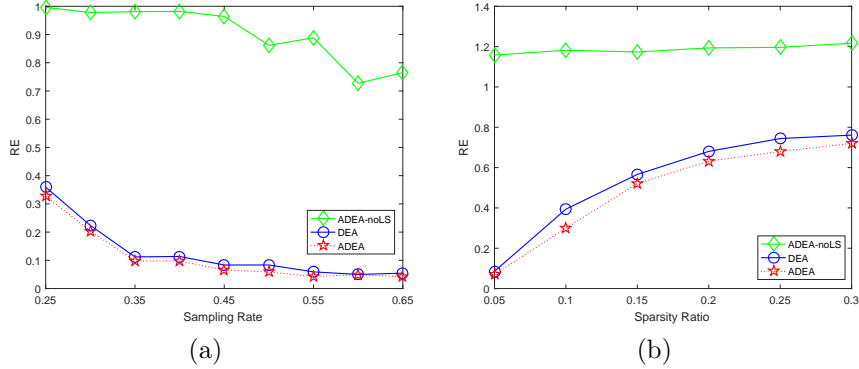


Figure 17: (a) Comparisons of DEA and ADEA under different sampling rates from 0 to 0.018 under noiseless condition, $n = 1000$, $k = 50$. (b) Comparisons of DEA and ADEA under different sparsity ratio from 0.05 to 0.30 under noiseless condition, $n = 1000$, $m = 400$.

the scheme’s exploitation ability and accelerates its convergence. The adaptive decomposition-based environmental selection operator acts on elitism selection with improved convergence and diversity. In particular, this reference vector adaptation strategy assists in selecting more valuable solutions in the knee region, which improves the convergence speed.

In this subsection, we present experimental results to further study the impact of IST-based local search operator and the reference vector adaptation strategy on reconstruction performance. We compare the reconstruction results for the proposed ADEA, the ASEA version without the IST-based local search operator (termed as ADEA-noLS), and the one without reference vector adaptation (termed as DEA).

Fig. 17 (a) and (b) shows the reconstruction results under different sampling rates and sparsity ratios respectively. The parameter settings are the same with those in Fig.(7) and Fig.(8) respectively. It can be seen that, ADEA-noLS fails to reconstruct the original signals, and ADEA achieves better reconstruction quality than DEA in most cases with various sampling rates or sparsity ratios. These results further validate the importance of IST-based local search operator and reference vector adaptation strategy in improving the convergence performance of ADEA.

5.2. Diversity of Solutions in the Knee Region

In this subsection, we study the diversity of solutions in the knee region for the three MOSR algorithms (i.e. StEMO, LBEA and ADEA). This further helps us understand the importance of the proposed reference vector adaptation strategy in ADEA. We choose two hard measurement cases, where the signal is generated in the same way as that in Section 4.2. In Case 1, some parameters of the signal are $n = 1000$, $m = 250$ and $k = 50$, without additive noise. In Case 2, the length and sparsity of signal are $n = 1000$ and $k = 50$ respectively,

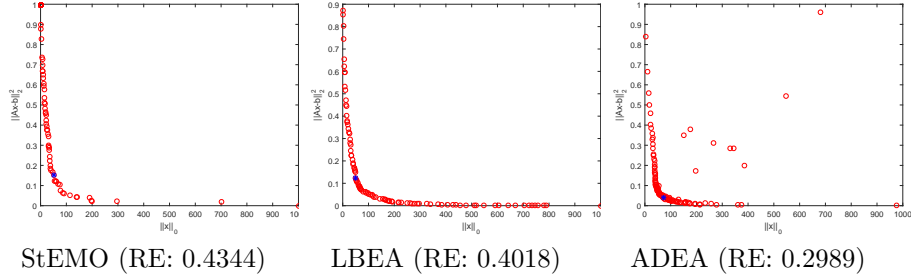


Figure 18: Comparison of the diversity in the knee region for MOSR algorithms in Case 1.

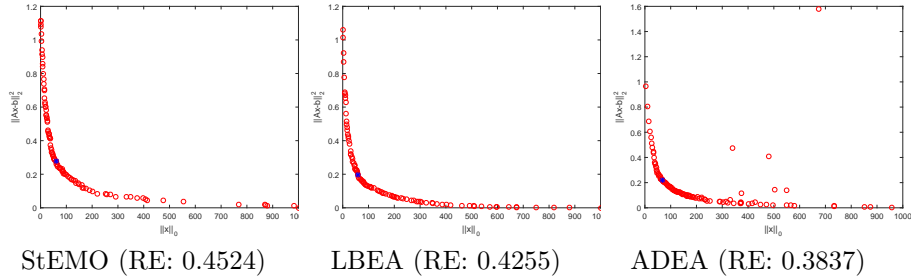


Figure 19: Comparison of the diversity in the knee region of MOSR algorithms in Case 2.

the sampling measurement m is set to 400, and the measurement is corrupted by white noise with mean zero and standard deviation 0.02.

515 The diversity of the final solutions for the MOSR algorithms in Case 1 is shown in Fig. 18. The red points denote all the final solutions and the blue point represents the knee point. ADEA achieves the best performance, and most of the solutions in ADEA are concentrated in the knee region. The curve fitted by those solutions in ADEA is more smooth than those in StEMO and LBEA. However, the solutions in StEMO's knee region are the most sparse and non-uniform, and the corresponding fitting curve is rough. This may lead to bad robustness and low accuracy, as we have seen from the reconstruction of the gausspike and gcrosspike benchmark problems in Fig. 13.

520 Fig. 19 illustrates the final solutions for the three MOSR algorithms in Case 2. ADEA has the best RE values and provides more solutions in the knee region, in comparison to StEMO and LBEA. The solutions in the knee region of ADEA are distributed more uniformly than those in the other two algorithms, which offers strong potential for resisting PF distortion (see Fig. 6(a) in [30]).

530 Therefore, it can be concluded that ADEA provides solutions that are more evenly distributed with higher density in the knee region, and hence improving the success rate of finding the knee point and the reconstruction accuracy.

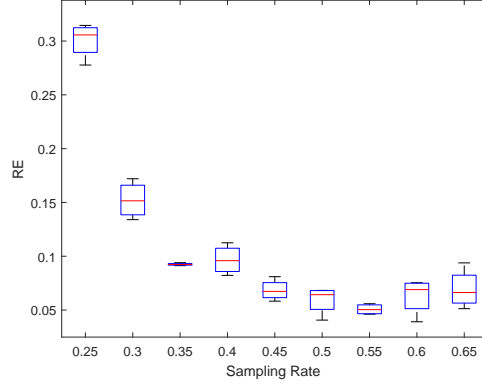


Figure 20: The boxplots of reconstruction results of thirty independent trials ($n = 1000$, $k = 50$ and t_{max} is 100)

5.3. Robustness of ADEA

Since the evolutionary searching process is random and unpredictable, and the λ value is not unique, the final solutions given by ADEA may be not unique. However, thanks to the exhaustive searching process in exploring multiple solution paths (i.e., we searched N solutions in the iteration process of ASEA, rather than adopting the steady-state search in conventional SR algorithms), a good approximate to the ground-truth signal can still be obtained. To verify this, we conducted thirty independent trials with varying sampling rates. The experimental settings are the same with those in Fig. 7 ($n = 1000$, $k = 50$ and t_{max} is 100). Fig. 20 presents the boxplot results for the reconstruction error. We can see that the final solutions have little RE difference at each sampling rate. Therefore, it can be concluded that ADEA is very robust.

6. Conclusions

In this paper, an adaptive decomposition-based evolutionary algorithm (ADEA) is proposed for the multiobjective sparse reconstruction problem. The ADEA scheme includes three major innovations: (1) a decomposition based evolutionary is introduced into ADEA to provide high computational efficiency and reconstruction accuracy; (2) an improved iterative soft thresholding-based local search is applied to further improve the exploitation ability of the algorithm; (3) an adaptive decomposition-based environmental selection operator with a reference vector adaptation strategy is developed to adaptively adjust the diversity in the knee region and to enable the finding of elitism solutions at low computational complexity.

The performance of ADEA on signal and image reconstruction problems is investigated by comparing it to five state-of-the-art algorithms. According to the experimental results, ADEA consistently achieves best reconstruction quality in

most cases. Furthermore, ADEA requires much less running time than existing evolutionary optimizers. Therefore, ADEA is a very competitive evolutionary optimizer for the SR problem. Given its medium computational complexity, ADEA can be applied to real-time applications, such as reconstructing slowly time-varying signals. To further accelerate the running speed of ADEA, parallel implementation [17] or distributed clusters [23] could be employed.

In the future, the decomposition-based selection strategy could be incorporated into other advanced global and local search operators to further improve the performance of evolutionary optimizers for MOSR. Since locating the knee point is the key to [identifying the final solution](#), future work can focus on studying more robust and effective methods for finding the knee point.

Acknowledgement

This work is supported by China Scholarship Council with Grant number 201706540025.

References

- [1] R. B. Agrawal, K. Deb, R. Agrawal, Simulated binary crossover for continuous search space, *Complex systems* 9 (2) (1995) 115–148.
- [2] R. G. Baraniuk, V. Cevher, M. F. Duarte, C. Hegde, Model-based compressive sensing, *IEEE Transactions on Information Theory* 56 (4) (2010) 1982–2001.
- [3] J. Barzilai, J. M. Borwein, Two-point step size gradient methods, *IMA journal of numerical analysis* 8 (1) (1988) 141–148.
- [4] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM journal on imaging sciences* 2 (1) (2009) 183–202.
- [5] T. Blumensath, Accelerated iterative hard thresholding, *Signal Processing* 92 (3) (2012) 752–756.
- [6] J. Branke, K. Deb, H. Dierolf, M. Osswald, Finding knees in multi-objective optimization, in: *PPSN*, vol. 3242, 2004.
- [7] T. T. Cai, L. Wang, Orthogonal matching pursuit for sparse signal recovery with noise, *IEEE Transactions on Information theory* 57 (7) (2011) 4680–4688.
- [8] E. J. Candès, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathématique* 346 (9-10) (2008) 589–592.

- [9] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Communications on pure and applied mathematics* 59 (8) (2006) 1207–1223.
- [10] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM review* 43 (1) (2001) 129–159.
- [11] Y. Chen, N. M. Nasrabadi, T. D. Tran, Hyperspectral image classification using dictionary-based sparse representation, *IEEE Transactions on Geoscience and Remote Sensing* 49 (10) (2011) 3973–3985.
- [12] R. Cheng, Y. Jin, M. Olhofer, B. Sendhoff, A reference vector guided evolutionary algorithm for many-objective optimization, *IEEE Transactions on Evolutionary Computation* 20 (5) (2016) 773–791.
- [13] I. Das, J. E. Dennis, Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems, *SIAM Journal on Optimization* 8 (3) (1998) 631–657.
- [14] K. Deb, M. Goyal, A combined genetic adaptive search (geneas) for engineering design, *Computer Science and informatics* 26 (1996) 30–45.
- [15] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints., *IEEE Trans. Evolutionary Computation* 18 (4) (2014) 577–601.
- [16] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multi-objective genetic algorithm: Nsga-ii, *IEEE transactions on evolutionary computation* 6 (2) (2002) 182–197.
- [17] A.-W. Deng, C.-H. Wei, C.-Y. Gwo, Algorithms for computing zernike moments and image reconstruction in parallel process, in: *Information Science and Control Engineering (ICISCE), 2015 2nd International Conference on*, IEEE, 2015.
- [18] D. Donoho, Y. Tsaig, Fast solution of l1-norm minimization problems when the solution may be sparse, 2006, Preprint 1 (2).
- [19] D. L. Donoho, Compressed sensing, *IEEE Transactions on information theory* 52 (4) (2006) 1289–1306.
- [20] J. J. Fuchs, More on sparse representations in arbitrary bases, *IFAC Proceedings Volumes* 36 (16) (2003) 1315–1320.
- [21] W. Gong, L. Hu, J. Li, W. Li, Combining sparse representation and local rank constraint for single image super resolution, *Information Sciences* 325 (2015) 1–19.

- 630 [22] E. T. Hale, W. Yin, Y. Zhang, Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence, *SIAM Journal on Optimization* 19 (3) (2008) 1107–1130.
- [23] P. Han, R. Niu, M. Ren, Distributed approximate message passing for compressed sensing, CoRR abs/1404.3766.
URL <http://arxiv.org/abs/1404.3766>
- 635 [24] K. K. Herrity, A. C. Gilbert, J. A. Tropp, Sparse approximation via iterative thresholding, in: *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3, IEEE, 2006.
- [25] B. Huang, S. Ma, D. Goldfarb, Accelerated linearized bregman method,
640 *Journal of Scientific Computing* 54 (2-3) (2013) 428–453.
- [26] S. J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale ℓ_1 -regularized least squares, *IEEE Journal of Selected Topics in Signal Processing* 1 (4) (2008) 606–617.
- [27] D. Li, Q. Wang, Y. Shen, Intelligent greedy pursuit model for sparse reconstruction based on ℓ_0 minimization, *Signal Processing* 122 (2016) 138–151.
645
- [28] H. Li, Y. Fan, Q. Zhang, Z. Xu, J. Deng, A multi-phase multiobjective approach based on decomposition for sparse reconstruction, in: *Evolutionary Computation (CEC), 2016 IEEE Congress on*, IEEE, 2016.
- [29] H. Li, Q. Zhang, J. Deng, Z.-B. Xu, A preference-based multiobjective evolutionary approach for sparse optimization, *IEEE transactions on neural networks and learning systems*.
650
- [30] L. Li, X. Yao, R. Stolkin, M. Gong, S. He, An evolutionary multiobjective approach to sparse reconstruction, *IEEE Transactions on Evolutionary Computation* 18 (6) (2014) 827–845.
- 655 [31] Q. Lyu, Z. Lin, Y. She, C. Zhang, A comparison of typical p minimization algorithms, *Neurocomputing* 119 (2013) 413–424.
- [32] I. Mierswa, M. Wurst, Information preserving multi-objective feature selection for unsupervised learning, in: *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, ACM, 2006.
- 660 [33] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, C. A. C. Coello, Survey of multiobjective evolutionary algorithms for data mining: Part ii, *IEEE Transactions on Evolutionary Computation* 18 (1) (2014) 20–35.
- [34] B. K. Natarajan, Sparse approximate solutions to linear systems, *SIAM journal on computing* 24 (2) (1995) 227–234.

- 665 [35] D. Needell, J. A. Tropp, Cosamp: Iterative signal recovery from incomplete and inaccurate samples, *Applied and Computational Harmonic Analysis* 26 (3) (2009) 301–321.
- [36] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization* 11 (4) (1997) 341–359.
- 670 [37] J. A. Tropp, A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Transactions on information theory* 53 (12) (2007) 4655–4666.
- [38] E. van den Berg, M. P. Friedlander, G. Hennenfent, F. J. Herrmann, R. Saab, Özgür Yilmaz, Algorithm 890: Sparco: A testing framework for sparse reconstruction, *ACM Transactions on Mathematical Software* 35 (4) (2009) 1–16.
URL <http://doi.acm.org/10.1145/1462173.1462178>
- [39] X. Wang, F. Liu, L. Jiao, J. Wu, J. Chen, Incomplete variables truncated conjugate gradient method for signal reconstruction in compressed sensing, *Information Sciences* 288 (2014) 387–411.
- 680 [40] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE transactions on image processing* 13 (4) (2004) 600–612.
- [41] K. Wei, Fast iterative hard thresholding for compressed sensing, *IEEE Signal Processing Letters* 22 (5) (2015) 593–597.
- [42] S. J. Wright, R. D. Nowak, M. A. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Transactions on Signal Processing* 57 (7) (2009) 2479–2493.
- 690 [43] Z. Xu, X. Chang, F. Xu, H. Zhang, $l_{1/2}$ regularization: A thresholding representation theory and a fast solver, *IEEE Transactions on neural networks and learning systems* 23 (7) (2012) 1013–1027.
- [44] B. Yan, Q. Zhao, Z. Wang, X. Zhao, A hybrid evolutionary algorithm for multiobjective sparse reconstruction, *Signal, Image and Video Processing* (2017) 1–8.
- 695 [45] G. You, Z.-H. Huang, Y. Wang, A theoretical perspective of solving phaseless compressed sensing via its nonconvex relaxation, arXiv preprint arXiv:1702.00110.
- [46] Y. Yuan, H. Xu, B. Wang, X. Yao, A new dominance relation-based evolutionary algorithm for many-objective optimization, *IEEE Transactions on Evolutionary Computation* 20 (1) (2016) 16–37.
- 700

- [47] Y. Yuan, H. Xu, B. Wang, B. Zhang, X. Yao, Balancing convergence and diversity in decomposition-based many-objective optimizers, *IEEE Transactions on Evolutionary Computation* 20 (2) (2016) 180–198.
- 705 [48] J. Zeng, S. Lin, Y. Wang, Z. Xu, $l_{1/2}$ regularization: Convergence of iterative half thresholding algorithm, *IEEE Transactions on Signal Processing* 62 (9) (2014) 2317–2329.
- [49] Q. Zhang, H. Li, Moea/d: A multiobjective evolutionary algorithm based on decomposition, *IEEE Transactions on evolutionary computation* 11 (6)
710 (2007) 712–731.
- [50] S. Zhang, H. Yao, X. Sun, K. Wang, J. Zhang, X. Lu, Y. Zhang, Action recognition based on overcomplete independent components analysis, *Information Sciences An International Journal* 281 (2014) 635–647.