Superposed Atomic Representation for Robust High-Dimensional Data Recovery of Multiple Low-Dimensional Structures

Yulong Wang[1,2,3]*

[1] College of Informatics, Huazhong Agricultural University, China
[2] Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, China
[3] Key Laboratory of Smart Farming Technology for Agricultural Animals, Ministry of Agriculture and Rural Affairs, China
wangyulong6251@gmail.com

Abstract

This paper proposes a unified Superposed Atomic Representation (SAR) framework for high-dimensional data recovery with multiple low-dimensional structures. The data can be in various forms ranging from vectors to tensors. The goal of SAR is to recover different components from their sum, where each component has a low-dimensional structure, such as sparsity, low-rankness or be lying a low-dimensional subspace. Examples of SAR include, but not limited to, Robust Sparse Representation (RSR), Robust Principal Component Analysis (RPCA), Tensor RPCA (TRPCA), and Outlier Pursuit (OP). We establish the theoretical guarantee for SAR. To further improve SAR, we also develop a Weighted SAR (WSAR) framework by paying more attention and penalizing less on significant atoms of each component. An effective optimization algorithm is devised for WSAR and the convergence of the algorithm is rigorously proved. By leveraging WSAR as a general platform, several new methods are proposed for high-dimensional data recovery. The experiments on real data demonstrate the superiority of WSAR for various data recovery problems.

Introduction

In many real-world applications, high-dimensional data is often a superposition of multiple components with different low-dimensional structures (Candès et al. 2011; Wright et al. 2009). For instance, a grayscale video data with moving objects can be regarded as a superposition of the background part and the foreground part containing moving objects. Since the frames of the background are highly correlated in general, it can be modeled as a low-rank tensor. While the foreground part can be modeled as a sparse tensor since the moving objects often occupy a small part of each video frame (Lu et al. 2020; Liu et al. 2023). Analogously, a face image with disguise (e.g., sunglasses) can be regarded as a superposition of a noiseless face image (lying in a low-dimensional subspace (Wright et al. 2009)) and the occlusion part (deemed as a sparse matrix). To recover the underlying components from the superposed data, a variety of data recovery methods have been proposed for distinct problems and applications (Chen, Wu, and Wang 2019; Dutta, Hanzely, and Richtárik 2019; Zheng et al. 2022; Wright and Ma 2022). We refer to such general high-dimensional data recovery problem as Multiple Low-Dimensional Structure Recovery (MLDSR).

RPCA and Tensor RPCA (TRPCA). Robust Principal Component Analysis (RPCA) (Candès et al. 2011) and its variants (Ding, He, and Carin 2011) belong to MLDSR and assume that data is a superposition of a low-rank matrix and a sparse matrix and aim to recover them from their sum. Since many real-data are often stored in tensors such as color images and hyperspectral images, RPCA is extended for tensor data and the problem changes to recovering a low-rank tensor and a sparse tensor from their sum. The resulting method is dubbed Tensor RPCA (TRPCA) (Lu et al. 2020). RPCA, TRPCA and their variants (Qiu et al. 2022) have a wide range of applications, such as image recovery and background substraction (Chang et al. 2020).

Robust Sparse Representation (RSR). As another example of MLDSR, RSR (Wright et al. 2009) assumes that the observed corrupted data (e.g., a face image with sunglasses) is a superposition of the noiseless data lying in a low-dimensional linear subspace and a sparse noise part. RSR methods have also achieved great success in various applications such as robust face recognition (Zhang et al. 2022), motion segmentation (Elhamifar and Vidal 2013) and hyperspectral image classification (Chen, Nasrabadi, and Tran 2011; Peng, Sun, and Du 2019).

Outlier Pursuit (OP). The OP problem (Xu, Caramanis, and Sanghavi 2012; Zhou and Feng 2017) also belongs to MLDSR and aims to identify the outliers from a mixture of inliers and outliers. The data is modeled as a superposition of a low-rank matrix (nonzero columns correspond to inliers lying in a low-dimensional subspace) and a column-sparse matrix (nonzero columns correspond to outliers out of the subspace). More examples of MLDSR include the line spectral estimation problem (Yang and Xie 2015).

The problems above have the commonality that all of them leverage the underlying multiple low-dimensional structures of high-dimensional data to recover the original data. *However, in the literature these problems above are tackled on a case-by-case basis.* These limitations of existing works motivate us to consider an interesting question: is it possible to propose a unified framework for the general MLDSR problem? This work gives a positive answer.

*Corresponding author.

Paper Contributions

The contributions of this work are summarized as follows.

1. We propose a unified framework dubbed SAR (Superposed Atomic Representation) to recover the high-dimensional data of multiple low-dimensional structures. Examples of SAR include but not limited to the RPCA, tensor RPCA (TRPCA), RSR and OP models.
2. To enhance SAR, we develop a Weighted SAR (WSAR) framework by taking into account the significance of different atoms in each atomic norm regularization. By leveraging WSAR as a general platform, several new data recovery methods are proposed for various applications.
3. We develop a unified optimization algorithm for the WSAR framework based on ADMM (Boyd et al. 2011). In addition, we have also rigorously prove the convergence of the proposed algorithm.

Superposed Atomic Representation

Atomic Norm

We first introduce the definition of Atomic Norm (AN).

Definition 1. *(Chandrasekaran et al. 2012) Given an atomic set $\mathcal{A}$, the atomic norm of $\mathbf{x}$ with respect to $\mathcal{A}$ is defined as*

$$\|\mathbf{x}\|_{\mathcal{A}} := \inf\{t > 0 : \mathbf{x} \in t \cdot conv(\mathcal{A})\}, \tag{1}$$

where $conv(\mathcal{A})$ is the convex hull of the set $\mathcal{A}$.

If $\mathcal{A}$ is centrally symmetric about the origin, the gauge function $\|\cdot\|_{\mathcal{A}}$ is a norm. In this case, we have

$$\|\mathbf{x}\|_{\mathcal{A}} := \inf\left\{\sum_{\mathbf{a}\in\mathcal{A}} c_{\mathbf{a}} : \mathbf{x} = \sum_{\mathbf{a}\in\mathcal{A}} c_{\mathbf{a}}\mathbf{a}, c_{\mathbf{a}} \geq 0, \forall \mathbf{a} \in \mathcal{A}\right\}. \tag{2}$$

The dual norm of $\|\cdot\|_{\mathcal{A}}$ is defined by

$$\|\mathbf{x}\|^*_{\mathcal{A}} := \sup\{\langle \mathbf{x}, \mathbf{a}\rangle, \mathbf{a} \in \mathcal{A}\}. \tag{3}$$

Many common norms such as the ℓ_1 norm and the nuclear norm are special cases of AN (Chandrasekaran et al. 2012).

Superposed Atomic Representation (SAR)

Consider two signals $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ with different low-dimensional structures, respectively. They can be vectors, matrices or tensors. Given only the superposition $\mathbf{x} = \mathbf{x}_1^* + \mathbf{x}_2^*$, the goal is to recover $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$. To this end, we propose the following SAR framework

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \|\mathbf{x}_1\|_{\mathcal{A}_1} + \lambda\|\mathbf{x}_2\|_{\mathcal{A}_2} \ \text{ s.t. } \ \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \tag{4}$$

where λ is a positive parameter balancing the two atomic norms.

Remark 1. *It is worth remarking that the proposed SAR framework can be naturally extended for more than two atomic norm regularization terms corresponding to multiple low-dimensional structures. The same goes for the proposed Weighted SAR (WSAR) framework and the optimization algorithm in the next section. For simplicity, in this work we only consider the case of two atomic norm regularization. It is also worth noting that the atomic representation model in the prior work (Wang et al. 2019) only considers* single special *low-dimensional structure and assumes that high-dimensional data lie in a union of low-dimensional subspaces.*

Model	Data form	$\mathcal{A}_1$	$\mathcal{A}_2$
RSR	vector	$\mathcal{A}_{\mathbf{D}}$	$\mathcal{A}_S$
AND	vector	$\mathcal{A}_F$	$\mathcal{A}_S$
RPCA	matrix	$\mathcal{A}_L$	$\mathcal{A}_{MS}$
OP	matrix	$\mathcal{A}_L$	$\mathcal{A}_J$
TRPCA	tensor	$\mathcal{A}_{TL}$	$\mathcal{A}_{TS}$
OR-TPCA	tensor	$\mathcal{A}_{TL}$	$\mathcal{A}_{LS}$

Table 1: Many high-dimensional data recovery methods are special cases of the unified SAR framework. The definitions of the atomic sets can be found in the context.

By specifying the atomic sets $\mathcal{A}_1$ and $\mathcal{A}_2$, a variety of existing popular data recovery models belong to SAR as special cases. Some of them are listed as below.

Example 1. (Robust Sparse Representation (RSR)) (Wright et al. 2009) The objective function of RSR is formulated as

$$\min_{\mathbf{c}\in\mathbb{R}^n,\mathbf{e}\in\mathbb{R}^m} \|\mathbf{c}\|_1 + \lambda\|\mathbf{e}\|_1 \ \text{ s.t. } \ \mathbf{y} = \mathbf{D}\mathbf{c} + \mathbf{e}, \tag{5}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector, $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_n] \in \mathbb{R}^{m\times n}$ denotes the dictionary, $\mathbf{c} \in \mathbb{R}^n$ is a sparse vector and $\mathbf{e} \in \mathbb{R}^m$ denotes the error term. Define $\mathcal{A}_{\mathbf{D}} = \{\pm\mathbf{d}_1, \cdots, \pm\mathbf{d}_n\}$ and $\mathcal{A}_S$ as the set of unit-Euclidean norm one-sparse vectors in $\mathbb{R}^n$. Let $\mathbf{x}_1 = \mathbf{D}\mathbf{c}$ and $\mathbf{x}_2 = \mathbf{e}$. Then we have $\|\mathbf{x}_1\|_{\mathcal{A}_{\mathbf{D}}} = \min_{\{\mathbf{c}:\mathbf{x}_1=\mathbf{D}\mathbf{c}\}} \|\mathbf{c}\|_1$ and $\|\mathbf{x}_2\|_{\mathcal{A}_S} = \|\mathbf{e}\|_1$. In this case, SAR (4) reduces to the RSR model with $\mathcal{A}_1 = \mathcal{A}_{\mathbf{D}}$ and $\mathcal{A}_2 = \mathcal{A}_S$.

Example 2. (Robust Principal Component Analysis (RPCA)) (Candès et al. 2011) RPCA aims to recover a low-rank matrix and a sparse matrix from their sum by solving the following problem

$$\min_{\mathbf{L},\mathbf{S}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \ \text{ s.t. } \ \mathbf{X} = \mathbf{L} + \mathbf{S}, \tag{6}$$

where $\|\mathbf{L}\|_*$ and $\|\mathbf{S}\|_1$ denote the nuclear norm and ℓ_1 norm of matrices, respectively. Analogously, we can define the low rankness inducing atomic set $\mathcal{A}_L := \{\mathbf{M} \in \mathbb{R}^{n_1\times n_2} | \ \text{rank}(\mathbf{M}) = 1, \|\mathbf{M}\|_F = 1\}$, and let $\mathcal{A}_{MS}$ be the set of unit-Frobenius norm one-sparse matrices in $\mathbb{R}^{n_1\times n_2}$. It can be verified that $\|\mathbf{M}\|_{\mathcal{A}_{MS}} = \|\mathbf{M}\|_1$. Thus, RPCA also belongs to SAR with $\mathcal{A}_1 = \mathcal{A}_L$ and $\mathcal{A}_2 = \mathcal{A}_{MS}$.

Example 3. (Outlier Pursuit (OP)) (Xu, Caramanis, and Sanghavi 2012) Given a set of liners and outliers, OP aims to learn the true low-dimensional subspace and identify the outliers simultaneously by solving the problem

$$\min_{\mathbf{L},\mathbf{S}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_{2,1} \ \text{ s.t. } \ \mathbf{X} = \mathbf{L} + \mathbf{S}, \tag{7}$$

where $\|\mathbf{S}\|_{2,1} = \sum_{j=1}^{n_2} \|\mathbf{S}(:, j)\|_2$. Define the atomic set

$$\mathcal{A}_J := \bigcup_{j=1}^{n} \{\mathbf{M} \in \mathbb{R}^{n_1\times n_2} | \ \|\mathbf{M}(:, j)\|_2 = 1, \mathbf{M}(:, j') = \mathbf{0}, j' \neq j\},$$

Then OP also belongs to SAR with $\mathcal{A}_1 = \mathcal{A}_L$ and $\mathcal{A}_2 = \mathcal{A}_J$.

Example 4. (Tensor RPCA (TRPCA)) (Lu et al. 2020) TRPCA aims to recover a low-rank tensor and a sparse tensor from their sum by solving the following problem

$$\min_{\mathcal{L},\mathcal{S}\in\mathbb{R}^{n_1\times n_2\times n_3}} \|\mathcal{L}\|_{\circledast} + \lambda\|\mathcal{S}\|_1 \ \text{ s.t. } \ \mathcal{X} = \mathcal{L} + \mathcal{S}, \tag{8}$$

where $\|\mathcal{L}\|_{\circledast}$ denotes the tensor nuclear norm (Kilmer et al. 2013; Lu et al. 2020) and $\|\mathcal{S}\|_1 = \sum_{ijk}|\mathcal{S}_{ijk}|$ is the tensor ℓ_1 norm. Analogously, we can define the low tubal-rankness inducing atomic set (Lu et al. 2018) $\mathcal{A}_{TL} := \left\{\mathcal{W} \in \mathbb{C}^{n_1\times n_2\times n_3},\ \mathcal{W} = n_3\mathcal{D}, \bar{\mathbf{D}} \in \mathcal{B}\right\}$, where $\bar{\mathbf{D}} \in \mathbb{C}^{n_1 n_3\times n_2 n_3}$ is a block diagonal matrix with the i-th block on the diagonal as the frontal slice $\bar{\mathbf{D}}^{(i)}$ of $\bar{\mathcal{D}}$. Here $\bar{\mathcal{D}}$ denotes the Fast Fourier Transformation (FFT) of $\mathcal{D}$ along the 3-rd dimension, and $\mathcal{B}$ denotes the set of the following matrices

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & & \\ & \ddots & \\ & & \mathbf{D}_{n_3} \end{bmatrix} \in \mathbb{C}^{n_1 n_3\times n_2 n_3},$$

where $\mathbf{D}_i \in \mathbb{C}^{n_1\times n_2}$ and there exists some k such that $\text{rank}(\mathbf{D}_k) = 1$, $\|\mathbf{D}_k\|_F = 1$ and $\mathbf{D}_j = \mathbf{0}$, $\forall j \neq k$. It can proved that $\|\mathcal{T}\|_{\mathcal{A}_{TL}} = \|\mathcal{T}\|_{\circledast}$ (Lu et al. 2018), which is the tensor nuclear norm. Let $\mathcal{A}_{TS}$ be the set of unit-Frobenius norm one-sparse tensors in $\mathbb{R}^{n_1\times n_2\times n_3}$. It can be verified that $\|\mathcal{T}\|_{\mathcal{A}_{TS}} = \|\mathcal{T}\|_1$ for any tensor $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$. Thus, TRPCA also belongs to SAR with $\mathcal{A}_1 = \mathcal{A}_{TL}$ and $\mathcal{A}_2 = \mathcal{A}_{TS}$.

More examples of SAR are summarized in Table 1. In the AND (Atomic Norm Denoising) model (Yang and Xie 2015) for line spectral estimation, the signal frequency related atomic set is defined by $\mathcal{A}_F := \left\{\left[e^{i\phi}, \cdots, e^{i(2\pi(n-1))t+\phi}\right]^T, t \in [0,1], \phi \in [0, 2\pi]\right\}$. For OR-TPCA (Zhou and Feng 2017), the lateral slice sparsity-inducing atomic set is defined as $\mathcal{A}_{LS} := \bigcup_{j=1}^{n_2}\{\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3} |\ \|\mathcal{T}(:,j,:)\|_F = 1, \mathcal{T}(:,j',:) = \mathbf{0}, j' \neq j\}$.

Recovery Guarantee

Theorem 1. *Let $\mathbf{x} = \mathbf{x}_1^* + \mathbf{x}_2^*$. Assume that for any $\mathbf{h} \neq \mathbf{0}$, there exists two subgradients $\mathbf{s}_1 \in \partial\|\mathbf{x}_1^*\|_{\mathcal{A}_1}$ and $\mathbf{s}_2 \in \partial\|\mathbf{x}_2^*\|_{\mathcal{A}_2}$ such that*

$$\langle \mathbf{h}, \mathbf{s}_1 - \lambda\mathbf{s}_2\rangle > 0.$$

Then $(\mathbf{x}_1^*, \mathbf{x}_2^*)$ is the unique minimizer of the SAR model.

Proof. Any feasible solution to the SAR model (4) can be rewritten as $(\mathbf{x}_1^* + \mathbf{h}, \mathbf{x}_2^* - \mathbf{h})$ for some $\mathbf{h}$. In what follows, we prove that for any $\mathbf{h} \neq \mathbf{0}$, $(\mathbf{x}_1^* + \mathbf{h}, \mathbf{x}_2^* - \mathbf{h})$ will have larger objective value than $(\mathbf{x}_1^*, \mathbf{x}_2^*)$. For any $\mathbf{h} \neq \mathbf{0}$ and any subgradient $\mathbf{s}_1$ of the function $\|\cdot\|_{\mathcal{A}_1}$, we have

$$\|\mathbf{x}_1^* + \mathbf{h}\|_{\mathcal{A}_1} \geq \|\mathbf{x}_1^*\|_{\mathcal{A}_1} + \langle \mathbf{h}, \mathbf{s}_1\rangle \tag{9}$$

Analogously, for any subgradient $\mathbf{s}_2$ of the function $\|\cdot\|_{\mathcal{A}_2}$, there holds

$$\|\mathbf{x}_2^* - \mathbf{h}\|_{\mathcal{A}_2} \geq \|\mathbf{x}_2^*\|_{\mathcal{A}_2} - \langle \mathbf{h}, \mathbf{s}_2\rangle \tag{10}$$

Combining Eqs. (9) and (10), for any $\mathbf{h} \neq \mathbf{0}$, we have

$$\begin{aligned} &\|\mathbf{x}_1^* + \mathbf{h}\|_{\mathcal{A}_1} + \lambda\|\mathbf{x}_2^* - \mathbf{h}\|_{\mathcal{A}_2} \\ \geq &\|\mathbf{x}_1^*\|_{\mathcal{A}_1} + \lambda\|\mathbf{x}_2^*\|_{\mathcal{A}_2} + \langle \mathbf{h}, \mathbf{s}_1\rangle - \lambda\langle \mathbf{h}, \mathbf{s}_2\rangle \\ = &\|\mathbf{x}_1^*\|_{\mathcal{A}_1} + \lambda\|\mathbf{x}_2^*\|_{\mathcal{A}_2} + \langle \mathbf{h}, \mathbf{s}_1 - \lambda\mathbf{s}_2\rangle \\ > &\|\mathbf{x}_1^*\|_{\mathcal{A}_1} + \lambda\|\mathbf{x}_2^*\|_{\mathcal{A}_2}. \end{aligned} \tag{11}$$

The last inequality comes from the theorem assumption. □

Weighted Superposed Atomic Representation

Note from the definition of atomic norm in Eq. (2) that it penalizes the atoms with large coefficients more heavily than those with smaller coefficients. Such imbalance may lead to suboptimal performance because atoms with large coefficients often correspond to significant components of data (Candès, Wakin, and Boyd 2008; Gu et al. 2017). To tackle this imbalance and derive more democratic method, in this section we propose the Weighted Atomic Norm (WAN). Based on WAN, we propose the Weighted SAR (WSAR) framework to further improve SAR by paying more attention and penalizing less on significant atoms.

Model

We define the weighted atomic norm (WAN) as

$$\|\mathbf{x}\|_{\mathbf{w},\mathcal{A}} := \inf\left\{\sum_{\mathbf{a}\in\mathcal{A}} w_{\mathbf{a}}c_{\mathbf{a}} : \mathbf{x} = \sum_{\mathbf{a}\in\mathcal{A}} c_{\mathbf{a}}\mathbf{a}, c_{\mathbf{a}} \geq 0, \forall \mathbf{a} \in \mathcal{A}\right\},$$

where $\mathbf{w}$ denotes the weight vector and its entry $w_{\mathbf{a}} \geq 0$ for any $\mathbf{a} \in \mathcal{A}$.

Remark 2. *If $w_{\mathbf{a}} = 1$, $\forall\ \mathbf{a} \in \mathcal{A}$, WAN reduces to AN. Note that since the atomic set $\mathcal{A}$ can be infinite (e.g., $\mathcal{A}_L$), the dimension of the weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ can also be infinite.*

Incorporating WAN into SAR, we have the Weighted Superposed Atomic Representation (WSAR) framework as

$$\min_{\mathbf{x}_1,\mathbf{x}_2} \|\mathbf{x}_1\|_{\mathbf{w}_1,\mathcal{A}_1} + \lambda\|\mathbf{x}_2\|_{\mathbf{w}_2,\mathcal{A}_2} \ \text{ s.t. } \ \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2. \tag{12}$$

By specifying different atomic sets, we can get new data recovery methods for different problems. Concretely, we refer to the specific methods in WSAR with $\mathcal{A}_1 = \mathcal{A}_{\mathbf{D}}$ and $\mathcal{A}_2 = \mathcal{A}_S$ as WSAR_RSR, with $\mathcal{A}_1 = \mathcal{A}_L$ and $\mathcal{A}_2 = \mathcal{A}_{MS}$ as WSAR_RPCA, and with $\mathcal{A}_1 = \mathcal{A}_{TL}$ and $\mathcal{A}_2 = \mathcal{A}_{TS}$ as WSAR_TRPCA, respectively.

Remark 3. *In fact, there some some prior works which consider the weights for different sparse entries in sparse recovery or singular values in low-rank matrix or tensor recovery. They include Weighted Sparse Representation (WSR) (Candès, Wakin, and Boyd 2008), Weighted Nuclear Norm Minimization (WNNM) (Gu et al. 2017), Enhanced TRPCA (ETRPCA) (Gao et al. 2021). However, there are two key differences between WSAR and these prior works. Firstly, WSAR is a* unified *data recovery framework while they are only designed for a* specific *recovery problem case by case. Secondly, WSAR considers the weights for atoms associated with* multiple *low-dimensional structures while they only consider the weights for* single *low-dimensional structure such as sparse entries or singular values.*

Optimization

The Lagrangian function of the WSAR problem (12) is

$$\begin{aligned} L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z}, \mu) = &\|\mathbf{x}_1\|_{\mathbf{w}_1,\mathcal{A}_1} + \lambda\|\mathbf{x}_2\|_{\mathbf{w}_2,\mathcal{A}_2} \\ &+ \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_1 - \mathbf{x}_2 + \mathbf{z}/\mu\|_F^2 - \frac{\mu}{2}\|\mathbf{z}/\mu\|_F^2, \end{aligned} \tag{13}$$

where $\mathbf{z}$ is the Lagrangian multiplier and μ is a positive parameter. Given the $\mathbf{x}_1^{(t)}$, $\mathbf{x}_2^{(t)}$, $\mathbf{z}^{(t)}$, and $\mu^{(t)}$ in the t-th iteration, we can update them alternatively by fixing others.

Algorithm 1: Weighted Superposed Atomic Representation

Input: Data $\mathbf{x}$, and the parameter λ.
Initialization: $\mathbf{x}_1^{(0)} = \mathbf{x}_2^{(0)} = \mathbf{0}$, $\mathbf{z}^{(0)} = \mathbf{0}$, $\mu^{(0)} = 10^{-2}$, $\rho = 1.1$, $\epsilon = 10^{-6}$, iteration counter $t = 0$.
Output: $\mathbf{x}_1^{(t)}$, $\mathbf{x}_2^{(t)}$.

1: **while** not converged **do**
2: Update $\mathbf{x}_1^{(t+1)}$ by Eq. (15).
3: Update $\mathbf{x}_2^{(t+1)}$ by Eq. (17).
4: Update the multiplier $\mathbf{z}^{(t+1)}$ by Eq. (18).
5: Update the parameter $\mu^{(t+1)}$ by $\mu^{(t+1)} = \rho\mu^{(t)}$.
6: Check the convergence condition:

$$\left\|\mathbf{x}_1^{(t+1)} - \mathbf{x}_1^{(t)}\right\|_\infty < \epsilon, \quad \left\|\mathbf{x}_2^{(t+1)} - \mathbf{x}_2^{(t)}\right\|_\infty < \epsilon,$$

$$\left\|\mathbf{x} - \mathbf{x}_1^{(t+1)} - \mathbf{x}_2^{(t+1)}\right\|_\infty < \epsilon,$$

7: **end while**

Step 1: Update $\mathbf{x}_1$ by

$$\min_{\mathbf{x}_1} \frac{1}{2}\left\|\mathbf{x}_1 - \left(\mathbf{x} - \mathbf{x}_2^{(t)} + \mathbf{z}^{(t)}/\mu^{(t)}\right)\right\|_F^2 + \frac{1}{\mu^{(t)}}\|\mathbf{x}_1\|_{\mathbf{w}_1,\mathcal{A}_1}. \quad (14)$$

The solution is

$$\mathbf{x}_1^{(t+1)} = \Pi_{\mathbf{w}_1,\mathcal{A}_1}\left(\mathbf{u}_1^{(t)}, 1/\mu^{(t)}\right), \quad (15)$$

where $\mathbf{u}_1^{(t)} = \mathbf{x} - \mathbf{x}_2^{(t)} - \mathbf{z}^{(t)}/\mu^{(t)}$. Here $\Pi_{\mathbf{w},\mathcal{A}}(\mathbf{u},\cdot) = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{u}\|_F^2 + \lambda\|\mathbf{x}\|_{\mathbf{w},\mathcal{A}}$ denotes the proximity operator with respect to $\mathcal{A}$ and $\mathbf{w}$. Several examples of $\Pi_{\mathbf{w},\mathcal{A}}(\cdot)$ are listed as follows.

- $\Pi_{\mathbf{w},\mathcal{A}_S}(\mathbf{v},\gamma) = S(\mathbf{v},\gamma\mathbf{w}) = \text{sign}(\mathbf{v}) \otimes (|\mathbf{v}| - \gamma\mathbf{w})_+$, $\mathbf{v} \in \mathbb{R}^n$,
- $\Pi_{\mathbf{w},\mathcal{A}_L}(\mathbf{M},\gamma) = \mathbf{U}\left(\Sigma - \gamma\text{diag}(\mathbf{w_M})\right)_+ \mathbf{V}^T$, $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$,
- $\Pi_{\mathbf{w},\mathcal{A}_J}(\mathbf{M},\gamma)(:,j) = \frac{\left(\|\mathbf{M}(:,j)\|_2 - \gamma w_j\right)_+}{\|\mathbf{M}(:,j)\|_2}\mathbf{M}(:,j)$, $j = 1,\cdots,n_2$,
- $\Pi_{\mathbf{w},\mathcal{A}_{TS}}(\mathcal{T},\gamma) = \text{sign}(\mathcal{T}) \otimes (|\mathcal{T}| - \gamma\mathbf{w})_+$, $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$.

Here $\text{sign}(\mathbf{z})$ denotes a vector of the sign of entries of $\mathbf{z}$ and $\otimes$ represents the Hadamard product or the entrywise product. $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ otherwise. $\mathbf{U}\Sigma\mathbf{V}^T$ is the SVD decomposition of a matrix $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$ where $\mathbf{U} \in \mathbb{R}^{n_1\times\bar{n}}$, $\Sigma \in \mathbb{R}^{\bar{n}\times\bar{n}}$, $\mathbf{V} \in \mathbb{R}^{n_2\times\bar{n}}$, and $\bar{n} = \min\{n_1, n_2\}$. Here $\mathbf{w_M} \in \mathbb{R}^{\bar{n}}$ denotes the subvector of $\mathbf{w}$ containing its entries corresponding to the atoms $\{\mathbf{u}_i\mathbf{v}_i^T, i = 1,\cdots,\bar{n}\}$ and $\text{diag}(\mathbf{w_M})$ denotes a diagonal matrix with the entries of $\mathbf{w_M}$ on its diagonal. The closed-form formula of $\Pi_{\mathbf{w},\mathcal{A}_{TL}}(\mathcal{T},\gamma)$ is analogous to the matrix case and can be found in (Gao et al. 2021).

Step 2: Update $\mathbf{x}_2$ by

$$\min_{\mathbf{x}_2} \frac{1}{2}\left\|\mathbf{x}_2 - \left(\mathbf{x} - \mathbf{x}_1^{(t+1)} + \mathbf{z}^{(t)}/\mu^{(t)}\right)\right\|_F^2 + \frac{\lambda}{\mu^{(t)}}\|\mathbf{x}_2\|_{\mathbf{w}_2,\mathcal{A}_2}. \quad (16)$$

Analogously, the solution can be expressed as

$$\mathbf{x}_2^{(t+1)} = \Pi_{\mathbf{w}_2,\mathcal{A}_2}\left(\mathbf{u}_2^{(t)}, \lambda/\mu^{(t)}\right), \quad (17)$$

where $\mathbf{u}_2^{(t)} = \mathbf{x} - \mathbf{x}_1^{(t+1)} - \mathbf{z}^{(t)}/\mu^{(t)}$.

Step 3: Update $\mathbf{z}$ by

$$\mathbf{z}^{(t+1)} = \mathbf{z}^{(t)} + \mu^{(t)}\left(\mathbf{x} - \mathbf{x}_1^{(t+1)} - \mathbf{x}_2^{(t+1)}\right), \quad (18)$$

Step 4: Update the parameter μ by

$$\mu^{(t+1)} = \rho\mu^{(t)},$$

where $\rho = 1.1$. The complete iterative procedure is summarized in Algorithm 1.

Convergence Analysis

This part is concerned with establishing the convergence analysis of Algorithm 1. To this end, we first introduce a necessary assumption and several useful lemmas.

Assumption 1. *For the atomic set $\mathcal{A}$ and weight vector $\mathbf{w}$, there exists a constant C such that the following inequality holds*

$$\left\|\Pi_{\mathbf{w},\mathcal{A}}(\mathbf{u},\lambda) - \mathbf{u}\right\|_F \leq \lambda C \max_{\mathbf{a}\in\mathcal{A}} \mathbf{w}_\mathbf{a}, \quad \forall\mathbf{u}. \quad (19)$$

In fact, it can be verified many common atomic sets including $\mathcal{A}_S$, $\mathcal{A}_L$, $\mathcal{A}_J$, and $\mathcal{A}_{TL}$ satisfy the assumption above.

Lemma 1. *The sequence $\left\{\mathbf{z}^{(t)}\right\}_{t=1}^\infty$ generated by Algorithm 1 are bounded.*

Proof. In light of the update rule of the Lagrangian multiplier $\boldsymbol{\mathcal{Y}}$ in Step 4 of Algorithm 1, we have

$$\begin{aligned}\left\|\mathbf{z}^{(t+1)}\right\|_F &= \left\|\mathbf{z}^{(t)} + \mu^{(t)}\left(\mathbf{x}_1^{(t+1)} + \mathbf{x}_2^{(t+1)} - \mathbf{x}\right)\right\|_F \\ &= \mu^{(t)}\left\|\mathbf{x}_2^{(t+1)} - \left(\mathbf{x} - \mathbf{x}_1^{(t+1)} - \mathbf{z}^{(t)}/\mu^{(t)}\right)\right\|_F \\ &= \mu^{(t)}\left\|\mathbf{x}_2^{(t+1)} - \mathbf{u}_2^{(t)}\right\|_F \\ &= \mu^{(t)}\left\|\Pi_{\mathbf{w}_2,\mathcal{A}_2}\left(\mathbf{u}_2^{(t)}, \lambda/\mu^{(t)}\right) - \mathbf{u}_2^{(t)}\right\|_F \\ &\leq \mu^{(t)}\frac{\lambda}{\mu^{(t)}}C_2M_2 = \lambda C_2M_2,\end{aligned} \quad (20)$$

for some constant C_2 and $M_2 = \max_{\mathbf{a}\in\mathcal{A}_2}\mathbf{w}_{2,\mathbf{a}}$. □

Lemma 2. *The sequences $\left\{\mathbf{x}_1^{(t)}\right\}_{t=1}^\infty$ and $\left\{\mathbf{x}_2^{(t)}\right\}_{t=1}^\infty$ generated by Algorithm 1 are bounded.*

Proof. The proof consists of the following two steps

1) **Step 1:** The sequence $\left\{L\left(\mathbf{x}_1^{(t+1)}, \mathbf{x}_2^{(t+1)}, \mathbf{z}^{(t)}, \mu^{(t)}\right)\right\}_{t=1}^\infty$ is bounded.
2) **Step 2:** The sequences $\left\{\mathbf{x}_1^{(t)}\right\}_{t=1}^\infty$ and $\left\{\mathbf{x}_2^{(t)}\right\}_{t=1}^\infty$ are bounded.

Proof of Step 1: Since $\mathbf{x}_1^{(t+1)}$ and $\mathbf{x}_2^{(t+1)}$ are optimal solutions to the corresponding subproblems, we have

$$L\left(\mathbf{x}_1^{(t+1)}, \mathbf{x}_2^{(t+1)}, \mathbf{z}^{(t)}, \mu^{(t)}\right) \leq L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t)}, \mu^{(t)}\right) \quad (21)$$

On the other hand, by the formulation of the Lagrangian function in Eq. (13) there holds

$$\begin{aligned}L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t)}, \mu^{(t)}\right) =& L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t-1)}, \mu^{(t-1)}\right) \\ &+ \left\langle\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x}, \mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\right\rangle \\ &+ \frac{\mu^{(t)} - \mu^{(t-1)}}{2}\left\|\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x}\right\|_F^2\end{aligned} \quad (22)$$

According to the update rule of the Lagrangian multiplier $\mathbf{z}^{(t)} = \mathbf{z}^{(t-1)} + \mu^{(t-1)}\left(\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x}\right)$, we have

$$\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x} = \frac{1}{\mu^{(t-1)}}\left(\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\right). \tag{23}$$

By instituting it into Eq. (22), we have

$$\begin{aligned} &L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t)}, \mu^{(t)}\right) \\ \leq &L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t-1)}, \mu^{(t-1)}\right) + \frac{\mu^{(t)} + \mu^{(t-1)}}{2(\mu^{(t-1)})^2}\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\|_F^2 \end{aligned} \tag{24}$$

According to Lemma 1, we know that the sequence $\{\|\mathbf{z}^{(t)}\|_F^2,\ t = 1, \cdots, \infty\}$ is upper bounded. Let B be the upper bound of $\left\{\left\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\right\|_F^2,\ t = 1, \cdots, \infty\right\}$. Combining Eqs. (21) and (24), we have

$$\begin{aligned} &L\left(\mathbf{x}_1^{(t+1)}, \mathbf{x}_2^{(t+1)}, \mathbf{z}^{(t)}, \mu^{(t)}\right) \\ \leq &L\left(\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \mathbf{z}^{(0)}, \mu^{(0)}\right) + B\sum_{t=1}^{\infty}\frac{\mu^{(t)} + \mu^{(t-1)}}{2(\mu^{(t-1)})^2} \\ \leq &L\left(\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \mathbf{z}^{(0)}, \mu^{(0)}\right) + B\sum_{t=1}^{\infty}\rho^{2-t} \\ = &L\left(\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \mathbf{z}^{(0)}, \mu^{(0)}\right) + \frac{B\rho^2}{\rho - 1} < \infty. \end{aligned} \tag{25}$$

Thus, the sequence $\left\{L\left(\mathbf{x}_1^{(t+1)}, \mathbf{x}_2^{(t+1)}, \mathbf{z}^{(t)}, \mu^{(t)}\right)\right\}_{t=1}^{\infty}$ is bounded and completes the proof of Step 1.

Proof of Step 2: Note that

$$\begin{aligned} &\left\|\mathbf{x}_1^{(t)}\right\|_{\mathbf{w}_1,\mathcal{A}_1} + \lambda\left\|\mathbf{x}_2^{(t)}\right\|_{\mathbf{w}_2,\mathcal{A}_2} \\ = &L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t-1)}, \mu^{(t-1)}\right) - \frac{\mu^{(t-1)}}{2}\left\|\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x} + \frac{\mathbf{z}^{(t-1)}}{\mu^{(t-1)}}\right\|_F^2 \\ &+ \frac{1}{2\mu^{(t-1)}}\|\mathbf{z}^{(t-1)}\|_F^2 \\ = &L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t-1)}, \mu^{(t-1)}\right) - \frac{\mu^{(t-1)}}{2}\left\|\frac{\mathbf{z}_t - \mathbf{z}^{(t-1)}}{\mu^{(t-1)}} + \frac{\mathbf{z}^{(t-1)}}{\mu^{(t-1)}}\right\|_F^2 \\ &+ \frac{1}{2\mu^{(t-1)}}\|\mathbf{z}^{(t-1)}\|_F^2 \\ = &L\left(\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t-1)}, \mu^{(t-1)}\right) - \frac{1}{2\mu^{(t-1)}}\|\mathbf{z}_t\|_F^2 + \frac{1}{2\mu^{(t-1)}}\|\mathbf{z}^{(t-1)}\|_F^2, \end{aligned} \tag{26}$$

where the third equation above follows from Eq. (23). Since the sequences $\left\{L\left(\mathbf{x}_1^{(t+1)}, \mathbf{x}_2^{(t+1)}, \mathbf{z}^{(t)}, \mu^{(t)}\right)\right\}_{t=1}^{\infty}$ and $\left\{\mathbf{z}^{(t)}\right\}_{t=1}^{\infty}$ are bounded, both $\left\{\mathbf{x}_1^{(t)}\right\}_{t=1}^{\infty}$ and $\left\{\mathbf{x}_2^{(t)}\right\}_{t=1}^{\infty}$ are bounded. □

With the results above, now we turn to proving the convergence of the proposed algorithm.

Theorem 2. *The sequences* $\left\{\mathbf{x}_1^{(t)}\right\}_{t=1}^{\infty}$ *and* $\left\{\mathbf{x}_2^{(t)}\right\}_{t=1}^{\infty}$ *generated by Algorithm 1 satisfy*

(1) $\lim_{t\to\infty}\left\|\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x}\right\|_F = 0$.

(2) $\lim_{t\to\infty}\left\|\mathbf{x}_1^{(t+1)} - \mathbf{x}_1^{(t)}\right\|_F = 0$.

(3) $\lim_{t\to\infty}\left\|\mathbf{x}_2^{(t+1)} - \mathbf{x}_2^{(t)}\right\|_F = 0$.

Proof. According to Lemmas 1 and 2, the sequences $\{\mathbf{z}^{(t)}\}_{t=1}^{\infty}$, $\{\mathbf{x}_1^{(t)}\}_{t=1}^{\infty}$ and $\{\mathbf{x}_2^{(t)}\}_{t=1}^{\infty}$ are bounded. There exists at least one accumulation point for $\{\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{z}^{(t)}\}_{t=1}^{\infty}$. Then we have

$$\lim_{t\to\infty}\left\|\mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} - \mathbf{x}\right\|_F = \lim_{t\to\infty}\frac{1}{\mu_{t-1}}\left\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\right\|_F = 0 \tag{27}$$

where Eq. (23) is used. Accordingly, the accumulation point is also feasible to the objective function.

Note from Eq. (23) that $\mathbf{x}_1^{(t)} = \mathbf{x} - \mathbf{x}_2^{(t)} + (\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)})/\mu^{(t-1)}$ and recall from Eq. (15) that $\mathbf{x}_1^{(t+1)} = \Pi_{\mathbf{w}_1,\mathcal{A}_1}\left(\mathbf{u}_2^{(t)}, 1/\mu^{(t)}\right)$ where $\mathbf{u}_2^{(t)} = \mathbf{x} - \mathbf{x}_2^{(t)} - \mathbf{z}^{(t)}/\mu^{(t)}$. According to Assumption 1, there exists a constant C_1 such that $\left\|\Pi_{\mathbf{w}_1,\mathcal{A}_1}(\mathbf{u}, \lambda) - \mathbf{u}\right\|_F \leq \lambda C_1 M_1,\ \forall\mathbf{u}$, where $M_1 = \max_{\mathbf{a}\in\mathcal{A}_1}\mathbf{w}_{1,\mathbf{a}}$. It follows that

$$\begin{aligned} &\left\|\mathbf{x}_1^{(t+1)} - \mathbf{x}_1^{(t)}\right\|_F \\ \leq &\left\|\Pi_{\mathbf{w}_1,\mathcal{A}_1}\left(\mathbf{u}_2^{(t)}, 1/\mu^{(t)}\right) - \mathbf{u}_2^{(t)}\right\|_F + \|\mathbf{u}_2^{(t)} - \mathbf{x}_1^{(t)}\|_F \\ \leq &\frac{C_1 M_1}{\mu^{(t)}} + \left\|(\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)})/\mu^{(t-1)} + \mathbf{z}^{(t)}/\mu^{(t)}\right\|_F \\ \leq &\frac{C_1 M_1}{\mu^{(t)}} + \frac{\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\|_F}{\mu^{(t-1)}} + \frac{\|\mathbf{z}^{(t)}\|_F}{\mu^{(t)}}. \end{aligned} \tag{28}$$

Since $\lim_{t\to\infty}\frac{1}{\mu^{(t)}} = 0$, and the sequence $\{\mathbf{z}^{(t)}\}_{t=1}^{\infty}$ is bounded, we have

$$\lim_{t\to\infty}\left\|\mathbf{x}_1^{(t+1)} - \mathbf{x}_1^{(t)}\right\|_F = 0. \tag{29}$$

Analogously, note from Eq. (23) that $\mathbf{x}_2^{(t+1)} = \mathbf{x} - \mathbf{x}_1^{(t)} + (\mathbf{z}^{(t)} - \mathbf{z}^{t-1})/\mu^{(t-1)}$, and recall from Eq. (15) that $\mathbf{x}_2^{(t+1)} = \Pi_{\mathbf{w}_2,\mathcal{A}_2}\left(\mathbf{u}_1^{(t+1)}, \frac{\lambda}{\mu^{(t)}}\right)$ where $\mathbf{u}_1^{(t+1)} = \mathbf{x} - \mathbf{x}_1^{(t+1)} + \mathbf{z}^{(t)}/\mu^{(t)}$. It follows that

$$\begin{aligned} &\|\mathbf{x}_2^{(t+1)} - \mathbf{x}_2^{(t)}\|_F = \|\mathbf{x}_2^{(t+1)} - \mathbf{u}_1^{(t+1)} \\ &+ \mathbf{x}_1^{(t)} - \mathbf{x}_1^{(t+1)} - \mathbf{z}^{(t)}/\mu^{(t)} - (\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)})/\mu^{(t-1)}\|_F \\ \leq &\left\|\Pi_{\mathbf{w}_2,\mathcal{A}_2}\left(\mathbf{u}_1^{(t+1)}, \frac{\lambda}{\mu^{(t)}}\right) - \mathbf{u}_1^{(t+1)}\right\|_F + \left\|\mathbf{x}_1^{(t)} - \mathbf{x}_1^{(t+1)}\right\|_F \\ &+ \left\|\frac{\mathbf{z}^{(t)}}{\mu^{(t)}} + \frac{\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}}{\mu^{(t-1)}}\right\|_F \\ \leq &\frac{\lambda C_2 M_2}{\mu^{(t)}} + \left\|\mathbf{x}_1^{(t)} - \mathbf{x}_1^{(t+1)}\right\|_F + \frac{\|\mathbf{z}^{(t)} - \mathbf{z}^{(t-1)}\|_F}{\mu^{(t-1)}} + \frac{\|\mathbf{z}^{(t)}\|_F}{\mu^{(t)}}. \end{aligned} \tag{30}$$

Combining Eq. (29) and $\lim_{t\to\infty}\frac{1}{\mu^{(t)}} = 0$, there holds

$$\lim_{t\to\infty}\left\|\mathbf{x}_2^{(t+1)} - \mathbf{x}_2^{(t)}\right\|_F = 0. \tag{31}$$

This completes the proof. □

The Setting of Weight Vectors

A key issue in WSAR is how to set the weight vectors $\mathbf{w}_i,\ i = 1, 2$. We expect to set them in an *adaptive and atomic* manner. This can be achieved by updating it iteratively based on the current estimate of $\mathbf{x}_1$ and $\mathbf{x}_2$. Assume that

$\mathbf{x}_i = \sum_{\mathbf{a}\in\mathcal{A}_i} c_{i,\mathbf{a}}\mathbf{a}$ for $i = 1, 2$. Since atoms with large coefficients often correspond to significant components of data, they should be less penalized. For this reason, we can set the weight $w_{i,\mathbf{a}}$ as a decreasing function of the coefficient $c_{i,\mathbf{a}}$ as

$$w_{i,\mathbf{a}} = f(c_{i,\mathbf{a}}) = \frac{1}{1 + c_{i,\mathbf{a}}/\tau}, \tag{32}$$

where τ is a constant and can be set in an adaptive way as (He et al. 2014). The value of the weight $w_{i,\mathbf{a}}$ measures the penalization strength of each atom. For the atom with $c_{i,\mathbf{a}} = 0$, the weight $w_{i,\mathbf{a}} = 1$. While for the atoms with large $c_{i,\mathbf{a}}$, the weights for the penalization strength is small. In the extreme case, when $c_{i,\mathbf{a}} \to +\infty$, $w_{i,\mathbf{a}} \to 0$. Therefore, the strategy (32) assigns lighter penalization on significant atoms.

Experiments

In this section, we explore the performance of special cases of WSAR in various applications. More experimental results can be found in the supplementary material due to space limitation.

Robust Face Restoration

In this part, we apply WSAR_RSR for robust face restoration. As mentioned before, a noisy face image can be regarded as a superposition of a noiseless face image (lying in a low-dimensional subspace (Wright et al. 2009)) and the noise part (deemed as a sparse vector). Consider a face image dataset of K subjects. Let $\mathbf{D}_k = \left[\mathbf{d}_1^k, \mathbf{d}_2^k, \cdots, \mathbf{d}_{n_k}^k\right] \in \mathbb{R}^{m\times n_k}$ denote a matrix composed of n_k vectorized facial image from the k-th subject for $k = 1, \cdots, K$. Denote by $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_K] \in \mathbb{R}^{m\times n}$ the dictionary of all training samples from the K subjects and $n = \sum_{k=1}^K n_k$. Given a possibly noisy facial image $\mathbf{y} \in \mathbb{R}^m$, we first compute the representation vector $\hat{\mathbf{x}}$ by RSR or WSAR_RSR. Then $\hat{\mathbf{y}} = \mathbf{D}\hat{\mathbf{x}}$ is regarded as the restored vectorized image of $\mathbf{y}$.

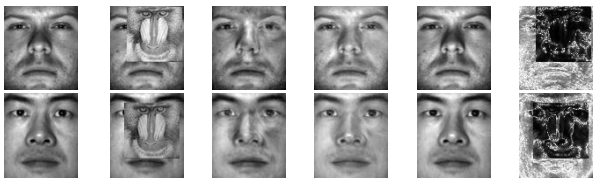
Datasets: We consider the widely used Extended Yale B (EYaleB) face image database (Lee, Ho, and Driegman 2005). It contains over 2400 frontal face images of 38 subjects taken under varying illumination. To form the dictionary $\mathbf{D}$, we randomly select half of images (32 images) per subject for training. In the experiment, for each test image a random square region is occluded by a baboon image.

Comparison methods: We compare WSAR_RSR with RSR (Wright et al. 2009) and EGSNR (Zhang et al. 2022) using the parameter recommended by the authors.

Results: Fig. 1 shows the original image and the restoration results on the two datasets by different methods. Note that the restored images of WSAR_RSR have better visual quality and less artifacts than RSR.

Robust Image Recovery

In this subsection, we evaluate the performance of WSAR_TRPCA for robust color image recovery in the presence of random noise. This is motivated by the fact that the natural color images can be approximated by low rank matrices or tensors while random noise can be regarded as sparse matrices or tensors (Lu et al. 2020).



(a) (b) (c) (d) (e) (f)

Figure 1: Face restoration of two randomly selected images with 50% occlusion on the EYaleB database. (a) Original images; (b) noisy images; (c) RSR; (d) EGSNR; (e) WSAR_RSR; (f) weight learned by WSAR_RSR (the darker means smaller value).

Image	1	2	3	4	5
RPCA	36.94	27.48	23.41	23.95	26.32
SNN	39.18	29.23	25.23	26.29	27.44
TRPCA	40.50	31.08	25.45	27.34	27.58
ETRPCA	41.15	31.75	26.26	27.76	28.27
WSAR_TRPCA	**43.75**	**36.31**	**32.23**	**30.02**	**33.29**

Table 2: Comparison of the PSNR values on the 5 images in Fig. 2. Best results are marked bold.

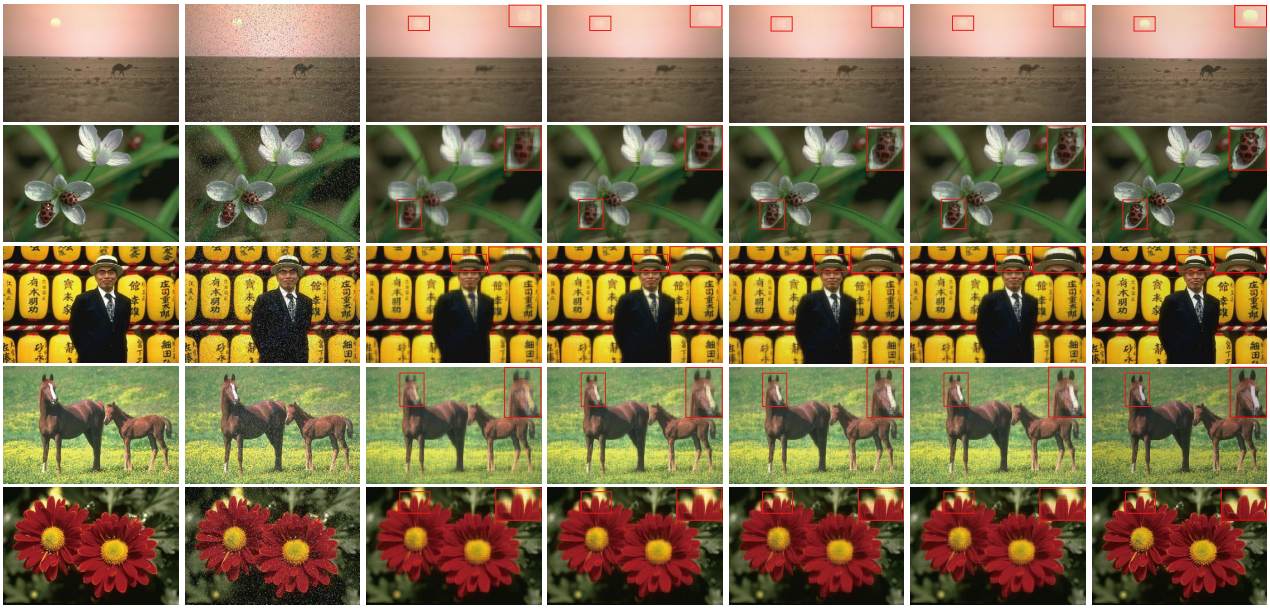
Datasets: In this experiment, we adopt the widely used Berkeley Segmentation Dataset (BSD) (Martin et al. 2001) for evaluation. To test the recovery performance of various methods, we randomly set 10% pixels in each color image to random values in $[0, 255]$. To make the problems more challenging, we make all the 3 channels of the images are corrupted at the same positions and the positions of the corrupted pixels are agnostic to the recovery algorithms.

Comparison methods: For comparison, we consider the following matrix and tensor based recovery methods: RPCA (Candès et al. 2011), Sum of Nuclear Norms (SNN) (Liu et al. 2013), TRPCA (Lu et al. 2020), ETRPCA (Gao et al. 2021) and WSAR_TRPCA. For fair comparison, we adopt the parameters recommended by the corresponding authors. For WSAR_TRPCA, we use the same parameter as TRPCA.

Results: Fig. 2 shows several sample images and the recovered results by different recovery algorithms. Their PSNR values are listed in Table 2. Fig. 3 presents the PSNR values of the recovered images by different comparison methods on 50 images of the BSD dataset. Note from the results that WSAR_TRPCA significantly outperforms other competing methods in terms of both visual quality and quantitative PSNR values.

Background Modeling

This part is to assess the performance of WSAR_TRPCA for background modeling. For a video with moving objects, the target of background modeling is to separate the foreground objects from the background. Since the frames of the background are highly correlated in general, it can be modeled as a low-rank tensor and the foreground moving objects can be modeled as a sparse tensor since the moving objects often occupy a small part of each video frame (Lu et al. 2020).



(a) (b) (c) (d) (e) (f) (g)

Figure 2: Recovery performance comparison on 5 sample images of the BSD dataset. (a) Original image; (b) observed image; (c)-(g) recovered images by RPCA, SNN, TRPCA, ETRPCA and WSAR_TRPCA, respectively. The demarcated area is enlarged in the top right corner for better visualization. The figure is better seen by zooming on a computer screen.

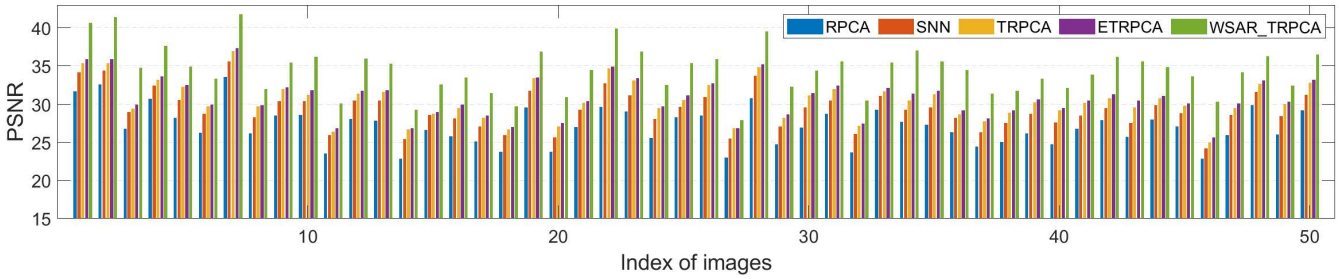


Figure 3: Comparison of the PSNR values of recovered results obtained by different methods on 50 images of the BSD dataset.

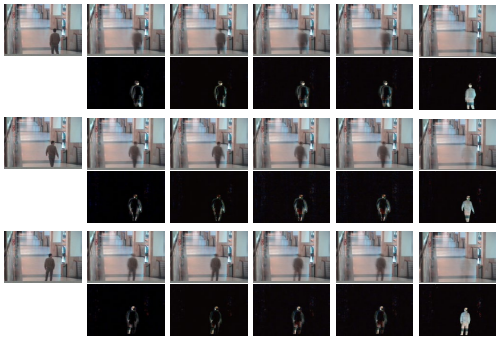
Datasets: We adopt the Scene Background Initialization (SBI) database (Maddalena and Petrosino 2014) and utilize the CAVIAR1 video with the first 100 frames.

Comparison methods: The comparison methods and parameters are the same as those for robust image recovery.

Results: Fig. 4 shows the background modeling results of three frames (60, 70 and 80) by competing algorithms. Note from Fig. 4 that WSAR_TRPCA can produce clean background and coherent foreground masks simultaneously. In comparison, the background images obtained by other competing methods are blurry and have severe ghosting effects.

Conclusion

In this paper, a unified data recovery framework dubbed SAR is proposed to recover high-dimensional data of multiple low-dimensional structures. We also extend SAR and propose a Weighted SAR (WSAR) framework with convergence guarantee. Experimental results verify the effectiveness of WSAR for various applications.



(a) (b) (c) (d) (e) (f)

Figure 4: Background modeling results of video sequences. (a) Original frames; (b)-(g) low rank and sparse components learned by RPCA, SNN, TRPCA, ETRPCA and WSAR_TRPCA, respectively.

Acknowledgments

We are grateful to the anonymous AAAI reviewers for their constructive comments. This work was supported in part by National Natural Science Foundation of China under Grant Nos. 62076041 and 62276111, and in part by the Fundamental Research Funds for the Central Universities under Project 2662023XXPY002, 2662023LXPY005, and SZYJY2023010.

References

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3: 1–122.

Candès, E.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis. *J. ACM*, 58(3): 11.

Candès, E.; Wakin, M. B.; and Boyd, S. P. 2008. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14(5): 877–905.

Chandrasekaran, V.; Recht, B.; Parrilo, P.; and Willsky, A. 2012. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12: 805–849.

Chang, Y.; Yan, L.; Zhao, X.; Fang, H.; Zhang, Z.; and Zhong, S. 2020. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE Trans. Cybern.*, 50(11): 4558–4572.

Chen, Q.; Wu, C.; and Wang, Y. 2019. Robust principal component analysis-based infrared small target detection. In *Proc. AAAI*, volume 33, 9925–9926.

Chen, Y.; Nasrabadi, N.; and Tran, T. 2011. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.*, 49(10): 3973–3985.

Ding, X.; He, L.; and Carin, L. 2011. Bayesian robust principal component analysis. *IEEE Trans. Image Process.*, 20(12): 3419–3430.

Dutta, A.; Hanzely, F.; and Richtárik, P. 2019. A nonconvex projection method for robust PCA. In *Proc. AAAI*, volume 33, 1468–1476.

Elhamifar, E.; and Vidal, R. 2013. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(11): 2765–2781.

Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2021. Enhanced Tensor RPCA and its Application. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(6): 2133–2140.

Gu, S.; Xie, Q.; Meng, D.; Zuo, W.; Feng, X.; and Zhang, L. 2017. Weighted nuclear norm minimization and its applications to low level vision. *Int. J. Comput. Vis.*, 121(2): 183–208.

He, R.; Zheng, W.; Tan, T.; and Sun, Z. 2014. Half-Quadratic-Based Iterative Minimization for Robust Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(2): 261–275.

Kilmer, M.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. Appl.*, 34(1): 148–172.

Lee, K.; Ho, J.; and Driegman, D. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5): 684–698.

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1): 208–220.

Liu, X.; Hou, J.; Peng, J.; Wang, H.; Meng, D.; and Wang, J. 2023. Tensor Compressive Sensing Fused Low-Rankness and Local-Smoothness. In *Proc. AAAI*, 8879–8887.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2020. Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(4): 925–938.

Lu, C.; Feng, J.; Lin, Z.; and Yan, S. 2018. Exact low tubal rank tensor recovery from Gaussian measurements. In *Proc. IJCAI*, 2504–2510.

Maddalena, L.; and Petrosino, A. 2014. Towards benchmarking scene background initialization. In *Proc. ICIAP*, 55–63.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. ICCV*, 416–423.

Peng, J.; Sun, W.; and Du, Q. 2019. Self-Paced Joint Sparse Representation for the Classification of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.*, 57(2): 1183–1194.

Qiu, H.; Wang, Y.; Tang, S.; Meng, D.; and Yao, Q. 2022. Fast and Provable Nonconvex Tensor RPCA. In *Proc. ICML*, 18211–18249.

Wang, Y.; Tang, Y.; Li, L.; Chen, H.; and Pan, J. 2019. Atomic Representation-based Classification: Theory, Algorithm and Applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(1): 6–19.

Wright, J.; A.Yang; Ganesh, A.; Sastry, S.; and Ma, Y. 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(2): 210–227.

Wright, J.; and Ma, Y. 2022. *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications*. Cambridge University Press.

Xu, H.; Caramanis, C.; and Sanghavi, S. 2012. Robust PCA via Outlier Pursuit. *IEEE Trans. Inf. Theory*, 58(5): 3047–3064.

Yang, Z.; and Xie, L. 2015. On gridless sparse methods for line spectral estimation from complete and incomplete data. *IEEE Trans. Signal Process.*, 63(12): 3139–3153.

Zhang, C.; Li, H.; Chen, C.; Qian, Y.; and Zhou, X. 2022. Enhanced group sparse regularized nonconvex regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(5): 2438–2452.

Zheng, J.; Zhang, X.; Wang, W.; and Jiang, X. 2022. Handling slice permutations variability in tensor recovery. In *Proc. AAAI*, volume 36, 3499–3507.

Zhou, P.; and Feng, J. 2017. Outlier-robust tensor PCA. In *Proc. CVPR*, 2263–2271.