A Simultaneous Sparse Learning Algorithm of Structured Approximation with Transformation Analysis Embedded in Bayesian Framework

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

26-01-2021 / 28-01-2021

CITATION

Wang, Guisheng (2021): A Simultaneous Sparse Learning Algorithm of Structured Approximation with Transformation Analysis Embedded in Bayesian Framework. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.13643054.v1>

DOI

[10.36227/techrxiv.13643054.v1](https://doi.org/10.36227/techrxiv.13643054.v1)

A Simultaneous Sparse Learning Algorithm of Structured Approximation with Transformation Analysis Embedded in Bayesian Framework

GUISHENG WANG^{1,*}, YEQUAN WANG², GUOCE HUANG¹, QINGHUA REN¹ and TINGTING REN³

¹ Air Force Engineering University, Xi'an 710077, China

² The 30th Research Institute of China Electronics Technology, Chengdu 610054, China

³ Air Force Research Institute, Beijing 102211, China

*Corresponding author: wgsfuyun@163.com

Sparse approximation is critical to the applications of signal or image processing, and it is conducive to estimate the sparse signals with the joint efforts of transformation analysis. In this study, a simultaneous Bayesian framework was extended for sparse approximation by structured shared support, and a simultaneous sparse learning algorithm of structured approximation (SSL-SA) is proposed with transformation analysis which leads to the feasible solutions more sensibly. Then the improvements of sparse Bayesian learning and iterative reweighting were embedded in the framework to achieve speedy convergence as well as high efficiency with robustness. Furthermore, the iterative optimization and transformation analysis were embedded in the overall learning process to obtain the relative optima for sparse approximation. Finally, compared to conventional reweighting algorithms for simultaneous sparse models with l_1 and l_2 , simulation results present the preponderance of the proposed approach to solve the sparse structure and iterative redundancy in processing sparse signals. The fact indicates that proposed method will be effective to sparsely approximate the various signals and images, which does accurately analyse the target in optimal transformation. It is envisaged that the proposed model could be suitable for a wide range of data in sparse separation and signal denoising.

Keywords: sparse approximation; Bayesian learning; transformation analysis; iterative optimization

Received: date; Accepted: date; Published: date

1. INTRODUCTION

With the flourishing development of digital communication and streaming media, we are surrounded in the explosion of information or data, and there are more and more urgent requirements for dimensionality reduction to acquire the interesting essentials. The sparse approximation [1,2] is the principle of linear inverse problems for signal estimation in multiple applications, such as compressed sensing, machine learning, and data processing. It enables the reconstruction of signals or images with measurements by the sparse model, which achieves signals denoising and compression in practice. More generally, the generated results

may be utilized for data separation and mode classification by sparse approximation. Therefore, the effects of sparse approximation have great impacts on the data representation, and the efficient methods of sparse approximation are imperative for effectiveness and robustness in data.

Various methods have been proposed including sparse subspace clustering, simultaneous sparse learning, and empirical Bayesian models [3-5]. To effectively deal with the miscellaneous issues for collections of high-dimension data in the real-world, sparse subspace clustering has an equal verification point that lies in a union of low-dimension subspaces. Among the infinitely possible representations, a sparse representation

is commensurate with selecting a few points from the same subspace [6]. Further, sparse subspace clustering exploits the assumption of sparsity. Thus, an essential assumption of this method is that requirements of low-dimension signals can be represented as the linear combination of other signals, similarly. In other words, each signal can be divided into a sparse combination of basic elements. A probabilistic subspace clustering approach is presented with rapid clustering for large signal collections [7]. Instead of using the sparse normalizer for sequential data with the existing methods of sparse subspace clustering, a quadratic normalizer is exploited for sparse representation in numerous application scenarios, and a block-diagonal prior is utilized for the spectral clustering affinity. Hence, the normalizer and the prior affinity are integrated into the mechanism to achieve a statistically significant improvement in accuracy [8]. For pursuing the higher clustering accuracies and lower computational load significantly, many novel algorithms are derived from the projection, sparsity and/or low-rank coefficients in the latent space with low-dimension for simultaneous dimensionality reduction and data clustering [9,10].

It has been verified effectively to improve the performance of subspace clustering [11], and a variety of approaches, such as the empirical Bayesian and the hierarchical Bayesian, are proposed to estimate the common support and the coefficients of signals [12]. However, it is limited to selecting the optimal transformations with the diversities of signals in time, and the assumption of common support is too restrictive and unreliable. For instance, the support of time-varying signals changes slowly over a period of time [13]. Recent work [14,15] gives the accountability and motivation needed to prepare non-convex algorithms for simultaneously structured models for simultaneous sparse approximation. While most centralized algorithms for simultaneous sparse approximation [16] are developed for the case where signals commonly share the same supports, it needs some appropriate methods to improve the adaptation in a decentralized manner.

In this paper, a simultaneous sparse learning algorithm of structured approximation (SSL-SA) is developed based on the optimization of transformation analysis. According to the simultaneous Bayesian framework extended by structured shared support, we can utilize the

sparse Bayesian learning, structured approximation and iterative ℓ_2 reweighting to update the iterative optimization and transformation analysis. Simulations show that the proposed framework further improves the validity for general and particular sparse signals more rapidly and accurately. The contributions of this paper are summarized as follows:

(i) A sparse Bayesian learning method for transformation analysis is proposed, where it does not only approximate the various signals and images sparsely, but also analyse the target in optimal transform domains accurately.

(ii) The structured approximation and its iterative version for transformation domain are derived from the sparse distribution, where the closed solution is obtained by the method of alternating optimization.

(iii) The optimal sparse approximation and transform domain are obtained by stochastic descent and flexible attenuation rate for cost function along with its convergence and computational complexity analysis.

The rest of the paper is structured as follows: Section 2 describes the backgrounds of transformation analysis, empirical Bayesian framework and sparse approximation for signals. In Section 3, we decompose signals with structured approximation based on analytic transformation and sparse distribution. In Section 4, a centralized ℓ_2 reweighting algorithm is proposed for the estimation of SSL-SA. Numerical results and discussions are presented in Section 5, where the performance is evaluated based on sparse interference signals with transformation optimization. Conclusions are provided in Section 6.

2. RELATED WORK

2.1. Transformation analysis

The problem of designing a suitable or general transformation for signals can be tracked back to the transformation of Fourier and Cosine as well as their local versions [17]. Plenty of pioneering work in the field refers to Fourier transform [18], Fractional orders [19,20], and wavelet [21], which is used to signals decomposition and analysis. In choosing an adequate analytic transformation, many transform coefficients are close to zero more or less, and thus possess sparse characteristics obviously [22]. Moreover, the orthogonal multi-scale transform dilates an elementary function. Further findings in the field

of transformation design set the stage to more efficient compression for images or codes. Besides, there are other various methods for transformation are used to analyse the spectral features in some aspects, such as Transform domain [23,24], sparse decomposition [25,26] and matching pursuit [27].

Currently, the methods of processing digital signals based on transformation analysis are primarily included Discrete Fourier Transform (DFT), Fractional Fourier Transform (FrFT), Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Short-Time Fourier Transform (STFT). In these cases, all multiform signals can be converted into optimal domains for processing to simplify the forms and characteristics as well as improve the effect of representation and classification. The fact is deserved to be noted that many practical signals can be compressed as collections of some basis elements, which means that they can be expressed with sparse representations properly after transformation by choosing an appropriate basis [28]. And it paves the way to investigate the sparse representation and reconstruction further. Therefore, the orthogonal basis on the determined optimal transformation should be addressed primarily in the sparse approximation of signals.

2.2. Empirical Bayesian

Currently, the methods based on likelihood are allowing robust statistical testing for inferring evolutionary rates [29], and they include Bayesian methods with prior distribution and maximum-likelihood methods without the prior. Among them, empirical Bayesian [30] is a valid structure for all systems in estimating the prior parameters, which will support the development of applications for data analysis and signal process.

In the empirical Bayesian method, the density of posterior probability is taken as the target distribution and the prior Gamma distribution is presumed in the case [31]. Within the whole Bayesian framework, the finer details of posterior probability can be derived from the functions of both the likelihood and the prior probability, which is given by

$$P(\Omega | \omega) = \frac{P(\omega | \Omega)P(\Omega | \gamma)}{\int P(\omega | \Omega)P(\Omega | \gamma)d\Omega} \quad (1)$$

where $P(\omega | \Omega)$ is the given probability rate, and γ is a parameter to control the prior property. It deserves to be considered that the empirical

Bayesian method is proved to be distinct from other Bayesian approaches, and to some extent, the prior distribution is greatly determined by the diversity and multi-factors of the given dataset.

However, it is an important task whereas generally difficult task for signal processing to immediately acquire the priors in practice [32], instead of Equation (1), the estimated rate is regarded as the expected value of the aforementioned parameters:

$$P(\Omega | \omega) = \frac{\int P(\omega | \Omega)P(\Omega | \gamma)P(\gamma)d\gamma}{\int \int P(\omega | \Omega)P(\Omega | \gamma)P(\gamma)d\Omega d\gamma} \quad (2)$$

While the process can be caught in the endless chain of conditional probabilities, the perfect solution of the aforementioned problems is utilized for the estimation of the case, and in further detail, the estimation method takes the place of γ with maximizing the marginal distribution $P(\omega | \gamma)$. Since the estimation might not necessarily stop at the specified steps but is as closed as possible, it leads to the feasible solutions more sensibly. Consequently, the estimation can be achieved by breaking the infinite chain of conditional probabilities [33] in the derivations of Equation (2).

2.3. Sparse approximation

Recently, researchers have focused their efforts continuously on the jointly sparse representations of multiple signals, which extend the applications of sparse approximation approaches, such as signal processing, imaging, and distributed compressed sensing [34-36]. The problem can be stated as following.

It is supposed that there are various signals described in the same phenomenon, and the impacts of noise are considered. We want to find the optimal sparse approximation of signals based on the same or partly the same set of the specifically elementary functions. That is, the problem aims to find the optimal approximation of the sparse signals on the condition that the number of functions is limited [37]. More extensive attentions are diverted for the strict formalization of sparsity model in the following:

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\varepsilon} \quad (3)$$

where the known matrix Φ is $n \times N$, the k sparse N -vector $\mathbf{x}$ is unknown, and the noise vector $\boldsymbol{\varepsilon}$ reflects the levels of potential uncertainty. In the circumstances, we want to seek an approximation $\hat{\mathbf{y}}$ to $\mathbf{y}$, whereas the robust condition of strict

sparsity for $\mathbf{x}$ is taken into account, especially for the realistic applications [38].

3. PROBLEM STATEMENT

It is exactly based on this consideration that the goal of the simultaneous sparse approximation problem is converted into that recovering the given matrix $\mathbf{S}$ and dictionary Φ , where the hypothesis is that each signal shares the same sparsity profile. And the overall process of measurements can be stated as

$$\mathbf{S} = \Phi \mathbf{C} + \varepsilon \quad (4)$$

where $\mathbf{C}$ is a coefficient matrix. The existing method of row-support function is convenient to evaluate the target number of non-zero rows in the matrix $\mathbf{C}$ [39], and the definition of the row-support function is

$$\|\mathbf{C}\|_{0,\text{row}} = \{i \in \{1, 2, \dots, M\} : c_{i,k} \neq 0\} \quad (5)$$

According to the aforementioned row-support function, the problem of simultaneous sparse approximation in Equation (4) can be depicted as

$$\begin{cases} \min_c \frac{1}{2} \|\mathbf{S} - \Phi \mathbf{C}\|_F^2 \\ \text{s.t. } \|\mathbf{C}\|_{0,\text{row}} \leq T \end{cases} \quad (6)$$

And its different forms is derived by

$$\begin{cases} \min_c \|\mathbf{C}\|_{0,\text{row}} \\ \text{s.t. } \frac{1}{2} \|\mathbf{S} - \Phi \mathbf{C}\|_F^2 \leq \varepsilon \end{cases} \quad (7)$$

According to one of the proposed versions in [40], the relaxed form of the function $\|\cdot\|_{0,\text{row}}$ is expressed as

$$J_{p,q}(\mathbf{C}) = \sum_i \|c_{i,\cdot}\|_q^p \quad (8)$$

where the parameters are commonly met the criteria $p \leq 1$ and $q \geq 1$. Hence, the formula in Equation (6) can be derived by the relaxation function:

$$\min_c \frac{1}{2} \|\mathbf{S} - \Phi \mathbf{C}\|_F^2 + \lambda J_{p,q}(\mathbf{C}) \quad (9)$$

where λ is the balancing parameter between the approximation error and sparsity condition.

Instead of just representing objects as superpositions of the traditional Fourier representation, we now have a wide range of additional transformations available to alternate signal representation schemes in various applications. Among them, there includes some derivations such as Wavelet, Wavelet Packets, Cosine Packets, Gabor, Chirplets [41] and etc. Many of these discoveries used multiple response vectors $\mathbf{s}_{\cdot k} \in \mathbb{C}^N$ as tools to explore how people

learn and observe jointly the unknown sparse matrix $\mathbf{C} \in \mathbb{C}^{M \times K}$ from response matrix $\mathbf{S} \in \mathbb{C}^{N \times K}$ under different conditions or transformation domains (e.g., spatial, temporal, or other analytic transformations $\mathbf{T}_\cdot$), which is normally explicated when the sensing matrix $\Phi \in \mathbb{C}^{N \times M}$ is variable

$$\mathbf{T}_\cdot(\mathbf{S}) = \Phi \mathbf{C} + \mathbf{E} \quad (10)$$

where $\mathbf{c}_{\cdot k} \in \mathbb{C}^M$ is in the set of $\mathbf{C}$, and $\mathbf{E}$ denotes the noise matrix for signals.

There are some certain structured subcollections of the elements investigated in previous work, such as time-frequency dictionaries, wavelet packets and cosine packets. Most of the subcollections correspond to certain orthogonal bases in a way. We can derive the basic representation from Equation (10)

$$\mathbf{S} = \tilde{\Phi} \mathbf{T}_\cdot^{-1}(\mathbf{C}) + \tilde{\mathbf{E}} \quad (11)$$

where exists the relationship between the basis $\mathbf{T}_\cdot^{-1}(\mathbf{C})_{B_2} =_{B_2} \mathbf{T}_\cdot^{-1} \mathbf{C}_{B_1}$.

An adaptive method of picking the optimal one from among these bases has been proposed in [42], which delivers near-optimal sparsity representations in the time order of $n \log(n)$. There is a valid expression formulated in accordance with

$$\min \Xi\{\mathbf{T}_\cdot^{-1}(\mathbf{C})\} \quad (12)$$

where the “entropy” means the sum of elements $\Xi\{\mathbf{T}_\cdot^{-1}(\mathbf{C})\} = \sum_i e[\mathbf{T}_\cdot^{-1}(\mathbf{C}_i)]$ and

$e[\mathbf{T}_\cdot^{-1}(\mathbf{C}_i)]$ is a scalar function.

We utilize the sparse distribution to achieve the decomposition for signals, and overcome strict restrictions to the sparsity of signals based on structured approximation, which is more suitable for the approximation of various signals in transformation analysis. In this case, the analytic transformation $\hat{\mathbf{T}}_\cdot^{-1}(\mathbf{C})$ can be divided into two parts: a row-sparse matrix and an element-sparse matrix in detail where they remain in the relationship between mutual embeddedness. Thus, the diversified modes of matrix can be constructed by dealing with relaxed constraints of target signals, and then they are not necessarily forced to share the same support. The aforementioned optimization problem is converted to a problem of multi-optimization:

$$\begin{aligned} \min_{\hat{\mathbf{T}}_\cdot^{-1}(\mathbf{C}), \Phi} & \quad \|\tilde{\Phi} \hat{\mathbf{T}}_\cdot^{-1}(\mathbf{C}) - \mathbf{S}\|_F^2 + \alpha \|\hat{\mathbf{T}}_\cdot^{-1}(\mathbf{C})\|_0 + \\ & \quad \beta \|\hat{\mathbf{T}}_\cdot^{-1}(\mathbf{C})\|_{0,\text{row}} \end{aligned} \quad (13)$$

where the weighting parameters $\alpha > 0$ and $\beta > 0$ adjust the row-sparse matrix and element-sparse

matrix to a certain state of equilibrium. While some of the concerns will be hard to deal with the nonconvexity of the target function, the convex functions l_1 norm and $l_{1,2}$ norm can be well applied in approaching of the multioptimization problem in [3], which is illustrated as following:

$$\min_{\hat{T}_\bullet^{-1}(C), \tilde{\Phi}} \left\| \tilde{\Phi} \hat{T}_\bullet^{-1}(C) - S \right\|_2^2 + \alpha \left\| \hat{T}_\bullet^{-1}(C) \right\|_1 + \beta \left\| \hat{T}_\bullet^{-1}(C) \right\|_{1,2} \quad (14)$$

According to the deprivation of sub-problems decomposition from Equation (14), the iterative optimization can be utilized to close to the approximate solutions through a loop:

$$\begin{cases} \hat{T}_\bullet^{-1*}(C) = \arg \min_{\hat{T}_\bullet^{-1}(C)} \left\| \tilde{\Phi} \hat{T}_\bullet^{-1}(C) - S \right\|_2^2 + \alpha \left\| \hat{T}_\bullet^{-1}(C) \right\|_1 + \beta \left\| \hat{T}_\bullet^{-1}(C) \right\|_{1,2} \\ \tilde{\Phi}^* = \min_{\tilde{\Phi}} \left\| \tilde{\Phi} \hat{T}_\bullet^{-1*}(C) - S \right\|_2^2 \end{cases} \quad (15)$$

To make the generated iterations appear in the bound $\left\| \tilde{\Phi}^* - \tilde{\Phi} \right\|_F \leq \varepsilon$, and the adjacent iteration of a feasible solution $\tilde{\Phi}^*$ is given by the method of gradient descent between the iterations of the t^{th} and $t+1^{\text{th}}$ in Equation (15):

$$\tilde{\Phi}_{t+1}^* = \tilde{\Phi}_t^* - \lambda \nabla J(T, \tilde{\Phi}_t^*) \quad (16)$$

Thereinto, nonnegativity of the attenuation parameter λ is satisfied at all times. As for the decomposition estimate of the cost function, we can gain the gradient of functions under some losses:

$$J(T, \tilde{\Phi}_t^*) = \left\| \tilde{\Phi}_t^* \hat{T}_\bullet^{-1}(C) - S \right\|_2^2 \quad (17)$$

$$= \sum_{k=1}^K \left\| \tilde{\Phi}_t^* \hat{T}_\bullet^{-1}(C) - s_{\bullet k} \right\|_2^2$$

$$\nabla J = \tilde{\Phi}_t^* \hat{T}_\bullet^{-1}(C) \hat{T}_\bullet^{-1}(C)^H \quad (18)$$

One of the most essentials in aforementioned problems deserves to be solved through learning the optimal approximation $\hat{T}_\bullet^{-1*}(C)$. Therefore, we can resolve the task by utilizing the learning method of simultaneous sparse Bayesian for structured approximation.

4. SIMULTANEOUS SPARSE LEARNING WITH STRUCTURED APPROXIMATION (SSL-SA)

In this section, we provide a simultaneous sparse learning approach with structured sparse approximation, and then propose an l_2 reweighting algorithm with Bayesian estimation to solve this task.

We consider the Gaussian model with noise variance σ^2 , and its priors are expressed by

$$\begin{cases} p(s_{\bullet k} | \hat{T}_\bullet^{-1}(C_{\bullet k}); \sigma^2) = \mathcal{N}(s_{\bullet k}; \tilde{\Phi} \hat{T}_\bullet^{-1}(C_{\bullet k}), \sigma^2 I_M) \\ p(\hat{T}_\bullet^{-1}(C_{\bullet k})) = \mathcal{N}(\hat{T}_\bullet^{-1}(C_{\bullet k}); \Gamma_{\bullet k}) \end{cases} \quad (19)$$

where the diagonal matrix $\Gamma_{\bullet k}$ is satisfied with restrictions $\text{diag}(\Gamma_{\bullet k}) = \gamma_{\bullet k}$, and an unknown parameter vector $\gamma_{\bullet k}$ is used for governing the k^{th} signal. Combing the likelihood and priors, then the j^{th} iteration of the posterior density of $\hat{T}_\bullet^{-1}(C_{\bullet k})$ is given by

$$\begin{cases} p(\hat{T}_\bullet^{-1}(C_{\bullet k}) | s_{\bullet k}; \gamma_{\bullet k}) = \frac{p(\hat{T}_\bullet^{-1}(C_{\bullet k}), s_{\bullet k}; \gamma_{\bullet k})}{\int p(\hat{T}_\bullet^{-1}(C_{\bullet k}), s_{\bullet k}; \gamma_{\bullet k}) dC_{\bullet k}} \\ = \mathcal{N}(M, \Sigma) \\ M = [\hat{T}_\bullet^{-1}(C_{\bullet 1}), \hat{T}_\bullet^{-1}(C_{\bullet 2}), \dots, \hat{T}_\bullet^{-1}(C_{\bullet k})] = \Gamma \tilde{\Phi}^T P^{-1} Y \\ \Sigma = \text{Cov}[\hat{T}_\bullet^{-1}(C_{\bullet k}) | s_{\bullet k}; \gamma_{\bullet k}] = \Gamma - \Gamma \tilde{\Phi}^T P^{-1} \Gamma \\ P = \sigma^2 I_M + \tilde{\Phi} \Gamma \tilde{\Phi}^T \end{cases} \quad (20)$$

where M and Σ are mean and covariance of the density respectively.

From Bayesian perspectives, we apply the maximum a posterior (MAP) to estimate $\hat{T}_\bullet^{-1}(C_{\bullet k})$

$$\begin{aligned} \hat{T}_\bullet^{-1}(C_{\bullet k}) &= \arg \max_{\hat{T}_\bullet^{-1}(C_{\bullet k})} p(\hat{T}_\bullet^{-1}(C_{\bullet k}) | s_{\bullet k}; \gamma_{\bullet k}) \\ &= \Gamma_{\bullet k} \tilde{\Phi}^T P^{-1} s_{\bullet k} \end{aligned} \quad (21)$$

It is noteworthy that the element-sparse part $\left\| \hat{T}_\bullet^{-1}(C_{\bullet k}) \right\|_1$ and row-sparse part $\left\| \hat{T}_\bullet^{-1}(C_{\bullet k}) \right\|_{1,2}$ have alternative representation in [3], and set $\gamma_{\bullet k}^a, \gamma^c \geq 0$ as scalar vectors that are common and unique to signals respectively, which give

$$\begin{cases} \left\| \hat{T}_\bullet^{-1}(C_{\bullet k}) \right\|_1 = \min_{\gamma_{\bullet k}^a} \frac{1}{2} \left\| \frac{\hat{T}_\bullet^{-1}(C_{\bullet k})}{\sqrt{\gamma_{\bullet k}^a}} \right\|_2^2 + \left\| \gamma_{\bullet k}^a \right\|_2^2 \\ \left\| \hat{T}_\bullet^{-1}(C) \right\|_{1,2} = \min_{\gamma^c} \sum \left[\frac{\left\| \hat{T}_\bullet^{-1}(C_{\bullet k}) \right\|_2^2}{2\gamma^c} \right] + \left\| \gamma^c \right\|_2^2 \end{cases} \quad (22)$$

Therefore, the convex optimization problem can be cast as

$$\begin{aligned}
& \min_{\hat{T}_\bullet^{-1}(C)} \frac{\|\tilde{\Phi}\hat{T}_\bullet^{-1}(C) - S\|_2^2}{2\sigma^2} + \alpha \|\hat{T}_\bullet^{-1}(C)\| + \beta \|\hat{T}_\bullet^{-1}(C)\|_{1,2} \\
&= \min_{\hat{T}_\bullet^{-1}(C), \gamma_\bullet^a, \gamma^c} \sum \left[\frac{\|\tilde{\Phi}\hat{T}_\bullet^{-1}(C_k) - s_{\bullet,k}\|_2^2}{2\sigma^2} + \frac{\beta \|\hat{T}_\bullet^{-1}(C_k)\|_2^2}{2\gamma^c} \right] + \\
&\quad \frac{\alpha}{2} \sum \left[\frac{\|\hat{T}_\bullet^{-1}(C_k)\|_2^2}{\sqrt{\gamma_\bullet^a}} + \|\gamma_\bullet^a\|_2^2 \right] + \beta \|\gamma^c\|_2^2 \\
&\equiv \min_{\hat{T}_\bullet^{-1}(C_k), \gamma_\bullet^a, \gamma^c} \left[\frac{1}{\sigma^2} \|\tilde{\Phi}\hat{T}_\bullet^{-1}(C_k) - s_{\bullet,k}\|_2^2 + \text{Tr}(\Gamma_\bullet^a) \right] + \vartheta^2 \text{Tr}(\Gamma^c) \\
&\quad \sum \left[\vartheta^2 (\hat{T}_\bullet^{-1}(C_k))^T (\Gamma^c)^{-1} \hat{T}_\bullet^{-1}(C_k) + (\hat{T}_\bullet^{-1}(C_k))^T (\Gamma_\bullet^a)^{-1} \hat{T}_\bullet^{-1}(C_k) \right] \quad (23)
\end{aligned}$$

where $\vartheta > 0$, and $\Gamma_\bullet^a, \Gamma^c$ are diagonal matrices with $\text{diag}(\Gamma_\bullet^a) = \gamma_\bullet^a$, $\text{diag}(\Gamma^c) = \gamma^c$.

With the relationship in [4], there exists

$$\begin{aligned}
S^T P^{-1} S &= \min_{\hat{T}_\bullet^{-1}(C_k)} \frac{1}{\sigma^2} \sum \|\tilde{\Phi}\hat{T}_\bullet^{-1}(C_k) - s_{\bullet,k}\|_2^2 + \\
&\quad \sum (\hat{T}_\bullet^{-1}(C_k))^T (\Gamma_\bullet^a)^{-1} \hat{T}_\bullet^{-1}(C_k) + \quad (24) \\
&\quad \sum (\hat{T}_\bullet^{-1}(C_k))^T (\Gamma^c)^{-1} \hat{T}_\bullet^{-1}(C_k)
\end{aligned}$$

Then the cost function can be derived from

$$L(\hat{T}_\bullet^{-1}(C_k), \gamma^c, \gamma_\bullet^a) = \min_{\hat{T}_\bullet^{-1}(C_k), \gamma_\bullet^a, \gamma^c} S^T P^{-1} S + \text{Tr}(\Gamma_\bullet^a) + \vartheta^2 \text{Tr}(\Gamma^c) \quad (25)$$

Meanwhile, we obtain the following bound based on Equation (22) and Equation (24), which gives by

$$\begin{aligned}
S^T P^{-1} S &\leq \frac{1}{\sigma^2} \sum \left[\|\tilde{\Phi}\hat{T}_\bullet^{-1}(C_k) - s_{\bullet,k}\|_2^2 \right] + \\
&\quad \sum \left[\frac{\|\hat{T}_\bullet^{-1}(C_k)\|_2^2}{\gamma_\bullet^a} + \frac{\|\hat{T}_\bullet^{-1}(C_k)\|_2^2}{\gamma^c} \right] \quad (26)
\end{aligned}$$

And the equality in Equation (26) holds if the relation is met

$$\hat{T}_\bullet^{-1}(C_k) = M_{\bullet,k} = \Gamma_{\bullet,k} \tilde{\Phi}^T P^{-1} s_{\bullet,k} \quad (27)$$

Furthermore, instead of minimizing the cost function over $\hat{T}_\bullet^{-1}(C_k)$, $\gamma_\bullet^a, \gamma^c$ in Equation (25), we treat $\hat{T}_\bullet^{-1}(C_k)$ as a hidden variable with Equation (27) alternatively, and then arrive at the following structured approximation

$$\left\{ \begin{aligned} L(\gamma^c, \gamma_\bullet^a) &\equiv \min_{\gamma_\bullet^a, \gamma^c} \sum \left[\frac{\|\hat{T}_\bullet^{-1}(C_k)\|_2^2}{\sqrt{\gamma_\bullet^a}} + \frac{\Sigma_\bullet^a}{\gamma_\bullet^a} \right] + \\ &\quad \vartheta^2 \sum \frac{\|\hat{T}_\bullet^{-1}(C_k)\|_2^2 + \Sigma^c}{\gamma^c} \quad (28) \\ \Sigma_\bullet^a &= \Gamma_\bullet^a - \Gamma_\bullet^a \tilde{\Phi}^T P^{-1} \Gamma_\bullet^a \\ \Sigma^c &= \Gamma^c - \Gamma^c \tilde{\Phi}^T P^{-1} \Gamma^c \end{aligned} \right.$$

Therefore, we can gain the solutions, which are expressed as

$$\begin{cases} \gamma_\bullet^a = \Sigma_\bullet^a + \|\hat{T}_\bullet^{-1}(C_k)\|_2^2 \\ \gamma^c = \Sigma^c + \vartheta^2 \|\hat{T}_\bullet^{-1}(C_k)\|_2^2 \end{cases} \quad (29)$$

Above all, according to the updating parameters acquired, we can obtain the close form solutions, and utilize the effectiveness under the different transformation domains to update the optimal transformation for corresponding signals, which will further improve the validity for approximation.

Algorithm 1: Simultaneous Sparse Learning with Structured Approximation (SSL-SA)

Inputs: dataset S , dimensionality N , initial attenuation rate α_0 , and iterations $t_{\max}$

Outputs : sensing matrix Φ and analytic transformation $T_\bullet$.

Initialize parameters: make the initial sensing matrix $\Phi^{(0)}$, and $\gamma_\bullet^a, \gamma^c \geq 0$

While $t \leq \text{iterations}$

 Computes the optimal analytic transformation $T_\bullet$ according to Equation (27)

 Cool the attenuation rate α_t , $\alpha_t = t^{-r} \alpha_0^{t-1}$

 Compute the optimal sensing matrix Φ^* ,

$$\tilde{\Phi}^* = \min_{\Phi} \|\tilde{\Phi}\hat{T}_\bullet^{-1}(C) - S\|_2^2$$

 For all $k=1, \dots, N$

 For all $l=1, \dots, k-1$

 Do update the parameters $\gamma_\bullet^a, \gamma^c$ expressed in Equation (29)

 End

 Do update the gradient of cost function:

$$\nabla J = \tilde{\Phi}_t \hat{T}_\bullet^{-1}(C) \hat{T}_\bullet^{-1}(C)^H$$

 Do update the sensing matrix Φ_t :

$$\tilde{\Phi}_{t+1}^* = \tilde{\Phi}_t^* - \alpha \nabla J(T_\bullet(C), \tilde{\Phi}_t^*)$$

 End

 Update the transformation:

$$\hat{T}_\bullet^{-1}(C_k) = M_{\bullet,k} = \Gamma_{\bullet,k} \tilde{\Phi}^T P^{-1} s_{\bullet,k}$$

End

Determine the optimal transformation $T_\bullet$.

$$\begin{aligned} \hat{T}_\bullet^{-1}(C) &= \arg \min_{\hat{T}_\bullet^{-1}(C)} \|\tilde{\Phi}\hat{T}_\bullet^{-1}(C) - S\|_2^2 + \alpha \|\hat{T}_\bullet^{-1}(C)\| + \\ &\quad \beta \|\hat{T}_\bullet^{-1}(C)\|_{1,2} \end{aligned}$$

5. RESULTS AND DISCUSSIONS

To implement design verification of the proposed approach SSL-SA, the effectiveness performance of simultaneously structured approximation is

evaluated based on sparse signals with transformation optimization. Our proposed method SSL-SA is compared with the similarly existing work for sparse approximation where the original Structured Sparse Models for l_1 or l_2 reweighting (SSM-1, SSM-2) [3] and the Bayesian Compressive Sensing (BCS) are included [5]. The number of initial random measurements $N=40$. For SSM-1 and SSM-2, we randomly selected $K=20$ nonzero rows of sparse signals for sparse representation where all the nonzero entries are independent with each other, then the nonzero rows of each column were randomly forced to be zeros. Considering the diversities of sparse signals, we compared the performance of general sparse signals and particular sparse signals under the same conditions, respectively. The same sensing matrix was utilized for all signals, and the noise vectors were generated randomly with Gaussian noise when the signal to noise ratio (SNR) is of 10 dB. For iterative parameters, we set the initial attenuation rate $\alpha_0=1$ and iterations $t_{\max}=20$.

In detail, the general sparse signals are limited in the common signals with lower sparsity whereas the particulars are mainly about the special function, such as radar signals and electromagnetic interference signals with lower or higher sparsity. It is worth noting that the sparse approximation is more suitable for data separation with higher sparsity, and the sparse interference signals are tested in the practical applications for approximation.

5.1. Convergence performance

In the experiments for particular sparse signals, we used a library of common interference sources for Class A impulse interference, which is the jamming type simulated extensively [43]. And the environmental noise was added on the signals with the white Gaussian noise. The basic convergence of performance is evaluated during iterations based on a relative error, which is defined by $\|\hat{C}_{t+1} - C_t\|_F / \|C_t\|_F$. Fig. 1 describes the tendency to converge to a stable state in the inner iterative loop of the all mentioned algorithms.

FIGURE 1. Iterative errors compared to different algorithms.

It is found that the proposed algorithm SSL-SA converges more rapidly than the other mentioned methods, and this phenomenon has been observed in different settings of analogous signals although we only show one example here. The distinctive

convergence of our proposed algorithm SSL-SA compared with SSM-2 is resulted from the combination of iterative optimization and transformation analysis, which determines the optimal transformation for the variable signals.

Meanwhile, the distinction between SSL-SA and BCS is caused by the different model and iterative optimization in the structured elements. In addition to all the foregoing, the convergent difference of the two SSMs is a result of the diversely component reweighting mechanisms.

More specifically, Table 1 illustrates the data statistics of different algorithms about iterative errors, and SSL-SA only needs the 9th iteration whereas the SSMs gets 10th and BCS is 12th when the differential error of iterations is reached the standard about 0.5. Compared with the final convergence, SSM-1, BCS and SSL-SA are in close proximity to 10%, which is obtained the improvement of 18% in the level of SSM-2 on account of the complicated forms of iteration. Furthermore, the median fairly reflects the effectiveness of iterative convergence and the proposed SSL-SA outperforms all the compared algorithms for its modified iteration and transformed optimization.

TABLE 1. Data statistics of iterative errors compared for various algorithms.

	<i>SSM-1</i>	<i>SSM-2</i>	<i>BCS</i>	<i>SSL-SA</i>
<i>Mean</i>	2.455	3.173	4.728	2.714
<i>Standard deviation</i>	7.006	8.979	13.54	8.828
<i>Minimum</i>	0.102	0.118	0.101	0.104
<i>Median</i>	0.4682	0.4239	0.6586	0.3581
<i>Range</i>	31.76	39.88	60.19	39.9

5.2. Time-effectiveness performance

To reflect the timeliness performance of algorithms, the running time of the proposed method is required for the time-sensitive evaluation of tests. For the ten groups of interference signals with the change in frequency, 100 Monte-Carlo simulations were executed when the signal-to-noise ratios within the scope of -5 dB to 15 dB, and the sampling frequency $f_s=1024$ MHz. The channel is assumed to be a Gaussian white noise channel and the initial SNR is 5 dB. All experiments are performed in the MATLAB R2013b environment and the simulation hardware platform is an Intel(R) Core(TM) i7 CPU (3.40 GHz), 4G memory PC.

The mean value of each iterative time for the same group of interference datasets is compared in Fig. 2, and the statistic details of whole groups for dealing with the various interference signals are shown in Fig. 3.

FIGURE 2. The computing time in the whole iterations compared for various algorithms.

FIGURE 3. The median value and standard deviation of computing time compared for various algorithms.

All the SSM-1, SSM-2 and our proposed SSL-SA algorithms take less running time by 14.6% than BCS for the improvement of structured approximation. Furthermore, the two SSMs and SSL-SA consume the approximate time, which means their similar time complexities. In comparison to the two SSMs methods, our SSL-SA improves the stability of iterative optimization about 5% for the updated parameters with Bayesian estimation, respectively.

5.3. Recovery performance

In this experimental verification, the number of sparse signals was fixed as $K=40$, and the number of random measurements ranged from 40 to 120. The sensing matrix was randomly corrupted by the additive Gaussian noise when the initial SNR is of 10 dB. And the average reconstruction error in this paper is generally regarded as $\|\hat{C}-C\|_F/\|C\|_F$, in which $\hat{C}$ is the estimation of C . Fig. 4 and Fig.

5 compared the recovery performance of the mentioned algorithms, which has exhibited the superiority of our proposed solutions.

It is observed in Fig. 4 that the reconstruction errors of all aforementioned algorithms tend to have a decrease together with the increasing number of measurements. In detail, with the convergence closed to an ideal state, the proposed method presents its superiorities of rapidness compared with other three algorithms, and it can obtain the improvement about 10%~20% in effective performance.

FIGURE 4. Comparison of recovery accuracy between different algorithms.

Meanwhile, an error bar of the concrete performance is illustrated in Fig. 5, the fluctuations of our SSM-SA decrease gradually and finally make convergence to the optimal sparse approximation. When the number of measurements remains the degree below 60, the BCS has poor performance for its neglecting estimation of the structured signals. However, for a large number of measurements above 80, the proposed SSL-SA and BCS algorithms perform better than the two SSMs due to the error prorogation of the reconstructing process initially. Considering the fluctuations of the whole process, SSA-SL presents steadier than BCS for the more feasible iteration and optimization, which demonstrates the robustness of SSL-SA with the change of measurements.

FIGURE 5. Fluctuations of recovery accuracy between different algorithms.

5.4. Practical approximating performance

Signal Approximation [20] is used to verify the performance of the proposed method, and in this case, each group of samples should be guaranteed the independence as far as possible. For the verified simulations more generally, the irregular dataset was generated by the different intensity of noise interference mixed into the interference signals in section 5.2, which results in the various dataset more complexly. Moreover, the power of signals was set in the scope of $1/10 P_{\max} \sim 2 P_{\max}$ with the interval of $1/10 P_{\max}$. 60 Monte-Carlo simulations were executed by the obtained 100 samples with the change of interference intensity when the SNR is 5 dB, and the sampling frequency $f_s=512$ MHz. The channel was utilized to simplify as an additive Gaussian noise channel. The root mean square error (RMSE) and the average RMSE were used to evaluate the approximation performance, and the simulation results are demonstrated in Fig. 6 ~ Fig.8.

FIGURE 6. Approximating performance of interference signals with lower power: (a)

The original interference; (b) The approximated sample

FIGURE 7. Approximating performance of interference signals with higher power: (a) The original interference; (b) The approximated sample

FIGURE 8. RMSE of approximated interference signals in numerous simulations

It is observed that the optimal approximation under the diversity of interference power remains an acceptable extent where the RMSEs are approximated below 2.5. And compared with the approximation of higher power samples, the samples with low power are more suitable for sparse approximation due to the rapid adjustments of the iterated parameters on the impact of smaller datum. As a result, the proposed SSL-SA can be extended to effectively deal with the unknown signals with low power. And the detailed relations between RMSE and power are depicted in Fig. 9.

FIGURE 9. Average RMSE of approximation varied with the power of interference signals.

With the increase of the signal intensities, the average RMSE increases gradually with a linear relationship approximately. Furthermore, the effectiveness of the optimal approximation can be judged by the linear fitness between RMS and power, which includes the maximum residual mode and maximum deviation. In our simulations, the fitting equation can be obtained of $y=0.022x+1.434$, and the maximum residual mode is 0.126. Then the significant tests of the linear hypothesis can be conducted by t -test, and the rejection domain is worth consideration. In the end, we can refer the approximated fitness to the analysis of unknown signals extensively.

6. CONCLUSIONS

With the combination of compressed sensing and machine learning, data separation and mode classification can be achieved in the measurements by sparse approximation. While simultaneous sparse learning is successful for sparse approximation problems, the strict assumption turns into essential restrictions for the performance of various and unknown signals processing. The proposed method develops an appropriate simultaneous Bayesian framework in the learning optimization process, structured shared supports and transformation analysis are utilized for the sparse approximate of general and particular signals. The improvement for convergence and efficiency are obtained greatly, which will broaden its applications in signals denoising and images compression. In further study, the derivations of the proposed model will be applied for interference separation, parameters estimation and representation classification.

FUNDING

This work was supported by in part by the University Cooperation Project of Key Laboratory for Aerospace Information Technology, grant number KX162600022, and in part by the Postdoctoral Science Foundation of China, grant number 2016M603044.

Conflict of Interest: The authors declared that there are no conflicts of interest.

REFERENCES

- [1] Abo-Zahhad, M., Hussein, A. and Mohamed, A. (2015) Compressive sensing algorithms for signal processing applications: A survey. *International Journal of Communications, Network and System Sciences*, 8,197-216.
- [2] Massa A., Rocca P., Oliveri G., et al. (2015) Compressive sensing in electromagnetics- A review. *IEEE Antennas and Propagation Magazine*, 57, 224-238.
- [3] Chen, W., Wipf, D., Wang, Y., et al. (2016) Simultaneous Bayesian sparse approximation with structured sparse models. *IEEE Transactions on Signal Processing*, 64, 6145–6159.
- [4] Wipf, D., Rao, B., and Nagarajan, S. (2011) Latent variable Bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57, 6236-6255.
- [5] Yang L., Liu Y., Zhu C., et al. (2018) Synthesis of frequency-invariant patterns by the fast multitask compressive sensing. *IEEE Transactions on Electrical and Electronic Engineering*, 13, 175-176.
- [6] Elhamifar E. and Vidal R. (2013) Sparse subspace clustering: algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 2765-2781.
- [7] Adler A., Elad M. and Hel-Or Y. (2013) Probabilistic subspace clustering via sparse representations. *IEEE Signal Processing Letters*, 20, 63-66.
- [8] Wu F., Hu Y., Gao J., et al. (2016) Ordered subspace clustering with block-diagonal priors. *IEEE Transactions on Cybernetics*, 46, 3209-3219.
- [9] Patel V.M., Nguyen H.V. and Vidal R. (2015) Latent space sparse and low-rank subspace clustering. *IEEE Journal of Selected Topics in Signal Processing*, 9, 691-701.
- [10] Adler A., Elad M. and Hel-Or Y. (2015) Linear-time subspace clustering via bipartite graph modeling. *IEEE Transactions on Neural Networks and Learning Systems*, 26, 2234-2246.
- [11] Vaswani N. and Lu W. (2010) Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Transactions on Signal Processing*, 58, 4595–4607.
- [12] Wipf D. and Rao B. (2007) An empirical bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing*, 55, 3704–3716.
- [13] Zhang, J., Li, Y., Gu, Z., et al. (2014). Recoverability analysis for modified compressive sensing with partially known support. *PloS one*, 9, e87985.
- [14] Oymak S., Jalali A., Fazel M., et al. (2015) Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, 61, 2886–2908.
- [15] Yan Y., Ricci E., Subramanian R., et al. A multi-task learning framework for head-pose estimation under target motion. (2016) *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 1070–1083.
- [16] Ling Q., Wen Z., and Yin W. (2013) Decentralized jointly sparse optimization by reweighted ℓ_q minimization. *IEEE Transactions on Signal Processing*, 61, 1165–1170.
- [17] Zhang Z., Xu Y., Yang J., et al. (2015) A survey of sparse representation: algorithms and applications. *IEEE Access*, 3, 490-530.
- [18] Wang H.J., Wang J.P., Yao F.Q., et al. (2019) Multi-band contour-let transform for adaptive remote sensing image denoising. *The Computer Journal*, 9, bxz073.
- [19] Hong Z., Ge J., Hai Z., et al. (2016) Existence detection of blanket jamming based on fractal characteristics in FRFT domain. *High Power Laser and Particle Beams*, 28,78-84.
- [20] Wang G.S., Wang Y.Q., Huang G.C., et al. (2020). Classification methods with signal approximation for unknown interference. *IEEE Access*, [https://doi.org/ 10.1109/ACCESS.2020.2976128](https://doi.org/10.1109/ACCESS.2020.2976128).
- [21] Kuzovnikov A. V. (2014) Study of the methods for developing jamming-immune communications systems with the use of wavelet-modulated signals. *Journal of*

- Communications Technology & Electronics*, 59, 61-70.
- [22] Maggu J., Aggarwal H.K., and Majumdar A. (2019) Label-consistent transform learning for hyper-spectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 16, 1502-1506.
- [23] Wang G.S., Ren Q.H., Jiang Z.G., et al. (2017) Jamming classification and recognition in transform domain communication system based on signal feature space. *Systems Engineering & Electronics*, 39, 1950-1958.
- [24] Tao C., Pan H.B., Li Y.S., et al. (2015) Unsupervised spectral-spatial feature learning with stacked sparse auto-encoder for hyper-spectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 12, 2438-2442.
- [25] Donoho D.L., Tsaig Y., Drori I., et al. (2012) Sparse solution of underdetermined systems of linear equations by stage-wise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58, 1094-1121.
- [26] Li J.C. (2015) A new robust signal recognition approach based on holder cloud features under varying SNR environment. *Bmc Genomics*, 9, 208-223.
- [27] Zhang G.L., Yang C.L., Wang J.L. (2014) Discrimination of exo-atmospheric targets based on optimization of probabilistic neural network and IR multispectral fusion. *Journal of Electronics & Information Technology*, 36, 896-902.
- [28] Simon F., and Holger R. (2013) *A Mathematical Introduction to Compressive Sensing*. Springer New York.
- [29] Wang M., Maruyama Y. (2016) Consistency of bayes factor for non-nested model selection when the model dimension grows. *Bernoulli*, 22, 2080-2100.
- [30] Marin J., Pudlo P., Robert C.P., et al. (2012) Approximate Bayesian computational methods. *Statistics and Computing*, 2012, 22, 1167-1180.
- [31] Liang F., Cheng Y., Lin G., et al. (2014) Simulated stochastic approximation annealing for global optimization with a square-root cooling schedule. *Journal of the American Statistical Association*, 109, 847-863.
- [32] Sevillalara L., Sun D., Jampani V., et al. (2016) Optical flow with semantic segmentation and localized layers. *IEEE conference on computer vision and pattern recognition*, 3889-3898. <https://doi.org/10.1109/CVPR.2016.422>.
- [33] Liang F. (2011) Annealing evolutionary stochastic approximation Monte-Carlo for global optimization. *Statistics and Computing*, 21, 375-393.
- [34] Gribonval R. and Nielsen M. (2006) Sparse approximations in signal and imaging processing. *Signal Processing*, 86, 415-416.
- [35] Bertrand A. (2015) Distributed signal processing for wireless EEG sensor networks. i *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23, 923-935.
- [36] Park H.R. and Li J. (2016) Sparse covariance-based high resolution time delay estimation for spread spectrum signals. *Electronics Letters*, 51, 155-157.
- [37] Donoho D. and Reeves G. (2012) The sensitivity of compressed sensing performance to relaxation of sparsity. *IEEE International Symposium on Information Theory Proceedings*, Cambridge, MA, 2211-2215.
- [38] Ravazzi C., Fosson S. M., and Magli E. (2015) Distributed iterative thresholding for ℓ_0/ℓ_1 -regularized linear inverse problems. *IEEE Transactions on Information Theory*, 61, 2081-2100.
- [39] Rakotomamonjy A. (2011) Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms. *Signal Processing*, 91, 1505-1526.
- [40] Chen W. (2017) Simultaneous sparse Bayesian learning with partially shared supports. *IEEE Signal Processing Letters*, 24, 1641-1645.
- [41] López-Oller D., Gomez A.M., Pérez-Córdoba J.L., et al. (2016) An error mitigation technique for erasure channels based on a wavelet representation of the speech excitation signal. *IEEE Transactions on Multimedia*, 18, 1245-1256.
- [42] Aoi M., Lepage K., Lim Y., et al. (2015) An approach to time-frequency analysis with ridges of the continuous chirplet transform. *IEEE Transactions on Signal Processing*, 63, 699-710.
- [43] Axell E., Tengstrand S.O., and Wiklundh K. (2017) Online classification of class A impulse interference. *IEEE Military Communications Conference (MILCOM)*, Baltimore, MD, USA, 180-184.