

Convergence analysis on the alternating direction method of multipliers for the cosparse optimization problem

Zisheng Liu ^a and Ting Zhang ^b

^aSchool of Statistics and Big Data, Henan University of Economics and Law, Zhengzhou, China;

^bSchool of Mathematics and Information Science, Henan University of Economics and Law, Zhengzhou, China.

ARTICLE HISTORY

Compiled November 23, 2023

ABSTRACT

From a dual perspective of the sparse representation model, Nam et al. proposed the cosparse analysis model. In this paper, we aim to investigate the convergence of the alternating direction method of multipliers (ADMM) for the cosparse optimization problem. First, we examine the variational inequality representation of the cosparse optimization problem by introducing auxiliary variables. Second, ADMM is used to solve cosparse optimization problem. Finally, by utilizing a tight frame with a uniform row norm and building upon lemmas and the strict contraction theorem, we establish a worst-case $\mathcal{O}(1/t)$ convergence rate in the ergodic sense.

KEYWORDS

sparse representation model, cosparse analysis model, alternating direction method of multipliers, variational inequality, convergence analysis

AMS CLASSIFICATION

90C25

1. Introduction

Low-dimensional signal recovery takes advantage of the inherent low-dimensionality of many natural signals, despite their high ambient dimension. Utilizing prior information about the low-dimensional space can significantly aid in recovering the signal of interest. Sparsity, a widely recognized form of prior information, serves as the foundation for the burgeoning field of compressive sensing (CS [1–5]). The recovery of sparse inputs has found numerous applications in areas such as imaging, speech, radar signal processing, sub-Nyquist sampling, and more [6–9]. A typical sparse recovery problem is associated with the following linear system:

$$y = Mx, \quad (1)$$

where $y \in R^m$ is an observed vector, $M \in R^{m \times d}$ is a measurement matrix and $x \in R^d$ is an unknown signal which would be estimated from y . According to the Nyquist-

Shannon sampling theorem, if the k -space data is undersampled so much that it fails to meet the Nyquist sampling criterion, then reconstructing the data can be difficult or impossible without prior knowledge of x .

1.1. Sparse synthesis model

Over the past decade, the application of compressed sensing significantly increased the image reconstruction speed and efficiency because of its capability to reconstruct images from highly undersampled signals. Sparse prior is widely used in CS-based reconstruction methods. For the sparse synthesis model, if a vector x is sufficient sparse, under the incoherence assumptions on the measurement matrix M , x can be robustly estimated by the problem

$$\begin{aligned} \min_x & \|x\|_\tau \\ \text{s.t. } & y = Mx, \end{aligned} \tag{2}$$

where $0 \leq \tau \leq 1$. The advanced ideas and methods have been explored by applications in signals and image processing [10–13]. After years of research, this model is becoming more and more mature and stable.

1.2. Cosparse analysis model

In the recent decade, the cosparse analysis model is an alternative approach has gained popularity [14,15,17–20]. Within this framework, a potentially redundant analysis operator $D \in \mathbb{R}^{n \times d}$ ($n \geq d$) is employed, and the analyzed vector Dx is expected to be sparse. This implies that a signal $x \in \mathbb{R}^d$ belongs to the cosparse analysis model with cosparsity ℓ if $\ell = n - \|Dx\|_0$. In this paper, the quantity ℓ represents the number of rows in D that are orthogonal to the signal. Consequently, x is referred to as ℓ -cosparse or simply cosparse. The specific definitions of cosparse and cosupport can be found in literature [15], for ease of reference, we have listed them below.

Definition 1.1 (Cosparse). A signal $x \in \mathbb{R}^d$ is said to be cosparse with respect to an analysis operator $D \in \mathbb{R}^{n \times d}$ if the analysis representation vector Dx contains many zero elements. Further, the number of zero elements

$$\ell = n - \|Dx\|_0$$

is called the cosparsity of x , we also say x is ℓ -cosparse.

Definition 1.2 (Cosupport). For a signal $x \in \mathbb{R}^d$ and a given analysis operator $D \in \mathbb{R}^{n \times d}$ with its rows $D_j \in \mathbb{R}^d$ ($1 \leq j \leq n$), the cosupport is defined by

$$\Lambda := \{j | \langle D_j, x \rangle = 0\}.$$

In this paper, D is a tight frame with uniform row norm. We remind the reader that a frame is defined as below.

Definition 1.3 (Frame[21,22]). Let $\Phi = \{\varphi_i\}_{i=1}^N \subseteq \mathbb{R}^n$ be a vector sequence of the

Hilbert space with $N \geq n$. If there exist constants $0 < A \leq B < \infty$ such that

$$\forall x \in R^n, \quad A\|x\|^2 \leq \sum_{i=1}^N |\langle x, \varphi_i \rangle|^2 \leq B\|x\|^2, \quad (3)$$

then Φ is referred to as a finite frame of R^n . The constants A and B in the above formula are known as the lower and upper bounds of the finite frame Φ , respectively. They are considered to be the optimal bounds, with A being the supremum in the lower bound and B being the infimum in the upper bound. If $A = B$, then the frame Φ is called an A -tight frame. If $A = B = 1$, then Φ is called a Parseval frame. If there exists a constant C such that each meta-norm $\|\varphi_i\| = C$ of the frame Φ , then Φ is called an iso-norm frame. In particular, for a tight frame, if $C = 1$, it is referred to as a uniformly tight frame.

According to the definition of cosparsity, the cosparse analysis model focuses on the zero elements of the analysis representation vector Dx , rather than the non-zero elements. This perspective contrasts with the sparse synthesis model. If the cosparsity ℓ is significantly large, meaning that the number of zeros ℓ is close to d , we say that x has a cosparse representation. The cosupport set is identified by iteratively removing rows from D for which $\langle D_j, x \rangle \neq 0$ until the index set Λ remains unchanged, with $|\Lambda| \geq \ell$.

If the analysis representation vector Dx is sparse, similar to the sparse model, the estimation of x from the measurements can be achieved by

$$\begin{aligned} \min_x \|Dx\|_0 \\ s.t. \quad y = Mx. \end{aligned} \quad (4)$$

The minimization problem (4) is known to be NP-hard [15], necessitating the use of approximation methods. Similar to the sparse model, one option is to use the greedy analysis pursuit (GAP) approach, which is inspired by the orthogonal matching pursuit (OMP) algorithm [14–16]. Alternatively, the nonconvex ℓ_0 norm can be approximated by the convex ℓ_1 norm, leading to the relaxed problem known as analysis basis pursuit (ABP) [23]. In this case, x can be estimated by solving a modified optimization problem

$$\begin{aligned} \min_x \|Dx\|_1 \\ s.t. \quad \|y - Mx\|_2 \leq \epsilon, \end{aligned} \quad (5)$$

where $\|\cdot\|_1$ is the ℓ_1 norm that sums the absolute values of a vector and ϵ is a upper bound on the noise level $\|v\|_2$.

ABP is equivalent to the unconstrained optimization

$$\min_x \|Dx\|_1 + \frac{\alpha}{2} \|y - Mx\|_2^2, \quad (6)$$

which we call analysis LASSO (ALASSO). It can be said that ABP and ALASSO are equivalent in the sense that for any $\epsilon > 0$, there exists an α such that the optimal solutions of ABP and ALASSO are identical. For the optimization problem (6), our previous work presented the modified GAP algorithm and error analysis [18,24]. The simulations we conducted demonstrated the advantages of the proposed method for the

cosparse optimization problem. These optimization problems can also be solved using interior point methods [25]. However, as the problem dimension increases, these techniques become time-consuming since they require solutions of linear systems. Other suggested approaches include the alternating direction method of multipliers (ADMM) [26–28,37] and the accelerated alternating minimization method (AAM) [29]. In this paper, we propose a new way to analyze the convergence theory of the cosparse optimization problem based on the variational inequality.

1.3. Organization of the paper

Our focus in this paper is on the cosparse optimization problem and its convergence study based on a variational inequality. The paper is structured as follows: In Section 2, we introduce auxiliary variables and investigate the variational inequality characterization of the cosparse optimization problem. In Section 3, we present several lemmas that establish the strict contraction of the ADMM for the cosparse optimization problem. Using these lemmas and the strict contraction theorem, we establish a worst-case $\mathcal{O}(1/t)$ convergence rate in the ergodic sense. Finally, Section 4 provides a brief conclusion.

2. Preliminaries

To apply the ADMM for solving the cosparse optimization problem (6), we convert the unconstrained optimization problem mentioned above into a constrained optimization problem as follows

$$\begin{aligned} & \min_{x,z} \|z\|_1 + \frac{\alpha}{2} \|y - Mx\|_2^2 \\ & \text{s.t. } Dx - z = 0, \end{aligned} \tag{7}$$

where an auxiliary variable $z \in R^n$ is introduced in (6) to transfer Dx out of the nondifferentiable term $\|\cdot\|_1$ and $\alpha > 0$ is a penalty parameter.

In this section, we summarize the variational inequality (VI) characterization of (7). Initially, we present the optimality condition of the constrained optimization problem (7), which forms the foundation for our subsequent convergence analysis [30,32]. We then proceed to express the Lagrangian function of (7) as follows

$$L(z, x, \lambda) = \|z\|_1 + \frac{\alpha}{2} \|y - Mx\|_2^2 - \lambda^T(Dx - z). \tag{8}$$

In (8), we assume that $x \in \mathcal{X}$, $z \in \mathcal{Z}$ and $\lambda \in R^n$ where $\mathcal{X} \subset R^d$ and $\mathcal{Z} \subset R^n$ are closed convex sets, we call $(z^*, x^*, \lambda^*) \in \Omega := \mathcal{Z} \times \mathcal{X} \times R^n$ to be a saddle point of $L(z, x, \lambda)$ if the following inequalities are satisfied

$$L(z^*, x^*, \lambda) \leq L(z^*, x^*, \lambda^*) \leq L(z, x, \lambda^*). \tag{9}$$

Obviously, a saddle point (z^*, x^*, λ^*) can be characterized by the system

$$\begin{cases} z^* = \arg \min \{L(z, x^*, \lambda^*) | z \in \mathcal{Z}\}, \\ x^* = \arg \min \{L(z^*, x, \lambda^*) | x \in \mathcal{X}\}, \\ \lambda^* = \arg \max \{L(z^*, x^*, \lambda) | \lambda \in R^n\}, \end{cases} \quad (10)$$

which can be rewritten as

$$\begin{cases} z^* \in \mathcal{Z}, L(z, x^*, \lambda^*) - L(z^*, x^*, \lambda^*) \geq 0, \\ x^* \in \mathcal{X}, L(z^*, x, \lambda^*) - L(z^*, x^*, \lambda^*) \geq 0, \\ \lambda^* \in R^n, L(z^*, x^*, \lambda^*) - L(z^*, x^*, \lambda) \geq 0. \end{cases} \quad (11)$$

Below, we present a summary of the method for expressing the optimality condition of the cosparse analysis model (7) via a variational inequality.

Proposition 2.1. Suppose $\mathcal{X} \subset R^d$ is a closed convex set, and $\theta(x) : R^d \rightarrow R$ is a convex function. Furthermore, let $f(x)$ be differentiable in \mathcal{X} . We assume that the set of solutions for the minimization problem $\min\{\theta(x) + f(x) | x \in \mathcal{X}\}$ is nonempty, then,

$$x^* = \arg \min \{\theta(x) + f(x) | x \in \mathcal{X}\} \quad (12)$$

if and only if

$$x, x^* \in \mathcal{X}, \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0. \quad (13)$$

The proof of Proposition 2.1 is available in [33]. Let $\theta_1(z) = \|z\|_1$ and $\theta_2(x) = \frac{\alpha}{2} \|y - Mx\|_2^2$, according to the above inequality (13), a saddle point (z^*, x^*, λ^*) of the Lagrangian function (8) can be characterized by a solution point of the following variational inequality

$$\omega, \omega^* \in \Omega, \theta(u) - \theta(u^*) + (\omega - \omega^*)^T F(\omega^*) \geq 0, \quad (14)$$

where

$$\theta(u) = \theta_1(z) + \theta_2(x), \quad \Omega = \mathcal{Z} \times \mathcal{X} \times R^n, \quad (15)$$

and

$$\omega = \begin{pmatrix} z \\ x \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} z \\ x \end{pmatrix}, \quad F(\omega) = \begin{pmatrix} \lambda \\ -D^T \lambda \\ Dx - z \end{pmatrix}, \quad (16)$$

Since F is an affine operator, and

$$F(\omega) = \begin{pmatrix} 0 & 0 & I \\ 0 & 0 & -D^T \\ -I & D & 0 \end{pmatrix} \begin{pmatrix} z \\ x \\ \lambda \end{pmatrix}, \quad (17)$$

According to the antisymmetry of the affine matrix, it follows that

$$(\omega - \bar{\omega})^T [F(\omega) - F(\bar{\omega})] \equiv 0, \quad \forall \omega, \bar{\omega} \in \Omega. \quad (18)$$

Using inequality (13) and combining (8), we derive the following conclusion with $(z^*, x^*, \lambda^*) \in \Omega$,

$$\begin{cases} \theta_1(z) - \theta_1(z^*) + (z - z^*)^T \lambda^* \geq 0, \\ \theta_2(x) - \theta_2(x^*) + (x - x^*)^T (-D^T \lambda^*) \geq 0, \\ (\lambda - \lambda^*)^T (Dx^* - z^*) \geq 0. \end{cases} \quad (19)$$

After conducting the aforementioned analysis, the linear constrained cosparse optimization problem is reformulated as a variational inequality. Consequently, the task is ultimately simplified to identifying a saddle point of the Lagrangian function. In the subsequent section, the convergence analysis of the ADMM method for addressing the cosparse optimization problem, as denoted by equation (7), will be discussed.

3. Convergence analysis of the cosparse optimization problem

3.1. Variational inequality characterization of ADMM

The augmented Lagrangian function of the problem (7) can be formulated as follows

$$\mathcal{L}_\beta(z, x, \lambda) = \|z\|_1 + \frac{\alpha}{2} \|y - Mx\|_2^2 - \lambda^T (Dx - z) + \frac{\beta}{2} \|Dx - z\|_2^2, \quad (20)$$

where λ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter for the linear constraints. Thus, applying directly the augmented Lagrangian function (20) and starting with an initial iterate $(x^0, \lambda^0) \in \mathcal{X} \times R^n$, the ADMM generates its sequence via following iterative scheme

$$\begin{cases} z^{k+1} = \arg \min \{\mathcal{L}_\beta(z, x^k, \lambda^k) | z \in \mathcal{Z}\}, \\ x^{k+1} = \arg \min \{\mathcal{L}_\beta(z^{k+1}, x, \lambda^k) | x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Dx^{k+1} - z^{k+1}), \quad \lambda \in R^n \end{cases} \quad (21)$$

the corresponding variational inequalities of (21) can be given as

$$\begin{cases} \theta_1(z) - \theta_1(z^{k+1}) + (z - z^{k+1})^T [\lambda^k - \beta(Dx^k - z^{k+1})] \geq 0, \\ \theta_2(x) - \theta_2(x^{k+1}) + (x - x^{k+1})^T [-D^T \lambda^k + \beta D^T (Dx^{k+1} - z^{k+1})] \geq 0, \\ (\lambda - \lambda^{k+1})^T [(Dx^{k+1} - z^{k+1}) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)] \geq 0. \end{cases} \quad (22)$$

For some reviews on the classical ADMM, one can refer to literatures [28,30,31,34–36].

3.2. Assertions

To establish that $\{\omega^k\}$ is strictly contractive with respect Ω , we first present several lemmas.

Lemma 3.1. *Let the sequence $\{\omega^k\}$ be generated by (21). Then, we have*

$$\begin{aligned} & \theta(u) - \theta(u^{k+1}) + (\omega - \omega^{k+1})^T F(\omega) \\ & \geq (z - z^{k+1})^T \beta(Dx^k - Dx^{k+1}) + \frac{1}{\beta}(\lambda - \lambda^{k+1})^T (\lambda^k - \lambda^{k+1}), \quad \forall \omega \in \Omega. \end{aligned} \quad (23)$$

Proof. From (22) we know that

$$\theta_1(z) - \theta_1(z^{k+1}) + (z - z^{k+1})^T [\lambda^k - \beta(Dx^k - z^{k+1})] \geq 0, \quad \forall z \in \mathcal{Z} \quad (24)$$

and

$$\theta_2(x) - \theta_2(x^{k+1}) + (x - x^{k+1})^T (-D^T \lambda^k + \beta D^T (Dx^{k+1} - z^{k+1})) \geq 0, \quad \forall x \in \mathcal{X}. \quad (25)$$

Using $\lambda^{k+1} = \lambda^k - \beta(Dx^{k+1} - z^{k+1})$ we can easily deduce

$$\lambda^k = \lambda^{k+1} + \beta(Dx^{k+1} - z^{k+1}) \quad (26)$$

and

$$(Dx^{k+1} - z^{k+1}) = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}). \quad (27)$$

Putting the formulations (26) and (27) into (24) and (25), respectively, then we have the following inequalities

$$\theta_1(z) - \theta_1(z^{k+1}) + (z - z^{k+1})^T [\lambda^{k+1} + \beta(Dx^{k+1} - z^{k+1}) - \beta(Dx^k - z^{k+1})] \geq 0, \quad (28)$$

$$\theta_2(x) - \theta_2(x^{k+1}) + (x - x^{k+1})^T (-D^T \lambda^{k+1}) \geq 0, \quad (29)$$

and

$$(\lambda - \lambda^{k+1})^T (Dx^{k+1} - z^{k+1}) \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}). \quad (30)$$

Combining (28), (29) and (30) we have

$$\begin{cases} \theta_1(z) - \theta_1(z^{k+1}) + (z - z^{k+1})^T \lambda^{k+1} \geq (z - z^{k+1})^T \beta(Dx^k - Dx^{k+1}), \\ \theta_2(x) - \theta_2(x^{k+1}) + (x - x^{k+1})^T (-D^T \lambda^{k+1}) \geq 0, \\ (\lambda - \lambda^{k+1})^T (Dx^{k+1} - z^{k+1}) \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \end{cases} \quad (31)$$

which is

$$\begin{aligned} & \theta(u) - \theta(u^{k+1}) + (\omega - \omega^{k+1})^T F(\omega^{k+1}) \\ & \geq (z - z^{k+1})^T \beta(Dx^k - Dx^{k+1}) + (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}). \end{aligned}$$

Note that the matrix in the operator F is skew-symmetric, then, using (18), we have

$$\begin{aligned} & \theta(u) - \theta(u^{k+1}) + (\omega - \omega^{k+1})^T F(\omega) \\ & \geq (z - z^{k+1})^T \beta(Dx^k - Dx^{k+1}) + (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}). \end{aligned} \quad (32)$$

The Lemma 3.1 is proved. \square

Lemma 3.2. Let the sequence $\{\omega^k\}$ be generated by (21). Then, we have

$$\begin{aligned} & \beta(z - z^{k+1})^T(Dx^k - Dx^{k+1}) + \frac{1}{\beta}(\lambda - \lambda^{k+1})^T(\lambda^k - \lambda^{k+1}) \\ &= -\frac{1}{2\beta}\|\lambda^k - \lambda\|_2^2 - \frac{\beta}{2}\|Dx^k - z\|_2^2 + \frac{1}{2\beta}\|\lambda^{k+1} - \lambda\|_2^2 + \frac{\beta}{2}\|Dx^{k+1} - z\|_2^2 \\ & \quad + \frac{\beta}{2}\|Dx^k - z^{k+1}\|_2^2. \end{aligned} \quad (33)$$

Proof. Applying the identity

$$(a - b)^T(c - d) = \frac{1}{2}\{\|a - d\|_2^2 - \|a - c\|_2^2\} + \frac{1}{2}\{\|c - b\|_2^2 - \|d - b\|_2^2\}$$

to the left-hand side in (33) with

$$a = z, \quad b = z^{k+1}, \quad c = Dx^k, \quad d = Dx^{k+1},$$

we obtain

$$\begin{aligned} & \beta(z - z^{k+1})^T(Dx^k - Dx^{k+1}) \\ &= \frac{\beta}{2}\{\|z - Dx^{k+1}\|_2^2 - \|z - Dx^k\|_2^2\} + \frac{\beta}{2}\{\|Dx^k - z^{k+1}\|_2^2 - \|Dx^{k+1} - z^{k+1}\|_2^2\}. \end{aligned} \quad (34)$$

Using the identity

$$b^T(b - a) = \frac{1}{2}(\|b\|_2^2 - \|a\|_2^2 + \|b - a\|_2^2),$$

and let

$$a = \lambda - \lambda^k, \quad b = \lambda - \lambda^{k+1},$$

we obtain

$$\frac{1}{\beta}(\lambda - \lambda^{k+1})^T(\lambda^k - \lambda^{k+1}) = \frac{1}{2\beta}\{\|\lambda - \lambda^{k+1}\|_2^2 - \|\lambda - \lambda^k\|_2^2 + \|\lambda^k - \lambda^{k+1}\|_2^2\}. \quad (35)$$

Using

$$\beta\|Dx^{k+1} - z^{k+1}\|_2^2 = \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|_2^2,$$

and combining (34) and (35), we complete the proof of this lemma. \square

Lemma 3.3. Let the sequence $\{x^k\}$, $\{z^k\}$ and $\{\lambda^k\}$ be generated by (21), then,

$$\beta\|Dx^k - z^{k+1}\|_2^2 \geq \beta\|Dx^k - Dx^{k+1}\|_2^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|_2^2. \quad (36)$$

Proof. Based on the second inequality of inequality (31), we can derive the following result

$$\begin{cases} \theta_2(x) - \theta_2(x^{k+1}) + (x - x^{k+1})^T(-D^T\lambda^{k+1}) \geq 0, \\ \theta_2(x) - \theta_2(x^k) + (x - x^k)^T(-D^T\lambda^k) \geq 0. \end{cases} \quad (37)$$

Let $x = x^k$ and $x = x^{k+1}$ in (37), respectively, then

$$\begin{cases} \theta_2(x^k) - \theta_2(x^{k+1}) + (x^k - x^{k+1})^T(-D^T\lambda^{k+1}) \geq 0, \\ \theta_2(x^{k+1}) - \theta_2(x^k) + (x^{k+1} - x^k)^T(-D^T\lambda^k) \geq 0. \end{cases}$$

From above inequalities, we have

$$(\lambda^k - \lambda^{k+1})^T(Dx^k - Dx^{k+1}) \geq 0. \quad (38)$$

Using

$$(Dx^{k+1} - z^{k+1}) = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}),$$

then we obtain

$$\begin{aligned} & \beta \|Dx^k - z^{k+1}\|_2^2 \\ &= \beta \|Dx^k - Dx^{k+1} + Dx^{k+1} - z^{k+1}\|_2^2 \\ &= \beta \|Dx^k - Dx^{k+1} + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|_2^2 \\ &\geq \beta \|Dx^k - Dx^{k+1}\|_2^2 + \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|_2^2. \end{aligned} \quad (39)$$

The proof of this lemma is completed. \square

3.3. Strict contraction

To present the main result of the paper, it is necessary to establish the strict contractility of the iterative sequence. The following subsection provides a proof of the strong contractility of the iterative sequence $\{\omega^k\}$, which relies on Lemma 3.1, Lemma 3.2, and Lemma 3.3.

Theorem 3.4. *Assuming that the sequence $\{\omega^k\}$ is generated by equation (21), we can state the following*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2 \quad (40)$$

where

$$v = \begin{pmatrix} \lambda \\ x \end{pmatrix}, \quad H = \begin{pmatrix} \frac{1}{\beta}I_m & 0 \\ 0 & \beta I_d \end{pmatrix}, \quad \mathcal{V}^* = \{(\lambda^*, x^*) | (z^*, x^*, \lambda^*) \in \Omega\}. \quad (41)$$

Proof. We can deduce from Lemma 3.1 and Lemma 3.2 that

$$\begin{aligned} & \theta(u^{k+1}) - \theta(u) + (\omega^{k+1} - \omega)^T F(\omega) \\ & \leq \frac{1}{2\beta} \|\lambda^k - \lambda\|_2^2 + \frac{\beta}{2} \|Dx^k - z\|_2^2 - \frac{1}{2\beta} \|\lambda^{k+1} - \lambda\|_2^2 - \frac{\beta}{2} \|Dx^{k+1} - z\|_2^2 \\ & \quad - \frac{\beta}{2} \|Dx^k - z^{k+1}\|_2^2. \end{aligned} \quad (42)$$

By utilizing Lemma 3.3, we can rewrite equation (42) as follows

$$\begin{aligned} 0 & \leq \theta(u^{k+1}) - \theta(u^*) + (\omega^{k+1} - \omega^*)^T F(\omega^*) \\ & \leq \frac{1}{2\beta} \|\lambda^k - \lambda^*\|_2^2 + \frac{\beta}{2} \|Dx^k - z^*\|_2^2 - \frac{1}{2\beta} \|\lambda^{k+1} - \lambda^*\|_2^2 - \frac{\beta}{2} \|Dx^{k+1} - z^*\|_2^2 \\ & \quad - \frac{1}{2\beta} \|\lambda^k - \lambda^{k+1}\|_2^2 - \frac{\beta}{2} \|Dx^k - Dx^{k+1}\|_2^2. \end{aligned} \quad (43)$$

That is

$$\begin{aligned} & \frac{1}{\beta} \|\lambda^{k+1} - \lambda^*\|_2^2 + \beta \|Dx^{k+1} - z^*\|_2^2 \\ & \leq \frac{1}{\beta} \|\lambda^k - \lambda^*\|_2^2 + \beta \|Dx^k - z^*\|_2^2 - \left(\frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|_2^2 + \beta \|Dx^k - Dx^{k+1}\|_2^2 \right). \end{aligned} \quad (44)$$

Let

$$Dx^* = z^*, \quad v = \begin{pmatrix} \lambda \\ x \end{pmatrix} \text{ and } H = \begin{pmatrix} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{pmatrix},$$

therefore, the left-hand side of inequality (44) becomes

$$\begin{aligned} & \frac{1}{\beta} \|\lambda^{k+1} - \lambda^*\|_2^2 + \beta \|Dx^{k+1} - z^*\|_2^2 \\ & = \begin{pmatrix} \lambda^{k+1} - \lambda^* \\ x^{k+1} - x^* \end{pmatrix}^T \begin{pmatrix} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{pmatrix} \begin{pmatrix} \lambda^{k+1} - \lambda^* \\ x^{k+1} - x^* \end{pmatrix} \\ & = (v^{k+1} - v^*)^T \begin{pmatrix} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{pmatrix} (v^{k+1} - v^*) \\ & = \|v^{k+1} - v^*\|_H^2. \end{aligned} \quad (45)$$

Likewise, the sum of the first two terms on the right-hand side of inequality (44) is

$$\begin{aligned} & \frac{1}{\beta} \|\lambda^k - \lambda^*\|_2^2 + \beta \|Dx^k - Dx^*\|_2^2 \\ & = \begin{pmatrix} \lambda^k - \lambda^* \\ x^k - x^* \end{pmatrix}^T \begin{pmatrix} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{pmatrix} \begin{pmatrix} \lambda^k - \lambda^* \\ x^k - x^* \end{pmatrix} \\ & = (v^k - v^*)^T \begin{pmatrix} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{pmatrix} (v^k - v^*) \\ & = \|v^k - v^*\|_H^2 \end{aligned} \quad (46)$$

and the sum of the last two terms on the right-hand side of inequality (44) is

$$\begin{aligned}
& \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|_2^2 + \beta \|Dx^k - Dx^{k+1}\|_2^2 \\
&= \left(\begin{array}{c} \lambda^k - \lambda^{k+1} \\ x^k - x^{k+1} \end{array} \right)^T \left(\begin{array}{cc} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{array} \right) \left(\begin{array}{c} \lambda^k - \lambda^{k+1} \\ x^k - x^{k+1} \end{array} \right) \\
&= (v^k - v^{k+1})^T \left(\begin{array}{cc} \frac{1}{\beta} I_m & 0 \\ 0 & \beta D^T D \end{array} \right) (v^k - v^{k+1}) \\
&= \|v^k - v^{k+1}\|_H^2.
\end{aligned} \tag{47}$$

Since $D \in R^{n \times d}$ is a unit tight frame, we have that $D^T D = I_d$. By combining formulas (45), (46), and (47), we complete the proof of the Theorem 3.4. \square

According to Theorem 3.4, we know that H is a positive definite matrix, and inequality (40) implies that the sequence $\{v^k\}$ is bounded. Assuming that the initial vector is $v_0 = (\lambda_0, x_0)^T$, we can obtain the following expression by summing both sides of inequality (40)

$$\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2. \tag{48}$$

The above equation indicates that $\lim_{k \rightarrow \infty} \|v^k - v^{k+1}\|_H^2 = 0$. Therefore, any subsequence v^{k_j} of v^k also has $\lim_{j \rightarrow \infty} \|v^{k_j} - v^{k_j+1}\|_H^2 = 0$. Suppose there exists a subsequence that converges to \bar{v} , then formula (23) implies that \bar{v} is the solution of formula (21). This shows that any accumulation point of the sequence v^k is a solution of (21). According to formula (40), v^k cannot have more than one accumulation point, and hence v^k converges to $\bar{v} \in \mathcal{V}^*$.

3.4. Convergence rate in ergodic sense

Combining with Theorem 3.4, we prove a worst-case $\mathcal{O}(1/t)$ convergence rate in a ergodic sense of the ADMM scheme (21) for cosparse signal reconstruction problem.

Theorem 3.5. *Let the sequence $\{\omega^k\}$ be generated by (21). Then, for any positive integer t , we have*

$$\begin{aligned}
& \theta(u^t) - \theta(u) + (\omega^t - \omega)^T F(\omega) \\
& \leq \frac{1}{2(t+1)} \left[\frac{1}{\beta} \|\lambda^0 - \lambda\|_2^2 + \beta \|Dx^0 - z\|_2^2 \right], \quad \forall \omega \in \Omega
\end{aligned} \tag{49}$$

where

$$\omega^t = \frac{1}{t+1} \left(\sum_{k=0}^t \omega^{k+1} \right). \tag{50}$$

Proof. For any integer k , by (42) we obtain

$$\begin{aligned} & \theta(u^{k+1}) - \theta(u) + (\omega^{k+1} - \omega)^T F(\omega) \\ & \leq \frac{1}{2\beta} \|\lambda^k - \lambda\|_2^2 + \frac{\beta}{2} \|Dx^k - z\|_2^2 - \frac{1}{2\beta} \|\lambda^{k+1} - \lambda\|_2^2 - \frac{\beta}{2} \|Dx^{k+1} - z\|_2^2. \end{aligned} \quad (51)$$

Suppose $k = 0, 1, 2, \dots, t$ are non-negative integers. By summing the left and right ends of the inequality (51), we deduce that

$$\begin{aligned} & \sum_{k=0}^t \theta(u^{k+1}) - (t+1)\theta(u) + \left[\sum_{k=0}^t \omega^{k+1} - (t+1)\omega \right]^T F(\omega) \\ & \leq \frac{1}{2\beta} \|\lambda^0 - \lambda\|_2^2 + \frac{\beta}{2} \|Dx^0 - z\|_2^2, \quad \forall \omega \in \Omega. \end{aligned} \quad (52)$$

The left and right ends of the inequality (52) are multiplied by $\frac{1}{t+1}$ at the same time, and let

$$\omega^t = \frac{1}{t+1} \sum_{k=0}^t \omega^{k+1}, \quad (53)$$

then, the inequality (52) is equivalent to

$$\begin{aligned} & \frac{1}{t+1} \sum_{k=0}^t \theta(u^{k+1}) - \theta(u) + (\omega^t - \omega)^T F(\omega) \\ & \leq \frac{1}{2(t+1)} \left[\frac{1}{\beta} \|\lambda^0 - \lambda\|_2^2 + \beta \|Dx^0 - z\|_2^2 \right]. \end{aligned} \quad (54)$$

Given that the function $\theta(u)$ is convex, let

$$u^t = \frac{1}{t+1} \sum_{k=0}^t u^{k+1} = \frac{1}{t+1} (u^1 + u^2 + \dots + u^t),$$

we can derive the following expression

$$\begin{aligned} \theta(u^t) &= \theta \left[\frac{1}{t+1} (u^1 + u^2 + \dots + u^t) \right] \\ &\leq \frac{1}{t+1} [\theta(u^1) + \theta(u^2) + \dots + \theta(u^t)] \\ &= \frac{1}{t+1} \sum_{k=0}^t \theta(u^{k+1}). \end{aligned} \quad (55)$$

By utilizing equations (54) and (55), we complete the proof of the Theorem 3.5. \square

After t -th iterations, then ω^t defined by (53) satisfies

$$\tilde{\omega} \in \Omega \text{ and } \sup_{\omega \in \mathcal{D}_{\tilde{\omega}}} \{\theta(\tilde{u}) - \theta(u) + (\tilde{\omega} - \omega)^T F(\omega)\} \leq \frac{d}{2t} = \mathcal{O}\left(\frac{1}{t}\right),$$

where

$$D_{\tilde{\omega}} = \{\omega \in \Omega \mid \|\omega - \tilde{\omega}\| \leq 1\}, \quad d := \sup\left\{\frac{1}{\beta} \|\lambda^0 - \lambda\|_2^2 + \beta \|Dx^0 - z\|_2^2 \mid \omega \in \mathcal{D}_{\tilde{\omega}}\right\},$$

and $v^0 = (\lambda^0, x^0)$ is the initial iteration point. That means ω^t is an $\mathcal{O}(\frac{1}{t})$ solution of (22).

4. Conclusions

This paper presents a novel approach to analyze the convergence of cosparse optimization problem. In order to complete the proof of the main theorem, we first give the overall framework of the ADMM to solve the cosparse optimization problem, secondly, through three lemmas, this paper gives the basic inequalities required for the theorem proof, and finally, our analysis establishes a worst-case convergence rate of $\mathcal{O}(1/t)$, which demonstrates the effectiveness of our approach.

Researchers currently rely on a range of methods to solve separable convex optimization problems. Two popular approaches are the generalized symmetric ADMM and parameterizable proximal point algorithms [37–39]. These methods have demonstrated their effectiveness and superiority in various experiments. In our future work, we plan to explore the potential of combining these methods to solve the cosparse signal reconstruction problem.

Funding

The authors were supported by the National Natural Science Foundation of China Mathematics Tian Yuan Fund under grant No. 12226323 and 12226315, the National Natural Science Foundation of China under grant No. 62103136, the Henan Province Undergraduate College Youth Backbone Teacher Training Program.

Acknowledgments

The authors wish to thank Professor Zheng-Hai Huang for providing his valuable comments which have significantly improved the quality of this paper.

References

- [1] D. L. Donoho, *Compressed Sensing*. IEEE Trans. Inform. Theory, 2006, 52(4), 1289-1306.
- [2] Y. Tsaig, David L. Donoho, *Extensions of compressed sensing*. Signal Processing, 2006, 86(3), 549-571.
- [3] E. J. Candés, T. Tao, *Decoding by linear programming*. IEEE Trans. Inform. Theory, 2004, 51(12), 4203-4215.

- [4] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, New York. 2013.
- [5] I. Mishra and S. Jain, *Soft computing based compressive sensing techniques in signal processing: A comprehensive review*, J. Intell. Syst., 2021, 30, 312-326.
- [6] A. M. Bruckstein, D. L. Donoho, M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Rev. 2009, 51 34-81.
- [7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer press, 2010.
- [8] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd edition*, Academic Press, 2008.
- [9] J. L. Starck, F. Murtagh, M. J. Fadili, *Sparse Image and Signal Processing-Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, 2010.
- [10] M. Elad, M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process, 2006, 15 (12), 3736-3745.
- [11] M. Elad, J. L. Starck, Querre, P, Donoho, D.L, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal. 2005, 19 (3), 340-358.
- [12] A. L. Casanovas, G. Monaci, P. Vandergheynst, R. Gribonval, *Blind audiovisual source separation based on sparse representations*, IEEE Trans. Multimedia. 2010, 12 (5), 358-371.
- [13] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, M. E. Davies, *Sparse representations in audio and music: from coding to source separation*, Proc. IEEE., 2010, 98 (6), 995-1005.
- [14] S. Nam, M. Davies, M. Elad, R. Gribonval, *Cosparse analysis modeling-uniqueness and algorithms*, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011, Prague, Czech Republic, May, 2011, 5804-5807.
- [15] S. Nam, M. Davies, M. Elad, R. Gribonval, *The cosparse analysis model and algorithms*, Appl. Comput. Harmon. Anal., 2013, 34 (1), 30-56.
- [16] T. Tirer and R. Giryes, *Generalizing CoSaMP to signals from a union of low dimensional linear subspaces*, Appl. Comput. Harmon. Anal., 2020, 49(1), 99-122.
- [17] R. Giryes, S. Nam, M. Elad, R. Gribonval, M. Davies, *Greedy-like algorithms for the cosparse analysis model*, Linear Algebra Appl., 2014, 441 (15), 22-60.
- [18] Z. Liu, J. Li, W. Li and P. Dai, *A modified greedy analysis pursuit algorithm for the cosparse analysis model*, Numer. Algor., 2017, 74, 867-887.
- [19] H. Song, X. Ren, Y. Lai, H. Meng, *Sparse analysis recovery via iterative cosupport detection estimation*, IEEE Access, 2021, 9, 38386-38395.
- [20] M. S. Kotzagiannidis and M. E. Davies. *Analysis vs synthesis with structure - An investigation of union of subspace models on graphs*, Appl. Comput. Harmon. Anal., 2022, 60, 293-332.
- [21] V. K. Goyal, J. Kovacevic, J. A. Kelner. *Quantized frame expansions with erasures*, Appl. Comput. Harmon. Anal., 2001, 10(3), 203-233.
- [22] D. Li, M. Xue. *Bases and frames in Banach spaces (in Chinese)*, Science Press, Beijing, 2007.
- [23] T. Zhao, C. E. Yonina, B. Amir, and N. Arye, *Smoothing and decomposition for analysis sparse recovery*, IEEE Trans. Signal Process., 2014, 62(7), 1762-1774.
- [24] J. Li, Z. Liu, W. Li, *The reweighted greedy analysis pursuit algorithm for the cosparse analysis model*, IEEE. The 2015 11th International Conference on Natural Computation, 2015, 1018-1022.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K, Cambridge Univ. Press, 2004.
- [26] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via alternating direction method of multipliers*, in Found. Trends Mach. Learn., 2010, 3(1), 1-122.
- [27] M. V. Afonso, J. M. Bioucas-Dias and M. A. T. Figueiredo, *Fast image recovery using*

- variable splitting and constrained optimization*, IEEE Trans. Image Process., 2010, 19(9), 2345-2356.
- [28] D-R. Han, *A survey on some Recent developments of alternating Direction method of multipliers*, J. Oper. Res. Soc. China, 2022, 10, 1-52.
 - [29] J. X. Xie, A. P. Liao and Y. Lei, *A new accelerated alternating minimization method for analysis sparse recovery*, Signal Process., 2018, 145, 167-174.
 - [30] B. S. He and X. M. Yuan, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 2012, 130 (3), 702-709.
 - [31] S-L. Hu and Z-H. Huang, *Alternating direction method for bi-quadratic programming*, Journal of Global Optimization, 2011, 51(3), 429-446.
 - [32] Z-H. Huang, G. H. Lin and N. H. Xiu, *Several developments of variational inequalities and complementarity problems, bilevel programming and MPEC*, Operations Research Transactions, 2014, 18(1), 113-133.
 - [33] B. S. He, F. Ma and X. Yuan, *Convergence study on the symmetric version of ADMM with large step sizes*, SIAM J. Imaging sciences, 2016, 9 (3), 1467-1501.
 - [34] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 2011, 3 (1), 1-122.
 - [35] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives*, Pacific J. Optim., 2015, 11, 619-644.
 - [36] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly nonlinear elliptic equations*, Technical Report STAN-CS-78-674, Computer Science Department, Stanford University, Stanford, CA, 1978.
 - [37] Z. H. Jia, G. Xue, X. J. Cai and D-R. Han, *The convergence rate analysis of the symmetric ADMM for the nonconvex separable optimization problems*, J. Ind. Manag. Optim., 2021, 17(4), 1943-1971.
 - [38] J. Bai, X. Chang, J. Li and F. Xu, *Convergence revisit on generalized symmetric ADMM*, Optim., 2021, 70, 149-168.
 - [39] M. M. Alves, J. Eckstein, M. Geremia and J. G. Melo, *Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms*, Comput. Optim. Appl., 2020, 75, 389-422.