J Fourier Anal Appl (2017) 23:1097–1117
DOI 10.1007/s00041-016-9500-z

CrossMark

The ℓ_1-Analysis in Phase Retrieval with Redundant Dictionary

Bing Gao[1]

Received: 9 January 2016 / Published online: 13 September 2016

Abstract This article presents new results concerning the recovery of a signal from the magnitude only measurements where the signal is not sparse in an orthonormal basis but in a redundant dictionary, which we call it phase retrieval with redundant dictionary for short. To solve this phaseless problem, we analyze the ℓ_1-analysis model. Firstly we investigate the noiseless case with presenting a null space property of the measurement matrix under which the ℓ_1-analysis model provides an exact recovery. Secondly we introduce a new property (S-DRIP) of the measurement matrix. By solving the ℓ_1-analysis model, we prove that this property can guarantee a stable recovery of real signals that are nearly sparse in overcomplete dictionaries.

Keywords Compressed sensing · Phase retrieval · Sparse recovery · ℓ_1-analysis

Mathematics Subject Classification 94A12

1 Introduction

1.1 Phase Retrieval

Phase retrieval is the process of recovering signals from phaseless measurements. It is of fundamental importance in numerous areas of applied physics and engineering

Communicated by Roman Vershynin.

✉ Bing Gao
gaobing@lsec.cc.ac.cn

1 LSEC, Institute of Computational Mathematics, Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing 100091, China

[11,14]. In general form, phase retrieval problem is to estimate the original signal $x_0 \in \mathbb{H}^n$ ($\mathbb{H} = \mathbb{C}$ or $\mathbb{R}$) from

$$|Ax| = |Ax_0| + e, \tag{1.1}$$

where $A = [a_1, \ldots, a_m]^* \in \mathbb{H}^{m\times n}$ is the measurement matrix and $e = [e_1, \cdots, e_m] \in \mathbb{H}^m$ is an error term. While only the magnitude of Ax_0 is available, it is important to note that the setup naturally leads to ambiguous solutions. For example, if $\hat{x} \in \mathbb{H}^n$ is a solution to (1.1), then any multiplication of $\hat{x}$ and a scalar $c \in \mathbb{H}$ ($|c| = 1$) is also a solution to (1.1). Hence, these global ambiguities are considered acceptable for this problem. In this paper, we recover the signal x_0 actually means that we reconstruct x_0 up to a unimodular constant.

It is known that, when $\mathbb{H} = \mathbb{R}$, at least $2n - 1$ measurements are needed to recover a signal $x \in \mathbb{R}^n$ [3]. For the complex case, the minimum number of measurements are proved to be at least $4n-4$ when n is in the form of $n = 2^k+1$, $k \in \mathbb{Z}_+$ [9]. However, for a general dimension n, the same question is still open. About the minimum number of observations, more details can be found in [4,20]. To reduce the measurement number, priori information must be given, such as sparsity, which means that only few elements in the target signal x_0 is nonzero. In view of such sparse signals, phase retrieval is also known as compressive phase retrieval, which have many applications in data acquisition [15,18]. The compressive phase retrieval problem is in fact the magnitude-only compressive sensing problem. For compressive phase retrieval, Wang and Xu explored the minimum number of measurements and extended the null space property in compressed sensing to compressive phase retrieval [20]. In [19], Voroniski and Xu gave the definition of *strong restricted isometry property* (Definition 2.2) and then many conclusions in compressed sensing can be extended to compressive phase retrieval, such as instance optimality [12].

1.2 Phase Retrieval with Redundant Dictionary

The above conclusions in compressive phase retrieval hold just for signals which are sparse in the standard coordinate basis. However, there are many examples in which a signal of interest is not sparse in an orthonormal basis but sparse in an overcomplete dictionary, such as radar images [13]. We refer to such signals as dictionary-sparse signals. In recent years, many researchers laid special stress on analysing these dictionary-sparse signals in compressed sensing [1,7,16]. However, the phase retrieval literature is lacking on this subject. Motivated by the wide application of redundant dictionaries and frames in signal processing and data analysis, we aim to build up a framework for the recovery of dictionary-sparse signals in phase retrieval, which we call it phase retrieval with redundant dictionary.

Suppose $D \in \mathbb{H}^{n\times N}$ is an overcomplete dictionary ($n < N$) or a redundant dictionary. When $n \ll N$, we say the dictionary D is highly overcomplete or highly redundant. Suppose the signal $x_0 \in \mathbb{H}^n$ is sparse in the overcomplete dictionary $D \in \mathbb{H}^{n\times N}$. I.e., there exists a sparse vector $z_0 \in \mathbb{H}^N$, such that $x_0 = Dz_0$. Thus the phase retrieval with redundant dictionary can be interpreted as recovering a signal $x_0 = Dz_0$ from the measurements $|ADz_0|$, where z_0 is sparse. That is to recover Dz_0 from

$$|ADz| = |ADz_0|. \tag{1.2}$$

1.3 The ℓ_1-Analysis Model

Suppose the signal $x_0 \in \mathbb{H}^n$ can be expressed as $x_0 = Dz_0$, where $D \in \mathbb{H}^{n\times N}$ is a redundant dictionary and $z_0 \in \mathbb{H}^N$ is a sparse vector. When $\mathbb{H} = \mathbb{C}$, we use D^* to represent the adjoint conjugate of D. When $\mathbb{H} = \mathbb{R}$, we use D^* to represent the transpose of D. In compressed sensing, to reconstruct the signal x_0, the most commonly used model is the ℓ_1-analysis model

$$\min \|D^*x\|_1 \qquad \text{subject to} \qquad \|Ax - Ax_0\|_2^2 \le \epsilon^2, \tag{1.3}$$

where ϵ is the upper bound of the noise. Due to the smaller dimension of the unknown, ℓ_1-analysis leads to a simple optimization problem, which is considerably easier to solve. That's why the ℓ_1-analysis model is widely used. We refer interested readers to [1,7,10] for more superiorities of ℓ_1-analysis model. In [7], Candès et al. proved that when D is a tight frame and D^*x_0 is almost k-sparse, the ℓ_1-analysis (1.3) can guarantee a stable recovery provided that the measurement matrix is Gaussian random matrix with $m = \mathcal{O}(k\log(n/k))$.

For the phase retrieval with redundant dictionary (1.2), we also consider the corresponding ℓ_1-analysis model

$$\min \|D^*x\|_1 \qquad \text{subject to} \qquad \||Ax| - |Ax_0|\|_2^2 \le \epsilon^2, \tag{1.4}$$

where ϵ is the upper bound of the noise level. In this paper, we aim to explore the conditions under which the ℓ_1-analysis model (1.4) can generate an accurate or a stable solution to (1.2). First for the noiseless case, we analyze the null space of the measurement matrix and give the conditions for exact recovery. Then for the noise case, we give a new property on the measurement matrix and prove that this property can guarantee a stable recovery.

Note that when $D = I$, the phase retrieval with redundant dictionary is reduced to the traditional phase retrieval and the ℓ_1-analysis model is reduced to

$$\min \|x\|_1 \qquad \text{subject to} \qquad \||Ax| - |Ax_0|\|_2^2 \le \epsilon^2. \tag{1.5}$$

For this case, when $\mathbb{H} = \mathbb{R}$, Gao et al. provided a detailed analysis of (1.5) in [12] and had the conclusion that a k-sparse signal can be stably recovered by $\mathcal{O}(k\log(n/k))$ Gaussian random measurements. Then a natural question that comes to mind is whether this conclusion still holds for a general frame D.

1.4 Organization

The rest of the paper is organized as follows. In Sect. 2, we give notations and recall some previous conclusions. In Sect. 3, for noiseless case ($\epsilon = 0$), we analyze the null space of the measurement matrix and give sufficient and necessary conditions for (1.4) to achieve an exact solution, which will be discussed in real and complex case separately. In general, it's hard to check whether a matrix satisfies the null space property or not. So in Sect. 4, we introduce a new property (S-DRIP) (Definition 4.1) on the measurement matrix, which is a natural generalization of the DRIP (see [7] for more details). Using this property, we prove that when the measurement matrix is real Gaussian random matrix with $m \geq \mathcal{O}(k \log(n/k))$, the ℓ_1-analysis (1.4) can guarantee a stable recovery of real signals which are k-sparse in a redundant dictionary. In Sect. 5, we discuss the drawbacks of our results and file out some proper directions for the coming study. Lastly, some proofs are given in the Appendix.

2 Notations and Previous Results

We use ℓ_0-norm to measure the cardinality of non-zeros of a vector z. We call a signal z is k-sparse, if there are at most k non-zero elements in the signal, i.e., $\|z\|_0 \leq k$. A set of vectors $\{d_1, \cdots, d_N\}$ in $\mathbb{H}^n$ is a frame of $\mathbb{H}^n$ if there exist constants $0 < s \leq t < \infty$ such that for any $f \in \mathbb{H}^n$,

$$s\|f\|_2^2 \leq \sum_{j=1}^{N} |\langle f, d_j \rangle|^2 \leq t\|f\|_2^2.$$

If $s = t$, the frame is a tight frame. We call $D \in \mathbb{H}^{n \times N}$ a frame in the sense that the columns of D form a frame. Let

$$\Sigma_k^N := \left\{x \in \mathbb{H}^N : \|x\|_0 \leq k\right\}$$

and

$$D\Sigma_k^N := \left\{x \in \mathbb{H}^n : \exists z \in \Sigma_k^N, x = Dz\right\}.$$

Suppose the target signal x_0 is in the set $D\Sigma_k^N$, which means that x_0 can be represented as $x_0 = Dz_0$, where $z_0 \in \Sigma_k^N$.

The best k-term approximation error is defined as

$$\sigma_k(x)_1 := \min_{z \in \Sigma_k} \|x - z\|_1.$$

For positive integers p, q with $p \leq q$, we use $[p : q]$ to represent the set $\{p, p + 1, \ldots, q - 1, q\}$. Suppose $T \subseteq [1 : m]$ is a subset of $[1 : m]$. We use T^c to represent the complement set of T and $|T|$ to denote the cardinal number of T. Let $A_T := [a_j, j \in T]^*$ denote the sub-matrix of A where only rows with indices in T are kept. Denote $\mathcal{N}(A)$ as the null space of A.

Definition 2.1 (DRIP) [7] Fix a dictionary $D \in \mathbb{R}^{n \times N}$ and a matrix $A \in \mathbb{R}^{m \times n}$. The matrix A satisfies the DRIP with parameters δ and k if

$$(1-\delta)\|Dz\|_2^2 \le \|ADz\|_2^2 \le (1+\delta)\|Dz\|_2^2$$

holds for all k-sparse vectors $z \in \mathbb{R}^N$.

The paper [7] have shown that Gaussian random matrices and other random compressed sensing matrices satisfy the DRIP of order k provided the number of measurements m on the order of $\mathcal{O}(k \log(n/k))$.

Definition 2.2 (SRIP)[19] We say the matrix $A = [a_1, \cdots, a_m]^\top \in \mathbb{R}^{m\times n}$ has the Strong Restricted Isometry Property of order k and constants $\theta_-,\ \theta_+ \in (0, 2)$ if

$$\theta_-\|x\|_2^2 \le \min_{I\subseteq[1:m],|I|\ge m/2} \|A_I x\|_2^2 \le \max_{I\subseteq[1:m],|I|\ge m/2} \|A_I x\|_2^2 \le \theta_+\|x\|_2^2$$

holds for all k-sparse signals $x \in \mathbb{R}^n$.

This property was first introduced in [19]. Voroninski and Xu also proved that the Gaussian random matrices satisfy SRIP with high probability.

Theorem 2.1 [19] *Suppose that $t > 1$ and $A \in \mathbb{R}^{m\times n}$ is a Gaussian random matrix with $m = \mathcal{O}(tk \log(n/k))$. Then there exist θ_-, θ_+, with $0 < \theta_- < \theta_+ < 2$, such that A satisfies SRIP of order tk and constants θ_-, θ_+, with probability $1 - exp(-cm/2)$, where $c > 0$ is an absolute constant and θ_-, θ_+ are independent of t.*

3 The Null Space Property

In this section, for any $x_0 \in D\Sigma_k^N$, we consider the noiseless case of (1.4),

$$\min \|D^* x\|_1 \qquad \text{subject to} \qquad |Ax| = |Ax_0|. \tag{3.6}$$

Similarly as the traditional compressed sensing problem, we analyze the null space of the measurement matrix A to explore conditions under which (3.6) can obtain cx_0 ($|c| = 1$).

3.1 The Real Case

We first restrict the signals and measurements to the field of real numbers. The next theorem provides a sufficient and necessary condition for the exact recovery of (3.6).

Theorem 3.1 *For a given matrix $A \in \mathbb{R}^{m\times n}$ and a dictionary $D \in \mathbb{R}^{n\times N}$, we claim that the following properties are equivalent.*

(A) For any $x_0 \in D\Sigma_k^N$,

$$\operatorname{argmin}_{x\in\mathbb{R}^n}\{\|D^* x\|_1 : |Ax| = |Ax_0|\} = \{\pm x_0\}.$$

(B) For any $T \subseteq [1:m]$*, it holds*

$$\|D^*(u+v)\|_1 < \|D^*(u-v)\|_1$$

for all

$$u \in \mathcal{N}(A_T)\backslash\{0\}, \quad v \in \mathcal{N}(A_{T^c})\backslash\{0\}$$

satisfying

$$u + v \in D\Sigma_k^N.$$

Proof (B)⇒(A). Assume (A) is false, namely, there exists a solution $\hat{x} \neq \pm x_0$ to (3.6). As $\hat{x}$ is a solution, we have

$$|A\hat{x}| = |Ax_0| \tag{3.7}$$

and

$$\|D^*\hat{x}\|_1 \leq \|D^*x_0\|_1. \tag{3.8}$$

Denote $a_j^\top, j = 1, \ldots, m$ as the rows of A. Then (3.7) implies that there exists a subset $T \subseteq [1:m]$ satisfying

$$j \in T, \quad \langle a_j, x_0 + \hat{x} \rangle = 0,$$

$$j \in T^c, \quad \langle a_j, x_0 - \hat{x} \rangle = 0,$$

i.e.,

$$A_T(x_0 + \hat{x}) = 0, \quad A_{T^c}(x_0 - \hat{x}) = 0.$$

Define

$$u := x_0 + \hat{x}, \quad v := x_0 - \hat{x}.$$

As $\hat{x} \neq \pm x_0$, we have $u \in \mathcal{N}(A_T)\backslash\{0\}$, $v \in \mathcal{N}(A_{T^c})\backslash\{0\}$ and $u + v = 2x_0 \in D\Sigma_k^N$. Then from (B), we know

$$\|D^*x_0\|_1 < \|D^*\hat{x}\|_1,$$

which contradicts with (3.8).

(A)⇒(B). Assume (B) is false, which means that there exists a subset $T \subseteq [1:m]$,

$$u \in \mathcal{N}(A_T)\backslash\{0\}, \quad v \in \mathcal{N}(A_{T^c})\backslash\{0\}, \tag{3.9}$$

such that

$$u + v \in D\Sigma_k^N$$

and

$$\|D^*(u+v)\|_1 \geq \|D^*(u-v)\|_1. \tag{3.10}$$

Let $x_0 := u + v \in D\Sigma_k^N$ be the signal we want to recover. Set $\tilde{x} := u - v$ and we have $\tilde{x} \neq \pm x_0$. Then from (3.10) we have

$$\|D^*\tilde{x}\|_1 \leq \|D^*x_0\|_1. \tag{3.11}$$

Let $a_j^\top$, $j = 1, \ldots, m$ denote the rows of A. Then from the definition of x_0 and $\tilde{x}$, we have

$$2\langle a_j, u\rangle = \langle a_j, x_0 + \tilde{x}\rangle,$$
$$2\langle a_j, v\rangle = \langle a_j, x_0 - \tilde{x}\rangle.$$

By (3.9), the subset T satisfies

$$j \in T, \quad \langle a_j, x_0\rangle = -\langle a_j, \tilde{x}\rangle$$

and

$$j \in T^c, \quad \langle a_j, x_0\rangle = \langle a_j, \tilde{x}\rangle,$$

which implies

$$|Ax_0| = |A\tilde{x}|. \tag{3.12}$$

Putting (3.11) and (3.12) together, we know $\tilde{x}$ is a solution to model (3.6). However, $\tilde{x} \neq \pm x_0$ contradicts with (A). □

3.2 The Complex Case

We now consider the same problem in complex case which means that the signals and measurements are all in the complex number field. Let $\mathcal{S} = \{S_1, \ldots, S_p\}$ be a partition of $[1:m]$. The next theorem is a generalization of Theorem 3.1.

Theorem 3.2 *For a given matrix $A \in \mathbb{C}^{m\times n}$ and a dictionary $D \in \mathbb{C}^{n\times N}$, we claim that the following two properties are equivalent.*

(A) For any given $x_0 \in D\Sigma_k^N$,

$$\operatorname{argmin}_{x\in\mathbb{C}^n}\{\|D^*x\|_1 : |Ax| = |Ax_0|\} = \{cx_0, c \in \mathbb{S}\}.$$

(B) Suppose $\mathcal{S} = \{S_1, \ldots, S_p\}$ *is a partition of* $[1 : m]$. *For any* $\eta_j \in \mathcal{N}(A_{S_j})\backslash\{0\}$, *if*

$$\frac{\eta_1 - \eta_l}{c_1 - c_l} = \frac{\eta_1 - \eta_j}{c_1 - c_j} \in D\Sigma_k^N\backslash\{0\}, \quad j, l \in [2 : p], \quad j \neq l \tag{3.13}$$

holds for some pairwise distinct $c_1, \ldots, c_p \in \mathbb{S}$, *we have*

$$\|D^*(\eta_j - \eta_l)\|_1 < \|D^*(c_l\eta_j - c_j\eta_l)\|_1.$$

Proof $(B) \Rightarrow (A)$. Suppose the statement (A) is false. That is to say, there exists a solution $\hat{x} \notin \{cx_0, c \in \mathbb{S}\}$ to (3.6) which satisfies

$$\|D^*\hat{x}\|_1 \leq \|D^*x_0\|_1 \tag{3.14}$$

and

$$|Ax_0| = |A\hat{x}|. \tag{3.15}$$

Denote a_j^*, $j = 1, \ldots, m$ as the rows of A. From (3.15) we have

$$\langle a_j, c_jx_0\rangle = \langle a_j, \hat{x}\rangle,$$

with $c_j \in \mathbb{S}$, $j = 1, \ldots, m$. We can define an equivalence relation on $[1 : m]$, namely $j \sim l$, when $c_j = c_l$. This equivalence relation leads to a partition $\mathcal{S} = \{S_1, \ldots, S_p\}$ of $[1 : m]$. For any S_j, we have

$$A_{S_j}(c_jx_0) = A_{S_j}\hat{x}.$$

Set $\eta_j := c_jx_0 - \hat{x}$. Then we have $\eta_j \in \mathcal{N}(A_{S_j})\backslash\{0\}$ and

$$\frac{\eta_1 - \eta_l}{c_1 - c_l} = \frac{\eta_1 - \eta_j}{c_1 - c_j} = x_0 \in D\Sigma_k^N, \quad \text{for all } j, l \in [2 : p], \quad j \neq l.$$

According to the condition (B), we can get

$$\|D^*(\eta_j - \eta_l)\|_1 < \|D^*(c_l\eta_j - c_j\eta_l)\|_1,$$

i.e.,

$$\|D^*(c_j - c_l)x_0\|_1 < \|D^*(c_j - c_l)\hat{x}\|_1.$$

That is equivalent to

$$\|D^* x_0\|_1 < \|D^* \hat{x}\|_1,$$

which contradicts with (3.14).

$(A) \Rightarrow (B)$. Assume (B) is false, namely, there exists a partition $\mathcal{S} = \{S_1, \ldots, S_p\}$ of $[1:m]$, $\eta_j \in \mathcal{N}(A_{S_j})\backslash\{0\}$, $j \in [1:p]$ and some pairwise distinct $c_1, \ldots, c_p \in \mathbb{S}$ satisfying (3.13) but

$$\|D^*(\eta_{j_0} - \eta_{l_0})\|_1 \geq \|D^*(c_{l_0}\eta_{j_0} - c_{j_0}\eta_{l_0})\|_1$$

holds for some distinct $j_0, l_0 \in [1:p]$. Set

$$\tilde{x} := c_{l_0}\eta_{j_0} - c_{j_0}\eta_{l_0}, \quad c_{l_0} \neq c_{j_0},$$
$$x_0 := \eta_{j_0} - \eta_{l_0} \in D\Sigma_k^N.$$

Then we have

$$\tilde{x} \notin \{cx_0, c \in \mathbb{S}\}$$

and

$$\|D^*\tilde{x}\|_1 \leq \|D^* x_0\|_1. \tag{3.16}$$

Let a_j^*, $j = 1, \ldots, m$ denote the rows of A. From $\eta_j \in \mathcal{N}(A_{S_j})\backslash\{0\}$, we obtain

$$\langle a_k, \eta_{j_0}\rangle = 0 \text{ and } \langle a_k, \eta_{l_0}\rangle = 0, \ k \in S_{l_0} \cup S_{j_0}.$$

The definition of x_0 and $\tilde{x}$ implies

$$|\langle a_k, x_0\rangle| = |\langle a_k, \tilde{x}\rangle|, \ k \in S_{l_0} \cup S_{j_0}. \tag{3.17}$$

While for $k \notin S_{l_0} \cup S_{j_0}$, we might as well suppose $k \in S_t$ $(t \neq l_0, j_0)$, i.e., $\langle a_k, \eta_t\rangle = 0$. From

$$\frac{\eta_1 - \eta_l}{c_1 - c_l} = \frac{\eta_1 - \eta_j}{c_1 - c_j},$$

we can obtain

$$\frac{\eta_j - \eta_l}{c_j - c_l} = \frac{\eta_m - \eta_n}{c_m - c_n},$$

here j, l, m, n are distinct integers. Set

$$y_0 := \frac{\eta_{j_0} - \eta_t}{c_{j_0} - c_t} = \frac{\eta_{l_0} - \eta_t}{c_{l_0} - c_t}.$$

Then we have

$$\eta_{j_0} = (c_{j_0} - c_t)y_0 + \eta_t,$$
$$\eta_{l_0} = (c_{l_0} - c_t)y_0 + \eta_t.$$

So $\tilde{x}$ and x_0 can be rewritten as

$$\tilde{x} = c_{l_0}\eta_{j_0} - c_{j_0}\eta_{l_0} = c_t(c_{j_0} - c_{l_0})y_0 + (c_l - c_j)\eta_t,$$
$$x_0 = \eta_{j_0} - \eta_{l_0} = (c_{j_0} - c_{l_0})y_0.$$

Then $\langle a_k, \eta_t\rangle = 0$ implies

$$|\langle a_k, \tilde{x}\rangle| = |\langle a_k, x_0\rangle|, \; k \in S_t.$$

Using a similar argument, we can prove that the claim is also true for other subset S_j. So we have

$$|\langle a_k, \tilde{x}\rangle| = |\langle a_k, x_0\rangle|, \text{ for all } k. \tag{3.18}$$

Combining (3.16) and (3.18), we know $\tilde{x}$ is also a solution to (3.6). However, $\tilde{x} \notin \{cx_0, c \in \mathbb{S}\}$ contradicts with (A). □

Remark 3.1 If we choose $D = I$, the null space property in Theorem 3.1 and Theorem 3.2 is consistent with the null space property which was introduced in paper [20].

According to the Theorems 3.1 and 3.2, if the measurement matrix satisfies the null space property, we can obtain an exact solution by solving model (3.6). But in general, condition (B) in Theorems 3.1 or 3.2 is difficult to be checked. So in Sect. 4, we provide another property (S-DRIP) of the measurement matrix which can also guarantee an exact recover of model (3.6) in noiseless case. In addition, we prove that this property can be satisfied by Gaussian random matrix.

4 S-DRIP and Stable Recovery

In compressed sensing, for any tight frame D, [7] had the conclusion that a signal $x_0 \in D\Sigma_k^N$ can be approximately reconstructed by ℓ_1-analysis (1.3) provided the measurement matrix satisfies DRIP and the best k-term approximation error of D^*x_0 is small. While in phase retrieval, when $\mathbb{H} = \mathbb{R}$, Gao et al. proved that if the measurement matrix satisfies SRIP, then the ℓ_1-analysis (1.5) can provide a stable solution to traditional phase retrieval problem [12]. For the phase retrieval with redundant dictionary, we combine the above two results to explore the conditions under which the ℓ_1-analysis model (1.4) can guarantee a stable recovery.

We first impose a natural property on the measurement matrix, which is a combination of DRIP and SRIP.

Definition 4.1 (S-DRIP) Let $D \in \mathbb{R}^{n\times N}$ be a frame. We say the measurement matrix A obeys the S-DRIP of order k with constants $\theta_-, \theta_+ \in (0, 2)$ if

$$\theta_- \|Dv\|_2^2 \le \min_{I\subseteq[1:m],|I|\ge m/2} \|A_I Dv\|_2^2 \le \max_{I\subseteq[1:m],|I|\ge m/2} \|A_I Dv\|_2^2 \le \theta_+ \|Dv\|_2^2$$

holds for all k-sparse signals $v \in \mathbb{R}^N$.

Thus a matrix $A \in \mathbb{R}^{m\times n}$ satisfying S-DRIP means that any $m' \times n$ submatrix of A, with $m' \ge m/2$ satisfies DRIP with appropriate parameters.

In fact any matrix $A \in \mathbb{R}^{m\times n}$ obeying

$$\mathbb{P}\left[c_- \|Dv\|_2^2 \le \min_{I\subseteq[1:m],|I|\ge m/2} \|A_I Dv\|_2^2 \le \max_{I\subseteq[1:m],|I|\ge m/2} \|A_I Dv\|_2^2 \le c_+ \|Dv\|_2^2\right] \ge 1 - 2e^{-\gamma m} \tag{4.19}$$

($0 < c_- < c_+ < 2$ and γ is a positive number constant) for fixed $Dv \in \mathbb{R}^n$ will satisfy the S-DRIP with high probability. This can be seen by a standard covering argument (see the proof of Theorem 2.1 in [19]). In [19], Voroninski and Xu proved that Gaussian random matrix satisfies (4.19) in Lemma 4.4. So we have the following conclusion.

Corollary 4.1 *For $t > 1$, Gaussian random matrix $A \in \mathbb{R}^{m\times n}$ with $m = \mathcal{O}(tk\log(n/k))$ satisfies the S-DRIP of order tk and constants $\theta_-, \theta_+ \in (0, 2)$ with probability $1 - 2e^{-\gamma m}$, where γ is an absolute positive constant and θ_-, θ_+ are independent of t.*

For any $x_0 \in D\Sigma_k^N$, we return to consider the solving model

$$\min \|D^* x\|_1 \qquad \text{subject to} \qquad \||Ax| - |Ax_0|\|_2^2 \le \epsilon^2, \tag{4.20}$$

where ϵ is the error bound. Here signals and matrices are all restricted to the real number field. The next theorem tells under what conditions the solution to (4.20) is stable.

Theorem 4.1 *Assume that $D \in \mathbb{R}^{n\times N}$ is a tight frame and $x_0 \in D\Sigma_k^N$. The matrix $A \in \mathbb{R}^{m\times n}$ satisfies the S-DRIP of order tk and level $\theta_-, \theta_+ \in (0, 2)$, with*

$$t \ge \max\{\frac{1}{2\theta_- - \theta_-^2}, \frac{1}{2\theta_+ - \theta_+^2}\}.$$

Then the solution $\hat{x}$ to (4.20) satisfies

$$\min\{\|\hat{x} - x_0\|_2, \|\hat{x} + x_0\|_2\} \le c_1\epsilon + c_2 \frac{2\sigma_k(D^* x_0)_1}{\sqrt{k}},$$

where $c_1 = \frac{\sqrt{2(1+\delta)}}{1-\sqrt{t/(t-1)}\delta}$, $c_2 = \frac{\sqrt{2}\delta+\sqrt{t(\sqrt{(t-1)/t}-\delta)\delta}}{t(\sqrt{(t-1)/t}-\delta)}+1$. *Here* δ *is a constant satisfying*

$$\delta \le \max\{1-\theta_-, \theta_+ - 1\} \le \sqrt{\frac{t-1}{t}}.$$

We first give a more general lemma, which is the key to prove Theorem 4.1.

Lemma 4.1 *Let* $D \in \mathbb{R}^{n\times N}$ *be an arbitrary tight frame,* $x_0 \in D\Sigma_k^N$ *and* $\rho \ge 0$. *Suppose that* $A \in \mathbb{R}^{m\times n}$ *is a measurement matrix satisfying the DRIP with* $\delta = \delta_{tk}^A \le \sqrt{\frac{t-1}{t}}$ *for some* $t > 1$. *Then for any*

$$D^*\hat{x} \in \{D^*x \in \mathbb{R}^N : \|D^*x\|_1 \le \|D^*x_0\|_1 + \rho,\ \|Ax - Ax_0\|_2 \le \epsilon\},$$

we have

$$\|\hat{x} - x_0\|_2 \le c_1\epsilon + c_2\frac{2\sigma_k(D^*x_0)_1}{\sqrt{k}} + c_2\cdot\frac{\rho}{\sqrt{k}},$$

where $c_1 = \frac{\sqrt{2(1+\delta)}}{1-\sqrt{t/(t-1)}\delta}$, $c_2 = \frac{\sqrt{2}\delta+\sqrt{t(\sqrt{(t-1)/t}-\delta)\delta}}{t(\sqrt{(t-1)/t}-\delta)} + 1.$

We put the proof of this Lemma in the Appendix.

Remark 4.1 When $D = I$, which corresponds to the case of standard compressive phase retrieval, Theorem 4.1 and Lemma 4.1 are consistent with Theorem 3.1 and Lemma 2.1 in [12], respectively.

Remark 4.2 The DRIP constant in Lemma 4.1 is better than the DRIP constants given in [2] and [7]. In [7], Candès et al. proved that the l_1-analysis (1.3) can guarantee a stable recovery of signals which are k-sparse in the tight frame D provided the measurement matrix satisfying DRIP with $\delta_{2k} < 0.08$. Then Baker improved the result by increasing the DRIP constant to $\delta_{2k} < \frac{2}{3}$ in [2]. Here we extended Baker's approach to get a better bound $\delta_{tk} \le \sqrt{\frac{t-1}{t}}$ for $t > 1$. As [6] shows, in the special case $D = I$, for any $t \ge 4/3$, the condition $\delta_{tk} \le \sqrt{\frac{t-1}{t}}$ is sharp for stable recovery in the noisy case. So it is not difficult to conclude that for any tight frame D, the condition $\delta_{tk} \le \sqrt{\frac{t-1}{t}}$ is also sharp when $t \ge 4/3$.

Proof of the Theorem 4.1 As $\hat{x}$ is the solution to (4.20), we have

$$\|D^*\hat{x}\|_1 \le \|D^*x_0\|_1 \tag{4.21}$$

and

$$\||A\hat{x}| - |Ax_0|\|_2^2 \le \epsilon^2. \tag{4.22}$$

Denote $a_j^\top$, $j \in \{1,\dots,m\}$ as the rows of A and divide $\{1,\dots,m\}$ into two groups:

$$T = \{j \mid \text{sign}(\langle a_j, \hat{x}\rangle) = \text{sign}(\langle a_j, x_0\rangle)\},$$

$$T^c = \{j \mid \text{sign}(\langle a_j, \hat{x}\rangle) = -\text{sign}(\langle a_j, x_0\rangle)\}.$$

Then either $|T| \geq m/2$ or $|T^c| \geq m/2$. Without loss of generality, we suppose $|T| \geq m/2$.
Then (4.22) implies that

$$\|A_T\hat{x} - A_T x_0\|_2^2 \leq \|A_T\hat{x} - A_T x_0\|_2^2 + \|A_{T^c}\hat{x} + A_{T^c}x_0\|_2^2 \leq \epsilon^2. \tag{4.23}$$

Combining (4.21) and (4.23), we have

$$D^*\hat{x} \in \{D^*x \in \mathbb{R}^N : \|D^*x\|_1 \leq \|D^*x_0\|_1, \|A_T x - A_T x_0\|_2 \leq \epsilon\}. \tag{4.24}$$

Recall that A satisfies S-DRIP of order tk with constants $\theta_-,\ \theta_+ \in (0, 2)$. Here

$$t \geq \max\left\{\frac{1}{2\theta_- - \theta_-^2}, \frac{1}{2\theta_+ - \theta_+^2}\right\} > 1.$$

So A_T satisfies DRIP of order tk with

$$\delta_{tk}^{A_T} \leq \max\{1 - \theta_-,\ \theta_+ - 1\} \leq \sqrt{\frac{t-1}{t}}. \tag{4.25}$$

Combining (4.24), (4.25) and Lemma 4.1, we obtain

$$\|\hat{x} - x_0\|_2 \leq c_1\epsilon + c_2\frac{2\sigma_k(D^*x_0)_1}{\sqrt{k}},$$

where c_1 and c_2 are defined as before in the Theorem 4.1.
If $|T^c| \geq \frac{m}{2}$, we can get the corresponding result

$$\|\hat{x} + x_0\|_2 \leq c_1\epsilon + c_2\frac{2\sigma_k(D^*x_0)_1}{\sqrt{k}}.$$

Then we have proved the theorem. □

According to Theorem 4.1, when $\epsilon = 0$ and D^*x_0 is k-sparse, the ℓ_1-analysis (4.20) can provide an exact recovery of the phase retrieval with redundant dictionary (1.2) provided the measurement matrix satisfies S-DRIP. Meanwhile, from Theorem 4.1 and Corollary 4.1, we conclude that the ℓ_1-analysis (4.20) can provide a stable solution to problem (1.2) if we use as many as $\mathcal{O}(k\log(n/k))$ Gaussian random measurements.

5 Discussion

To solve the phase retrieval with redundant dictionary (1.2), we analyze the ℓ_1-analysis model and give two conditions on the measurement matrix that each of them can guarantee an exact recovery in noiseless case. Theorems 3.1 and 3.2 give the null space property as a sufficient and necessary condition for exact recovery. For the ℓ_1-synthesis model, we can also use the same analysis to give a null space property of the measurement matrix. A more detailed description of the ℓ_1-synthesis model is provided in [8]. Theorem 4.1 shows that the ℓ_1-analysis model is accurate when the measurement matrix satisfies S-DRIP and $\|D^* x_0\|_0 \le k$. In theory, the ℓ_1-analysis model has a good performance on solving the phase retrieval with redundant dictionary (1.2). However, the ℓ_1-analysis is a non-convex optimization for phase retrieval with redundant dictionary due to the non-convex feasible solution set. When $D = I$, the algorithms of this model have been studied in [15,17,22]. These algorithms all demonstrate empirical success, but the convergence issue remains a difficult problem. Extending these algorithms to a redundant dictionary D and giving a convergence analysis is one direction of our future research. Another key drawback of our results is that Theorem 4.1 only holds in the real number field. The point is that the phase changes continuously and there is no proper definition of SRIP in the complex number field. The extension of this result to complex number field is another direction of our future work.

Acknowledgements My deepest gratitude goes to Professor Zhiqiang Xu, my academic supervisor, for his guidance and many useful discussions.

Appendix

The following two lemmas are useful in the proof of Lemma 4.1.

Lemma 6.1 (Sparse representation of a polytope [6,21]): *Suppose $\alpha > 0$ is a constant and $s > 0$ is an integer. Set*

$$T(\alpha, s) := \{v \in \mathbb{R}^n : \|v\|_\infty \le \alpha,\ \|v\|_1 \le s\alpha\}.$$

For any $v \in \mathbb{R}^n$, set

$$U(\alpha, s, v) := \{u \in \mathbb{R}^n : \text{supp}(u) \subseteq \text{supp}(v),\ \|u\|_0 \le s,\ \|u\|_1 = \|v\|_1,\ \|u\|_\infty \le \alpha\}.$$

Then $v \in T(\alpha, s)$ if and only if v is in the convex hull of $U(\alpha, s, v)$. In particular, any $v \in T(\alpha, s)$ can be expressed as

$$v = \sum_{i=1}^{M} \lambda_i u_i \quad \text{and} \quad 0 \le \lambda_i \le 1,\ \sum_{i=1}^{M} \lambda_i = 1,$$
$$u_i \in U(\alpha, s, v).$$

Lemma 6.2 (Lemma 5.3 in [5]): *Suppose $m \ge r$, $a_1 \ge a_2 \ge \cdots \ge a_m \ge 0$ and $\sum_{i=1}^{r} a_i \ge \sum_{i=r+1}^{m} a_i$. Then for all $\alpha \ge 1$, we have*

$$\sum_{j=r+1}^{m} a_j^\alpha \le \sum_{i=1}^{r} a_i^\alpha.$$

Now we are ready to prove Lemma 4.1.

Proof of the Lemma 4.1 We assume that the tight frame $D \in \mathbb{R}^{n\times N}$ is normalized, i.e., $DD^* = I$ and $\|y\|_2 = \|D^* y\|_2$ for all $y \in \mathbb{R}^n$. For a subset $T \subseteq \{1, 2, \ldots, N\}$, we denote D_T as the matrix D restricted to the columns indexed by T (replacing other columns by zero vectors).

Set $h := \hat{x} - x_0$. Let T_0 denote the index set of the largest k coefficients of $D^* x_0$ in magnitude. Then

$$\begin{aligned}\|D^* x_0\|_1 + \rho \geq \|D^* \hat{x}\|_1 &= \|D^* x_0 + D^* h\|_1\\ &= \|D^*_{T_0} x_0 + D^*_{T_0} h + D^*_{T_0^c} x_0 + D^*_{T_0^c} h\|_1\\ &\geq \|D^*_{T_0} x_0\|_1 - \|D^*_{T_0} h\|_1 - \|D^*_{T_0^c} x_0\|_1 + \|D^*_{T_0^c} h\|_1,\end{aligned}$$

which implies

$$\begin{aligned}\|D^*_{T_0^c} h\|_1 &\leq \|D^*_{T_0} h\|_1 + 2\|D^*_{T_0^c} x_0\|_1 + \rho\\ &= \|D^*_{T_0} h\|_1 + 2\sigma_k(D^* x_0)_1 + \rho.\end{aligned}$$

Suppose S_0 is the index set of the k largest entries in absolute value of $D^* h$. We get

$$\begin{aligned}\|D^*_{S_0^c} h\|_1 \leq \|D^*_{T_0^c} h\|_1 &\leq \|D^*_{T_0} h\|_1 + 2\sigma_k(D^* x_0)_1 + \rho\\ &\leq \|D^*_{S_0} h\|_1 + 2\sigma_k(D^* x_0)_1 + \rho.\end{aligned}$$

Set

$$\alpha := \frac{\|D^*_{S_0} h\|_1 + 2\sigma_k(D^* x_0)_1 + \rho}{k}.$$

We divide $D^*_{S_0^c} h$ into two parts $D^*_{S_0^c} h = h^{(1)} + h^{(2)}$, where

$$h^{(1)} := D^*_{S_0^c} h \cdot I_{\{i:\, |D^*_{S_0^c} h(i)| > \alpha/(t-1)\}}, \quad h^{(2)} := D^*_{S_0^c} h \cdot I_{\{i:\, |D^*_{S_0^c} h(i)| \leq \alpha/(t-1)\}}.$$

Then a simple observation is that $\|h^{(1)}\|_1 \leq \|D^*_{S_0^c} h\|_1 \leq \alpha k$. Set

$$\ell := |\mathrm{supp}(h^{(1)})| = \|h^{(1)}\|_0.$$

Since all non-zero entries of $h^{(1)}$ have magnitude larger than $\alpha/(t-1)$, we have

$$\alpha k \geq \|h^{(1)}\|_1 = \sum_{i \in \mathrm{supp}(h^{(1)})} |h^{(1)}(i)| \geq \sum_{i \in \mathrm{supp}(h^{(1)})} \frac{\alpha}{t-1} = \ell \cdot \frac{\alpha}{t-1},$$

which implies $\ell \leq (t-1)k$.

Note that

$$\|h^{(2)}\|_1 = \|D^*_{S_0^c}h\|_1 - \|h^{(1)}\|_1 \le k\alpha - \ell \cdot \frac{\alpha}{t-1} = (k(t-1)-\ell)\frac{\alpha}{t-1},$$
$$\|h^{(2)}\|_\infty \le \frac{\alpha}{t-1}.$$

Then in Lemma 6.1, by setting $s := k(t-1)-\ell$, we can express $h^{(2)}$ as a weighted mean:

$$h^{(2)} = \sum_{i=1}^{M} \lambda_i u_i,$$

where $0 \le \lambda_i \le 1$, $\sum_{i=1}^M \lambda_i = 1$, $\|u_i\|_0 \le k(t-1)-\ell$, $\|u_i\|_\infty \le \alpha/(t-1)$ and $\mathrm{supp}(u_i) \subseteq \mathrm{supp}(h^{(2)})$. Thus

$$\begin{aligned}\|u_i\|_2 \le \sqrt{\|u_i\|_0} \cdot \|u_i\|_\infty &= \sqrt{k(t-1)-\ell} \cdot \|u_i\|_\infty \\ &\le \sqrt{k(t-1)} \cdot \|u_i\|_\infty \\ &\le \alpha\sqrt{k/(t-1)}.\end{aligned}$$

Recall that $\alpha = \frac{\|D^*_{S_0}h\|_1 + 2\sigma_k(D^*x_0)_1 + \rho}{k}$. Then

$$\begin{aligned}\|u_i\|_2 &\le \alpha\sqrt{k/(t-1)} \\ &\le \frac{\|D^*_{S_0}h\|_2}{\sqrt{t-1}} + \frac{2\sigma_k(D^*x_0)_1 + \rho}{\sqrt{k(t-1)}} \\ &\le \frac{\|D^*_{S_0}h + h^{(1)}\|_2}{\sqrt{t-1}} + \frac{2\sigma_k(D^*x_0)_1 + \rho}{\sqrt{k(t-1)}} \\ &= \frac{z+R}{\sqrt{t-1}},\end{aligned} \tag{6.26}$$

where $z := \|D^*_{S_0}h + h^{(1)}\|_2$, $R := \frac{2\sigma_k(D^*x_0)_1+\rho}{\sqrt{k}}$.

Now we suppose $0 \le \mu \le 1$, $d \ge 0$ are two constants to be determined. Set

$$\beta_j := D^*_{S_0}h + h^{(1)} + \mu \cdot u_j, \quad j = 1, \ldots, M.$$

Then for any fixed $i \in [1 : M]$,

$$\begin{aligned}\sum_{j=1}^{M} \lambda_j \beta_j - d\beta_i &= D^*_{S_0}h + h^{(1)} + \mu \cdot h^{(2)} - d\beta_i \\ &= (1-\mu-d)(D^*_{S_0}h + h^{(1)}) - d\mu u_i + \mu D^* h.\end{aligned} \tag{6.27}$$

For $\sum_{i=1}^M \lambda_i = 1$, we have the following identity

$$(2d-1)\sum_{1\le i<j\le M}\lambda_i\lambda_j\|AD(\beta_i-\beta_j)\|_2^2$$
$$=\sum_{i=1}^M\lambda_i\|AD(\sum_{j=1}^M\lambda_j\beta_j-d\beta_i)\|_2^2-\sum_{i=1}^M\lambda_i(1-d)^2\|AD\beta_i\|_2^2. \quad (6.28)$$

In (6.27), we chose $d=1/2$ and $\mu=\sqrt{t(t-1)}-(t-1)<1/2$. Then (6.28) implies

$$0=\sum_{i=1}^M\lambda_i\|AD(\sum_{j=1}^M\lambda_j\beta_j-d\beta_i)\|_2^2-\sum_{i=1}^M\frac{\lambda_i}{4}\|AD\beta_i\|_2^2$$
$$\overset{(2.27)}{=}\sum_{i=1}^M\lambda_i\|AD\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i+\mu D^*h\right)\|_2^2-\sum_{i=1}^M\frac{\lambda_i}{4}\|AD\beta_i\|_2^2$$
$$=\sum_{i=1}^M\lambda_i\|AD\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i\right)\|_2^2$$
$$+2\left\langle AD\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}h^{(2)}\right),\mu ADD^*h\right\rangle+\mu^2\|ADD^*h\|_2^2$$
$$-\sum_{i=1}^M\frac{\lambda_i}{4}\|AD\beta_i\|_2^2$$
$$=\sum_{i=1}^M\lambda_i\|AD\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i\right)\|_2^2 \quad (6.29)$$
$$+\mu(1-\mu)\left\langle AD(D_{S_0}^*h+h^{(1)}),ADD^*h\right\rangle-\sum_{i=1}^M\frac{\lambda_i}{4}\|AD\beta_i\|_2^2.$$

We next estimate the three terms in (6.29). First we give the following useful relation:

$$\left\langle D(D_{S_0}^*h+h^{(1)}),Dh^{(2)}\right\rangle$$
$$=\left\langle D(D_{S_0}^*h+h^{(1)}),D(D^*h-D_{S_0}^*h-h^{(1)})\right\rangle$$
$$=\left\langle D(D_{S_0}^*h+h^{(1)}),h\right\rangle-\left\langle D(D_{S_0}^*h+h^{(1)}),D(D_{S_0}^*h+h^{(1)})\right\rangle$$
$$=\left\langle D_{S_0}^*h+h^{(1)},D^*h\right\rangle-\|D(D_{S_0}^*h+h^{(1)})\|_2^2$$
$$=\|D_{S_0}^*h+h^{(1)}\|_2-\|D(D_{S_0}^*h+h^{(1)})\|_2^2. \quad (6.30)$$

Noting that $\|D_{S_0}^*h\|_0\le k$, $\|h^{(1)}\|_0=\ell\le(t-1)k$ and $\|u_i\|_0\le s=k(t-1)-\ell$, we obtain

$$\|D_{S_0}^*h+h^{(1)}\|_0\le\ell+k\le t\cdot k,\quad \|\beta_i\|_0\le\|D_{S_0}^*h\|_0+\|h^{(1)}\|_0+\|u_i\|_0\le t\cdot k,$$

and

$$\|(\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i\|_0 \le t\cdot k.$$

Here we assume $t \cdot k$ as an integer first. Since A satisfies the DRIP of order $t \cdot k$ with constant δ, we can obtain

$$\begin{aligned}
&\sum_{i=1}^{M}\lambda_i\|AD\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i\right)\|_2^2\\
&\le \sum_{i=1}^{M}\lambda_i(1+\delta)\|D\left((\frac{1}{2}-\mu)(D_{S_0}^*h+h^{(1)})-\frac{\mu}{2}u_i\right)\|_2^2\\
&= (1+\delta)\Bigg((\frac{1}{2}-\mu)^2\|D(D_{S_0}^*h+h^{(1)})\|_2^2+\frac{\mu^2}{4}\sum_{i=1}^{M}\lambda_i\|Du_i\|_2^2\\
&\quad -\mu(\frac{1}{2}-\mu)\left\langle D(D_{S_0}^*h+h^{(1)}), Dh^{(2)}\right\rangle\Bigg)\\
&\overset{(7.30)}{=} (1+\delta)\Bigg(\frac{1}{2}(\frac{1}{2}-\mu)\|D(D_{S_0}^*h+h^{(1)})\|_2^2+\frac{\mu^2}{4}\sum_{i=1}^{M}\lambda_i\|Du_i\|_2^2\\
&\quad -\mu(\frac{1}{2}-\mu)\|D_{S_0}^*h+h^{(1)}\|_2^2\Bigg),
\end{aligned}$$

$$\begin{aligned}
\left\langle AD(D_{S_0}^*h+h^{(1)}), ADD^*h\right\rangle &= \left\langle AD(D_{S_0}^*h+h^{(1)}), Ah\right\rangle\\
&\le \sqrt{1+\delta}\cdot\|D(D_{S_0}^*h+h^{(1)})\|_2\cdot\epsilon
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{M}\lambda_i\|AD\beta_i\|_2^2\\
&= \sum_{i=1}^{M}\lambda_i\|AD(D_{S_0}^*h+h^{(1)}+\mu\cdot u_i)\|_2^2\\
&\ge (1-\delta)\sum_{i=1}^{M}\lambda_i\|D(D_{S_0}^*h+h^{(1)}+\mu\cdot u_i)\|_2^2\\
&= (1-\delta)\left(\|D(D_{S_0}^*h+h^{(1)})\|_2^2+\mu^2\sum_{i=1}^{M}\lambda_i\|Du_i\|_2^2+2\mu\left\langle D(D_{S_0}^*h+h^{(1)}), Dh^{(2)}\right\rangle\right)\\
&\overset{(7.30)}{=} (1-\delta)\left((1-2\mu)\|D(D_{S_0}^*h+h^{(1)})\|_2^2+\mu^2\sum_{i=1}^{M}\lambda_i\|Du_i\|_2^2+2\mu\|D_{S_0}^*h+h^{(1)}\|_2^2\right).
\end{aligned}$$

Combining the above results with (6.26) and (6.29), we get

$$
\begin{aligned}
0 &\le \frac{1}{2}(1+\delta)\left(\frac{1}{2}-\mu\right)\|D(D_{S_0}^*h+h^{(1)})\|_2^2+\frac{1+\delta}{4}\mu^2\sum_{i=1}^M\lambda_i\|Du_i\|_2^2\\
&\quad-(1+\delta)\mu\left(\frac{1}{2}-\mu\right)\|D_{S_0}^*h+h^{(1)}\|_2^2\\
&\quad+\mu(1-\mu)\sqrt{1+\delta}\|D(D_{S_0}^*h+h^{(1)})\|_2\cdot\epsilon\\
&\quad-\frac{1}{4}(1-\delta)(1-2\mu)\|D(D_{S_0}^*h+h^{(1)})\|_2^2-\frac{1-\delta}{4}\mu^2\sum_{i=1}^M\lambda_i\|Du_i\|_2^2\\
&\quad-\frac{1-\delta}{2}\mu\|D_{S_0}^*h+h^{(1)}\|_2^2\\
&=\delta(\frac{1}{2}-\mu)\|D(D_{S_0}^*h+h^{(1)})\|_2^2+(\mu^2(1+\delta)-\mu)\|D_{S_0}^*h+h^{(1)}\|_2^2\\
&\quad+\frac{\delta}{2}\mu^2\sum_{i=1}^M\lambda_i\|Du_i\|_2^2+\mu(1-\mu)\sqrt{1+\delta}\|D(D_{S_0}^*h+h^{(1)})\|_2\cdot\epsilon\\
&\le(\delta(\frac{1}{2}-\mu)+\mu^2(1+\delta)-\mu)z^2+\frac{\delta}{2}\mu^2\sum_{i=1}^M\lambda_i\|u_i\|_2^2+\mu(1-\mu)\sqrt{1+\delta}\cdot z\cdot\epsilon\\
&\overset{(7.26)}{\le}\left((1+\delta)(\frac{1}{2}-\mu)^2-\frac{1-\delta}{4}\right)z^2+\frac{\delta}{2}\mu^2\frac{(z+R)^2}{t-1}+\mu(1-\mu)\sqrt{1+\delta}\cdot z\cdot\epsilon\\
&=\left((\mu^2-\mu)+\delta\left(\frac{1}{2}-\mu+(1+\frac{1}{2(t-1)})\mu^2\right)\right)z^2\\
&\quad+\left(\mu(1-\mu)\sqrt{1+\delta}\cdot\epsilon+\frac{\delta\mu^2R}{t-1}\right)z+\frac{\delta\mu^2R^2}{2(t-1)}\\
&=-t\left((2t-1)-2\sqrt{t(t-1)}\right)(\sqrt{\frac{t-1}{t}}-\delta)z^2\\
&\quad+\left(\mu^2\sqrt{\frac{t}{t-1}}\sqrt{1+\delta}\cdot\epsilon+\frac{\delta\mu^2R}{t-1}\right)z+\frac{\delta\mu^2R^2}{2(t-1)}\\
&=\frac{\mu^2}{t-1}\left(-t(\sqrt{\frac{t-1}{t}}-\delta)z^2+(\sqrt{t(t-1)(1+\delta)}\epsilon+\delta R)z+\frac{\delta R^2}{2}\right),
\end{aligned}
$$

which is a quadratic inequality for z. Recall that $\delta < \sqrt{(t-1)/t}$. So by solving the above inequality, we get

$$
\begin{aligned}
z &\le \frac{(\sqrt{t(t-1)(1+\delta)}\epsilon+\delta R)+\left((\sqrt{t(t-1)(1+\delta)}\epsilon+\delta R)^2+2t(\sqrt{(t-1)/t}-\delta)\delta R^2\right)^{1/2}}{2t(\sqrt{(t-1/t)}-\delta)}\\
&\le \frac{\sqrt{t(t-1)(1+\delta)}}{t(\sqrt{(t-1)/t}-\delta)}\epsilon+\frac{2\delta+\sqrt{2t(\sqrt{(t-1)/t}-\delta)\delta}}{2t(\sqrt{(t-1)/t}-\delta)}R.
\end{aligned}
$$

We know $\|D^*_{S_0^c}h\|_1 \leq \|D^*_{S_0}h\|_1 + R\sqrt{k}$. In the Lemma 6.2, if we set $m = N$, $r = k$, $\lambda = R\sqrt{k} \geq 0$ and $\alpha = 2$, we can obtain

$$\|D^*_{S_0^c}h\|_2 \leq \|D^*_{S_0}h\|_2 + R.$$

So

$$\begin{aligned}
\|h\|_2 &= \|D^*h\|_2 \\
&= \sqrt{\|D^*_{S_0}h\|_2^2 + \|D^*_{S_0^c}h\|_2^2} \\
&\leq \sqrt{\|D^*_{S_0}h\|_2^2 + (\|D^*_{S_0}h\|_2 + R)^2} \\
&\leq \sqrt{2\|D^*_{S_0}h\|_2^2} + R \leq \sqrt{2}z + R \\
&\leq \frac{\sqrt{2(1+\delta)}}{1-\sqrt{t/(t-1)}\delta}\epsilon + \left(\frac{\sqrt{2}\delta + \sqrt{t(\sqrt{(t-1)/t}-\delta)\delta}}{t(\sqrt{(t-1)/t}-\delta)} + 1\right)R.
\end{aligned}$$

Substituting R into this inequality, we can get the conclusion. For the case where $t \cdot k$ is not an integer, we set $t^* := \lceil tk \rceil / k$, then $t^* > t$ and $\delta_{t^*k} = \delta_{tk} < \sqrt{\frac{t-1}{t}} < \sqrt{\frac{t^*-1}{t^*}}$. We can prove the result by working on δ_{t^*k}. □

References

1. Aldroubi, A., Chen, X., Powell, A.M.: Perturbations of measurement matrices and dictionaries in compressed sensing. Appl. Comput. Harmon. Anal. **33**(2), 282–291 (2012)
2. Baker, C.A.: A note on sparsification by frames. arXiv:1308.5249 (2013)
3. Balan, R., Casazza, P., Edidin, D.: On signal reconstruction without phase. Appl. Comput. Harmon. Anal. **20**(3), 345–356 (2006)
4. Bodmann, B.G., Hammen, N.: Stable phase retrieval with low-redundancy frame. Adv. Comput. Math. **41**(2), 317–331 (2015)
5. Cai, T.T., Zhang, A.: Sharp RIP bound for sparse signal and low-rank matrix recovery. Appl. Comput. Harmon. Anal. **35**(1), 74–93 (2013)
6. Cai, T.T., Zhang, A.: Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. IEEE Trans. Inf. Theory **60**(1), 122–132 (2014)
7. Candès, E.J., Eldar, Y.C., Needell, D., Randall, P.: Compressed sensing with coherent and redundant dictionaries. Appl. Comput. Harmon. Anal. **31**(1), 59–73 (2011)
8. Chen, X.M., Wang, H.C., Wang, R.R.: A null space analysis of the ℓ_1-synthesis method in dictionary-based compressed sensing. Appl. Comput. Harmon. Anal. **37**(3), 492–515 (2014)
9. Conca, A., Edidin, D., Hering, M., Vinzant, C.: An algebraic characterization of injectivity in phase retrieval. Appl. Comput. Harmon. Anal. **38**(2), 346–356 (2015)
10. Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. Inverse Probl. **23**(3), 947968 (2007)
11. Fienup, J.: Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. J. Opt. Soc. Am. **4**(1), 118–123 (1987)
12. Gao, B., Wang, Y., Xu, Z.Q.: Stable signal recovery from phaseless measurements. J. Fourier Anal. Appl. **22**(4), 787–808 (2016)
13. Herman, M., Strohmer, T.: High resolution radar via compressed sensing. IEEE Trans. Signal Process. **57**(6), 22752284 (2009)
14. Kohler, D., Mandel, L.: Source reconstruction from the modulus of the correlation function: a practical approach to the phase problem of optical coherence theory. J. Opt. Soc. Am. **63**(2), 126–134 (1973)

15. Moravec, M., Romberg, J., Baraniuk, R.: Compressive phase retrieval. Proc. SPIE **6701**, 670120–670120 (2007). doi:10.1117/12.736360
16. Rauhut, H., Schnass, K., Vandergheynst, P.: Compressed sensing and redundant dictionaries. IEEE Trans. Inf. Theory **54**(5), 2210–2219 (2008)
17. Schniter, P., Rangan, S.: Compressive phase retrieval via generalized approximate message passing. IEEE Trans. Signal Process. **63**(4), 1043–1055 (2015)
18. Shechtman, Y., Eldar, Y.C., Szameit, A., Segev, M.: Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing. Opt. Express **19**(16), 14807–14822 (2011)
19. Voroninski, V., Xu, Z.Q.: A strong restricted isometry property, with an application to phaseless compressed sensing. Appl. Comput. Harmon. Anal. **40**(2), 386–395 (2016)
20. Wang, Y., Xu, Z.Q.: Phase retrieval for sparse signals. Appl. Comput. Harmon. Anal. **37**(3), 531–544 (2014)
21. Xu, G.W., Xu, Z.Q.: On the ℓ_1-norm invariant convex k-sparse decomposition of signals. J. Oper. Res. Soc. China **1**(4), 537–541 (2013)
22. Yang, Z., Zhang, C.S., Xie, L.H.,: Robust compressive phase retrieval via L1 minimization with application to image reconstruction. arXiv:1302.0081 (2013)