Noise Folding based on General Complete Perturbation in Compressed Sensing

Limin Zhou*, Xinxi Niu†, Jing Yuan‡
*†‡Information Security Center, Beijing University of Posts and Telecommunications
Beijing 100876, China
Email: zhoulimin.s@163.com.xxniu@bupt.edu.cn.

Abstract

This paper first presents a new general completely perturbed compressed sensing (CS) model \( y=(A+E)(x+u)+e \), called noise folding based on general completely perturbed CS system, where \( y \in \mathbb{R}^m \), \( u \in \mathbb{R}^m \), \( u \neq 0 \), \( e \in \mathbb{R}^m \), \( A \in \mathbb{R}^{m \times n} \), \( m \ll n \), \( E \in \mathbb{R}^{m \times n} \) with incorporating general nonzero perturbation \( E \) to sensing matrix \( A \) and noise \( u \) into signal \( x \) simultaneously based on the standard CS model \( y=Ax+e \). Our constructions mainly will whiten the new proposed CS model and explore into RIP, coherence for \( A+E \) of the new CS model after being whitened.

Index Terms

Compressed Sensing (CS), general perturbation, restricted isometry property (RIP), coherence

I. INTRODUCTION

COMPRESSED sensing (CS) model, which was proposed by Candes etc[13] and Donoho[14], has become a hot topic so as to attract a lot of researcher to study it over the past years because it can recover a signal as a technique. Thus it has been widely applied in many areas such as radar systems[15] signal processing[16], image processing[17] etc. These applications depend on the main function of CS model to recover the original signal with some related algorithms including convex relaxation[12][20] and greedy pursuits [20], which estimates the best approximation values of the original signal.

The classic and basic CS model in an unperturbed scenario can be formulated as

\[ y = Ax \] (1)

Here \( y \in \mathbb{R}^n \) is the measurement vector or observation value, \( A \in \mathbb{R}^{m \times n} \) is a full rank measurement matrix with \( m \ll n \). The signal \( x \in \mathbb{R}^n \) is assumed to be \( k \)-sparse that is no more than \( k \) entries of \( x \) are nonzero that \( x \) is called a \( k \)-sparse signal. We will assume throughout that measurement matrix \( A \in \mathbb{R}^{m \times n} \) with \( m \ll n \).

Roughly speaking, the basic model has mature theory and utilized in many areas[1][2][3][4] and there are a lot of different algorithms introduced in[12][20] such as match pursuit (BP)[6][21] and orthogonal match pursuit (OMP)[18][22][23][24][25]. Compressive Sampling matching Pursuit (CoSaMP)[19] and so on.

But in practical applications, the measurement vector \( y \) in (1) is often contaminated by noise or error. More concretely, a noise term \( e \in \mathbb{R}^n \), called an additive noise, is incorporated into \( y = Ax + e \) to result in a partially perturbed model[5][6][7]

\[ y = Ax + e \] (2)

where noise or error \( e \) is uncorrelated with signal \( x \). There are two methods to model noise \( e \) mentioned in [8]. Here, noise \( e \) is randomly sampling from Gaussian distribution. This model is used in many areas[5][6][7] and naturally has more mature theory in recent years. A number of concrete recovery accuracy algorithms on (2) have emerged e.g.BP[5][6], OMP[5], CoSaMP[19], e.t.c. in recent years.

In 2010, Matthew A.Herman et al in [9] first incorporated an unknown nontrivial random perturbation \( E \) into matrix \( A \) in (2) leading to general completely perturbed model[9][10][11] with \( E \neq 0 \), \( e \neq 0 \) as follows

\[ y = (A+E)x + e \] (3)

where \( E \in \mathbb{R}^{m \times n} \) is called general perturbation or multiplicative noise. They studied influence of \( E \) on signal \( x \) and other related theory indicating that it is a must to consider this noise[9][10][11]. However, intuitively, it is more harder to analyze the multiplicative noise \( E \) than the additive noise \( e \) because \( E \) is related to the signal \( x \) with \( Ex \).

As for (3), there are two different scenarios from different perspectives of views[9][10][11]. The first is from user’s perspective of view, measuring an undiscovered model to get its inaccurate matrix. The sensing process can be formulated as \( \hat{y} = Ax + e \), \( \hat{A} = A + E \)

corresponding to recovery process with

\[ (N_1) \quad \hat{x} = (\hat{y}, \hat{A}, \cdots) \]
Thus the useful measurement matrix is the perturbed matrix $\hat{A}$ not the original measurement matrix $A$. The system have been researched on the recovery signal with BP by Matthew A. Herman et al in [9] [19] and with OMP by Jie Dinng, etc. in [10][11].

The second model is from designer’s perspective[9][10][11]. The sensing process is just as
\[
\hat{y} = \hat{A}x + e, \quad \hat{A} = A + E
\]
and the recovery process is as
\[
(N_1') \quad \hat{x} = R(\hat{y}, A, \cdots)
\]
The useful sensing matrix is $A$ not $\hat{A}$ and the observation value is $\hat{y}$ not the observation value without perturbation $E$. To our best knowledge, no works focus on the recovery signal in the context of general perturbation $E$ except for [9] [10][11].

But in some practical applications, signal itself is often contaminated by noise, one of such cases is applied in sub-Nyquist converter. Though introducing noise to signal is significant, no prolific papers studied such signal noise $u$ except for [8] in 2011, first adding an unknown random noise $u \in R^n$ to sinal $x$ based on $y = Ax + e$ to produce noise folding CS model[8]
\[
y = A(x + u) + e
\]
They analyze the RIP and coherence of the equivalent system after whitening and show that the difference of the RIP and coherence between original $A$ and whiten matrix is small. The related conclusion about this model can be seen in [8]. Based on the theory [8][9][10][11], we proposed a new CS model and study its related properties.

Our new CS model

As mentioned above, note that as for (2)(3)(8), only one noise e.g. noise $e$ or noise folding $u$ or perturbation $E$ affect the CS model. Maybe noise $e$, noise folding $u$ and perturbation $E$ simultaneously affect the CS, although no paper studied this. Based on this new idea, together [8] with [9][10][11] motivate us to introduce noise $u$ to general completely perturbed model (3) to result in the noise folding in general completely perturbed situation or to incorporate nontrivial perturbation $E$ into (8) to produce completely perturbed model with folding noise in CS, of which for the first time yield so called the folding noise-general completely perturbed CS model to be formulated as
\[
y = (A + E)(x + u) + e
\]
Assume that $e \in R^n$ is a random noise vector with covariance $\sigma^2 I$ and $u \in R^n$ present random pre-measurement noise vector whose covariance is $\sigma_0^2 I$ independent of $e$. Here $e$ and $u$ is regarded as additive noise. $E \in R^{m \times n}$ is random matrix and more details on perturbation $E$ can be seen in [9]. It’s nature that we call CS model (5) noise folding based on complete perturbation in CS model. Analogous to (3) in [9][10][11], (5) can also be considered two different situations. Similarly, from user’s perspective of view, an incorrect sensing matrix can be gotten via unknown measurement model
\[
\hat{y} = A(x + u) + e, \quad \hat{A} = A + E
\]
and the recovery process algorithm proceed as
\[
(N_2) \quad \hat{x} = R(\hat{y}, \hat{A}, u, \cdots)
\]
The only difference between (N_1) and (N_2) is noise $u$ belong to (N_2). From the designer’s view, the sensing process can be formulated as
\[
\hat{y} = \hat{A}(x + u) + e, \quad \hat{A} = A + E
\]
and its recovery process as
\[
(N_2') \quad \hat{x} = R(\hat{y}, \hat{A}, u, \cdots)
\]
Similarly, compared to (N_1') noise $u$ belong to (N_2'). In this paper, we only study simply its properties such as RIP, coherence et al after whitening. Obviously, (5) can be extended to the multi-perturbation general CS model
\[
y = (A + \sum_{i=1}^{n} E_i)(x + u) + e, \quad i = 1, 2, \cdots, n
\]
with $E_i$ is perturbation. The system (6) can be viewed as an generalization of our new proposed CS system, which implies that the general conclusion of (6) can be obtained from the special conclusion of (5). The concrete results can be seen in Appendix. Simultaneously, other general CS systems can be conjectured naturally as follows
\[
y = (\sum_{i=1}^{s} A_i + E)(x + u) + e, \quad y = (\sum_{i=1}^{s} A_i + E)(x + \sum_{i=1}^{s} u_i) + e, \quad y = (\sum_{i=1}^{s} A_i + \sum_{i=1}^{s} E_i)(x + \sum_{i=1}^{s} u_i) + e
\]
Although their properties seem plausible but we don’t know how exploit and analysis them, we leave them as open problems. Here we mainly study relative theory on (5)(6). In section 3, we give the more general results.
II. Preliminaries

In this paper, we will restrict our attention to RIP and coherence, C-stable. By convention, sensing matrix A and perturbation E are assumed to sample independent and identically distributed (i.i.d) Gaussian random variables since such matrix satisfies RIP[20][8] and coherence, C-stable [20] e.t.c with probability one.

Definition 1. [20] A sensing matrix A satisfies the restricted isometry property (RIP) of order k if there exists a δk ∈ (0, 1) s.t.

\[ (1 - \delta_k) \| x \|_2^2 \leq \| Ax \|_2^2 \leq (1 + \delta_k) \| x \|_2^2 \]

for any k-sparse vector with k = 1, 2, 3 ⋯, where δk is the smallest nonnegative number called the restricted isometry constant (RIC).

Definition 1’. [8] For (1) (2), there is another equivalent statement for the RIP’ for A, denoted by RIP’, in some special cases. For any index set Λ ⊂ \{1, ⋯, N\} of size k, let AΛ denote the submatrix of A consisting of the column vectors indexed by Λ, the matrix A possesses RIP’ with constants 0 < αk ≤ βk, if

\[ \alpha_k \| h \|_2^2 \leq \| A_h \|_2^2 \leq \beta_k \| h \|_2^2 \]

for any index set Λ ⊂ \{1, ⋯, N\} of size k where N is a positive integer. For (8), there exist another form of RIP for A in (8) since A is whitened due to signal noise u to signal x that has been given by lemma 2[8].

Lemma 1[8]. As for folding noise model (8), RIP for whitened A can be formulated as

\[ \alpha_k (1 - \rho_1) \| h \|_2^2 \leq B \| h \|_2^2 \leq \beta_k (1 + \rho_1) \| h \|_2^2 \]

where \( \rho_1 = \frac{\sigma_m}{\sigma_0} \) with 0 < \( \rho < \frac{1}{2} \) and B is obtained after whitening sensing matrix A.

The perturbation E and sensing matrix A in (3) can be quantified in [9][10][11] as

\[ \frac{\| E \|_2}{\| A \|_2} \leq \varepsilon_A, \quad \frac{\| E \|_2^{(k)}}{\| A \|_2^{(k)}} \leq \varepsilon_A^{(k)}, \quad \| A \|_2^{(k)} = \sigma_{max}(A) \]

where the symbols \| A \|_2 denotes spectral norm of a matrix A, and \| A \|_2^{(k)} denotes the largest spectral norm taken over all k-column submatrice of matrix A. \( \sigma_{max}(A) \) denote the largest nonzero singular value taken over all k-column submatrice of matrix A. It is appropriate to assume that 0 < \( \varepsilon_A, \varepsilon_A^{(k)} \leq 1 \).

Lemma 2[9]. (RIP for A) For k = 1, 2, ⋯, given the RIC associated with matrix A in (3) and the relative perturbation \( \varepsilon_A^{(k)} \), fix the constant

\[ \delta_{k,max} = (1 + \delta_k)(1 + \varepsilon_A^{(k)})^2 - 1 \]

Assume that the RIC \( \hat{\delta}_k \leq \delta_{k,max} \) for matrix \( \hat{A} = A + E \) is the smallest nonnegative number, the RIP for \( \hat{A} \) can be formulated as

\[ (1 - \hat{\delta}_k) \| x \|_2^2 \leq \| \hat{A}x \|_2^2 \leq (1 + \hat{\delta}_k) \| x \|_2^2 \]

for any k-sparse vector x.

From Lemma 2 with (8), there is another equivalent statement for the RIP for \( \hat{A} \) for some need in some special cases. We give it in Lemma 1’.

Lemma 2’. For any index set \( \Lambda \subset \{1, \cdots, N\} \) of size k, let \( \hat{A}_\Lambda \) denote the submatrix of \( \hat{A} \) consisting of the column vectors indexed by \( \Lambda \). A matrix A possesses RIP with constants 0 < \( \hat{\alpha}_k \leq \hat{\beta}_k \), if

\[ \hat{\alpha}_k \| h \|_2^2 \leq \| \hat{A}_\Lambda h \|_2^2 \leq \hat{\beta}_k \| h \|_2^2, \quad \forall h \in \mathbb{R}^k \]

for any index set \( \Lambda \subset \{1, \cdots, N\} \) of size k where N is a positive integer.

Definition 2[8][20]. The coherence of a matrix A, \( \mu(A) \), is the largest absolute inner product between any two columns \( A_i, A_j, i \neq j \) of matrix A as follows

\[ \mu(A) = \max_{1 \leq i < j \leq n} \frac{| A_i^T A_j |}{\| A_i \|_2 \| A_j \|_2} \]

III. Constructions

A. Problem Formulation

For (5), our goal is to analyse the effect of the pre-measurement noise u and E on the behavior of CS recovery methods with its RIP, coherence.

Throughout this paper, assume that e is a random noise vector with covariance \( \sigma^2 I \), and similarly u is a random noise vector with covariance \( \sigma_u^2 I \) independent with e. Under these assumptions, (5) will be proved to be equivalent to \( y = Bx + w \) where \( B \) is a matrix whose coherence and RIP constants are very close to that of A and w is white noise with variance \( (\sigma^2 + \frac{m}{\delta} \sigma_0^2) I \) where I is identity matrix.
B. Equivalent Formulation

To set up our conclusion, (5) can be expressed as

\[ y = (A + E)x + w \quad \text{with} \quad w = (A + E)u + e \]  \hspace{1cm} (10)

By hypothesis of whiten noise, the covariance of effective vector \( w \) is \( Q \) of which \( Q = \sigma^2 I + \sigma_0^2 (A + E)(A + E)^T \). Obviously, it's easy to see that noise \( w \) is not whiten that the recovery process analysis become complicate. If \( w \) still preserve whitening, one case is that \( A = A + E \) must be proportional to identity matrix. For example, suppose that A+E consists of \( r = n/m \) orthogonal basis such as \( A = [A_1+E_1, A_2+E_2, \ldots, A_r+E_r] \) in which \( A_i+E_i, i=1,2,\ldots, r \), is \( m \times m \) orthogonal matrix. Therefore, we have \((A + E)(A + E)^T = (A_1 + E_1) \cdot (A_1 + E_1)^T + \cdots + (A_r + E_r)(A_r + E_r)^T = rI = \frac{1}{m} I \) and that the noise covariance of \( w \) is \( Q = \gamma I \) with \( \gamma = \sigma^2 + \frac{2}{m^2} \sigma_0^2 \). Under the special case, \( y = (A+E)(x+u)+e \) or \( y = (A+E)x+w \) is equivalent to \( y = Ax+e \). Compared with noise covariance of \( e \), noise covariance of \( w \) had increased by \( \frac{2}{m^2} \) if \( \sigma_0^2 \approx \sigma^2 \), the noise of \( w \) is increased by \( \frac{\sigma_0^2}{m^2} \), which is called noise folding[8].

C. RIP, Coherence

We show that the conclusion holds generally, that is \((A + E)(A + E)^T \) is not proportional to the identity, (5) and (10) are roughly equivalent really. Now we describe it more detail.

We will discuss that if \( E \) are random arbitrary matrix thus \( A+E \) is an arbitrary matrix with low coherence, low RIP and low stable. To study RIP, coherence, we must whiten noise \( w \) by multiplying \( Q_1^{1/2} \), in which \( Q_1 = \frac{Q}{\gamma} \) to get the equivalent system

\[ y = \hat{B}x + v, \quad \text{where} \quad \hat{B} = Q_1^{-1/2} (A + E), \quad v = Q_1^{-1/2} w \]

Note that noise vector \( v \) is whiten with covariance matrix \( \gamma I \) exactly under the context of \((A + E)(A + E)^T \) being proportional to identity matrix. But the biggest difference lies in measurement matrix changing from original matrix \( A+E \) to \( \hat{B} \) by whitening. The changing range is measured through three important indexes: the RIP constant, coherence and stable. Our theory mainly depends on approximating \((A + E)(A + E)^T \) with \( \frac{1}{m} I \) even \( A = A + E \) is arbitrary matrix. Let

\[ \eta = \|I - \frac{m}{n}(A + E)(A + E)^T\|_2 \]

measure accuracy of the approximating, in which \( \| \cdot \|_2 \) denote the standard operator norm in \( R^n \). For derivation convenient, in this paper, assume \( \eta \) is very small in order to show that the coherence, stable and RIP constant of \( \hat{B} \) are very close to that of \( A \) and \( A \). By convention, the entries of \( A \) are i.i.d. randomly sampling from gaussian distribution with mean zero and variance \( \frac{1}{m} \) that do good to test \( \eta \) is always small.

Another useful formula can be formulated

\[ \eta_0 = \|I - \frac{m}{n} AA^T\|_2 \]

that is introduced in [8]. That \( \eta_0 \) is very small has been proved in [8] with restrictions on \( A \). It's a natural part of our thought process that whether the difference between \( \eta \) and \( \eta_0 \) is very small. **Theorem 1** confirms that our conjecture is correct and further inspire us to think whether the distinct coherence, stable and RIP between \( \hat{B} \) and \( A \), \( A \) are very small. The later related theorems will give us the positive answers.

**Theorem 1.** Assume that sensing matrix \( A \in R^{m \times n} \) and an unknown random matrix \( E \in R^{m \times n} \) with \( m \ll n \). Let

\[ \frac{\|E\|_2}{\|A\|_2} \leq \varepsilon_A \text{ with } 0 < \varepsilon_A < 1, \quad \eta_0 = \|I - \frac{m}{n} AA^T\|_2 \text{ with } 0 < \eta_0 < \frac{1}{2}, \quad \eta = \|I - \frac{m}{n}(A + E)(A + E)^T\|_2, \quad \|A\| = \sigma_1, \quad \sigma_1 \]

is the largest non-zero positive singular value of \( A \), then

\[ \eta_0 - \frac{m}{n}(2\varepsilon_A + \varepsilon_A^2)\sigma_1^2 \leq \eta \leq \eta_0 + \frac{m}{n}(2\varepsilon_A + \varepsilon_A^2)\sigma_1^2 \]

\hspace{1cm} (11)

**Proof.** On the one hand

\[ \eta = \|I - \frac{m}{n}(A + E)(A + E)^T\|_2 \]

\[ = \|I - \frac{m}{n} AA^T - \frac{m}{n}(AE^T + EA^T + EE^T)\|_2 \]

\[ \geq \|I - \frac{m}{n} AA^T\|_2 - \frac{m}{n} \|AE^T\|_2 - \frac{m}{n} \|EA^T\|_2 - \frac{m}{n} \|EE^T\|_2 \]

\[ \geq \eta_0 - \frac{m}{n} \|A\|_2 \|E^T\|_2 - \frac{m}{n} \|E\|_2 \|A^T\|_2 - \frac{m}{n} \|E\|_2 \|E^T\|_2 \]

\[ \geq \eta_0 - \frac{m}{n} \|A\|_2 \varepsilon_A \|A^T\|_2 - \frac{m}{n} \varepsilon_A \|A\|_2 \varepsilon_A \|A^T\|_2 \]

\[ = \eta_0 - \frac{m}{n}(2\varepsilon_A + \varepsilon_A^2) \|A\|_2 \|A^T\|_2 \]

\[ = \eta_0 - \frac{m}{n}(2\varepsilon_A + \varepsilon_A^2)\sigma_1^2 \]

\hspace{1cm} (12)
The last equation holds because \( \| A \|_2 = \sigma_1 \) and \( \| A \|_2 = \| A^T \|_2 \).

On the other hand

\[
\eta = \| I - \frac{m}{n} (A + E)(A + E)^T \|_2 \\
= \| I - \frac{m}{n} AA^T - \frac{m}{n} (AE^T + EA^T + EE^T) \|_2 \\
\leq \| I - \frac{m}{n} AA^T \|_2 + \frac{m}{n} (\| AE^T \|_2 + \| EA^T \|_2 + \| EE^T \|_2) \\
\leq \eta_0 + \frac{m}{n} (\| A \|_2 + \| E \|_2 + \| A \|_2 + \| E \|_2) \\
\leq \eta_0 + \frac{m}{n} (2\| A \|_2 + 2\| E \|_2) \\
= \eta_0 + \frac{m}{n} (2\| A \|_2 + 2\| E \|_2) \\
(13)
\]

The last equation holds because \( \| A \|_2 = \sigma_1 \) and \( \| A \|_2 = \| A^T \|_2 \). Combine (12) with (13) to obtain (11).

**Remark 1.** According to (11), by assumption \( m \ll n \), \( n \to \infty \) then \( \frac{m}{n} \to 0 \), further

\[
\frac{m}{n} (2\| A \|_2 + 2\| E \|_2) \to 0
\]

due to \( 0 < \varepsilon_A \ll 1 \), \( \sigma_1 \) is a positive number. Thus, \( \eta_0 \leq \eta_1 \leq \eta_0 \), that is \( \eta_0 = \eta \) in the case of \( m \ll n \), \( n \to \infty \) according to \( 0 < \varepsilon_A \ll 1 \) and \( \sigma_1 \) is a positive number. **Theorem 1** show that relation between \( \eta \) and \( \eta_0 \) which implies that \( \eta = \eta_0 \) under some special conditions. Therefore, we can let \( \eta < \frac{1}{2} \) like \( \eta_0 < \frac{1}{2} \) in [8].

**Theorem 2.** Assume that sensing matrix \( A \in \mathbb{R}^{m \times n} \) and an unknown random matrix \( E \in \mathbb{R}^{m \times n} \) with \( m \ll n \). Let

\[
\frac{\| \varepsilon_A \|_2}{\| A \|_2} \leq \varepsilon_A
\]

with \( 0 < \varepsilon_A \ll 1 \), \( 0 < \eta < \frac{1}{2} \), \( 0 < \eta_0 < \frac{1}{2} \) in

\[
\eta = \| I - \frac{m}{n} (A + E)(A + E) \|_2, \quad \eta_0 = \| I - \frac{m}{n} AA^T \|_2
\]

and that \( \hat{A} \) satisfies the **RIP of order** \( k \) with \( 0 < \hat{\alpha}_A < \hat{\beta}_A \), \( \sigma_1 \) is the largest singular value of matrix \( A \), that is \( \| A \|_2 = \sigma_1 > 0 \). Then \( \hat{B} \) satisfies the **RIP of order** \( k \) with different constants below:

\[
\hat{\alpha}_k (1 - \mu_1') \| h \|_2 \leq \| \hat{B}_h \|_2 \leq \hat{\beta}_k (1 + \mu_1') \| h \|_2, \quad \mu_1' = \frac{\mu_1}{1 - \mu_1}
\]

\[
\hat{\alpha}_k (1 - \eta_1') \| h \|_2 \leq \| \hat{B}_h \|_2 \leq \hat{\beta}_k (1 + \eta_1') \| h \|_2, \quad \eta_1' = \frac{\eta_1}{1 - \eta_1}
\]

\[
\hat{\alpha}_k (1 - \mu_2') \| h \|_2 \leq \| \hat{B}_h \|_2 \leq \hat{\beta}_k (1 + \mu_2') \| h \|_2, \quad \mu_2' = \frac{\mu_2}{1 - \mu_2}
\]

**Proof.** The proof depends on one fact that \( Q_1 = \frac{Q}{\gamma} \) is close to \( 1 \) due to the definition of \( \eta \). Suppose that

\[
\gamma = \sigma^2 + \frac{n}{m} \sigma_0, Q = \sigma^2 I + \sigma_0^2 (A + E)(A + E)^T
\]

Assume that (5) can be written as \( y = \hat{B}x + v \) where \( \hat{B} = Q_1^{\frac{1}{2}} (A + E) = Q_1^{\frac{1}{2}} \hat{A}, \quad v = Q_1^{\frac{1}{2}} (A + E)u + e \). There are three different results of whitening on \( \hat{A} \) due to the different proving process.

**Case 1.**

\[
\| Q_1 - I \|_2 = \| Q \gamma - I \|_2 = \frac{1}{\gamma} \| Q - \gamma I \|_2 \\
= \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 (A + E)(A + E)^T - \gamma I \|_2 \\
= \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 A^2 + \sigma_0^2 AE^T + \sigma_0^2 EE^2 - \gamma I \|_2 \\
= \frac{\sigma_0^2}{\gamma} \| \frac{m}{n} AA^T + \frac{m}{n} AE^T + \frac{m}{n} EE^T - I \|_2 \\
\leq \frac{\sigma_0^2}{\gamma} \| \frac{m}{n} (\| A \|_2 A^T \|_2 + \| A \|_2 E \|_2 + \| E \|_2) + \| I \|_2) \\
= \frac{\sigma_0^2}{\sigma^2 + \frac{n}{m} \sigma_0} (\frac{m}{n} (\sigma_1^2 + 2\varepsilon_A \sigma_1^2 + \varepsilon_A^2 \sigma_1^2) + 1) \hat{\mu}_1
\]
The last equation holds because \( \| A \|_2 = \sigma_1 \) and \( \| A \|_2 = \| A^T \|_2. \)

From (14), since \( 0 < \varepsilon_A \ll 1, m \ll n, \| A \|_2 = \sigma_1 \) is a positive number, that is \( \frac{m}{n}(\sigma^2_1 + 2\varepsilon_A \sigma^2_1 + \varepsilon^2_A \sigma^2_1) + 1 \to 1 \) when \( n \to \infty \), therefore (14) \( \frac{m}{n}(\sigma^2_0 + \varepsilon \sigma^2_1) < 1 \). That is \( \| Q_1 - I \|_2 < \mu_1 < 1 \) when \( n \to \infty \).

**Case 2.**

\[
\| Q_1 - I \|_2 = \| \frac{Q}{\gamma} - I \|_2 = \frac{1}{\gamma} \| Q - \gamma I \|_2
\]

\[
= \frac{1}{\gamma} \| \sigma^2 I + \sigma^2_0 (A + E)(A + E)^T - \sigma^2 I - \frac{n}{m} \sigma_0 I \|_2
\]

\[
= \frac{n \sigma^2_0}{\sigma^2 + \frac{n}{m} \sigma_0} \| \frac{m}{n} (A + E)(A + E)^T - I \|_2
\]

\[
\leq \frac{n \sigma^2_0}{\sigma^2 + \frac{n}{m} \sigma_0} \eta \Delta \eta_1 < \eta < \frac{1}{2}
\]

(15)

That is case 3.

\[
\| Q_1 - I \|_2 \leq \eta_1 < \eta < \frac{1}{2}
\]

**Case 3.**

\[
\| Q_1 - I \|_2 = \| \frac{Q}{\gamma} - I \|_2 = \frac{1}{\gamma} \| Q - \gamma I \|_2
\]

\[
= \frac{1}{\gamma} \| \sigma^2 I + \sigma^2_0 (A + E)(A + E)^T - \gamma I \|_2
\]

\[
= \frac{1}{\gamma} \| \sigma^2 I + \sigma^2_0 AA^2 + \sigma^2_0 EE^T + \sigma^2_0 EAT + \sigma^2_0 ETA - \sigma^2 I - \frac{n}{m} \sigma_0 I \|_2
\]

\[
\leq \frac{n \sigma^2_0}{\gamma} (\| \frac{m}{n} AA^T \|_2 + \| \frac{m}{n} (A + E)(A + E)^T \|_2 + \| \frac{m}{n} E \|_2 + \| \frac{m}{n} E^T \|_2))
\]

\[
= \frac{n \sigma^2_0}{\gamma} (\eta_0 + \frac{m}{n} (2\varepsilon_A \sigma^2_1 + \varepsilon^2_A \sigma^2_1)) \Delta \mu_2
\]

(16)

The last equation holds because \( \| A \|_2 = \sigma_1 \) and \( \| A \|_2 = \| A^T \|_2. \)

From (16), since \( 0 < \varepsilon_A \ll 1, m \ll n, \| A \|_2 = \sigma_1 \) is positive, \( \frac{m}{n}(2\varepsilon_A \sigma^2_1 + \varepsilon^2_A \sigma^2_1) \to 0 \) when \( n \to \infty \), thus (16) \( \frac{n \sigma^2_0}{\gamma} (\eta_0 + \frac{m}{n} (2\varepsilon_A \sigma^2_1 + \varepsilon^2_A \sigma^2_1)) \Delta \mu_2 \)

As mentioned above, \( \| Q_1 - I \|_2 < 1 \) under some conditions. Using the above three cases (Case 1, Case 2, Case 3), we can obtain three different results, denote by three results **Case 1’, Case 2’, Case 3’** respectively.

**Case 1’.** \( (Q_1^{-1} - I) \) can be expressed as follows

\[
Q_1^{-1} - I = \frac{I - Q_1}{Q_1} = \left( \frac{Q_1}{I - Q_1} \right)^{-1} = \frac{I - Q_1}{I - (1 - Q_1)} = \sum_{k \geq 1} (I - Q_1)^k
\]

that converges from (14) \( \| I - Q_1 \|_2 < \mu_1 < 1 \) where \( \| \cdot \|_2 \) is an operator norm. Take such norm on both sides of the above equality and utilize the triangle inequality to get

\[
\| Q_1^{-1} - I \|_2 = \| \sum_{k \geq 1} (I - Q_1)^k \|_2 \leq \sum_{k \geq 1} \| I - Q_1 \|_2^k \leq \sum_{k \geq 1} \mu_1^k = \frac{\mu_1}{1 - \mu_1} \Delta \mu_1
\]

Let \( \Lambda \) be an index set of size \( k \), \( \forall h \in R^k \),

\[
\| \hat{B}_h h \|_2^2 - \| \hat{A}_h h \|_2^2 = h^T \hat{A}_h^T (Q_1^{-1} - I) \hat{A}_h h
\]

holds.

Since

\[
| h^T \hat{A}_h^T (Q_1^{-1} - I) \hat{A}_h h | \leq \| Q_1^{-1} - I \|_2 \| \hat{A}_h h \|_2 \leq \mu_1 \| \hat{A}_h h \|_2^2 \leq \mu_1 \| \hat{A}_h h \|_2^2
\]

we obtain

\[
\| \hat{B}_h h \|_2^2 - \| \hat{A}_h h \|_2^2 \leq \mu_1 \| \hat{A}_h h \|_2^2
\]

Remove the absolute value to get

\[
(1 - \mu_1) \| \hat{A}_h h \|_2^2 \leq \| \hat{B}_h h \|_2^2 \leq (1 + \mu_1) \| \hat{A}_h h \|_2^2
\]

Due to

\[
\hat{\alpha}_k \| h \|_2 \leq \| \hat{A}_h h \|_2 \leq \hat{\beta}_k \| h \|_2, \forall h \in R^n
\]
we have
\[ \hat{\alpha}_k(1 - \mu_1) \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_1) \left\| h \right\|_2^2, \quad \mu_1 = \frac{\mu_1}{1 - \mu_1} \]

**Case 2’**. It’s easy to see that \((Q_1^{-1} - I)\) converges from (15) \(\left\| I - Q_1 \right\|_2 < \eta_1 < \frac{1}{2}\). Take norm on both sides of \((Q_1^{-1} - I)\) and utilize the triangle inequality to get
\[ \left\| Q_1^{-1} - I \right\|_2 = \left\| \sum_{k \geq 1} (I - Q_1) k \right\|_2 \leq \sum_{k \geq 1} \left\| I - Q_1 \right\|_2^k < \sum_{k \geq 1} \eta_k^k = \frac{\eta_1}{1 - \eta_1} = \eta_1 \]

The remaining proof process of **Case 2’** is the same as that of **Case 1’** except for \(\eta_1'\) instead of \(\mu_1'\) At last we have
\[ \hat{\alpha}_k(1 - \mu_1') \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_1') \left\| h \right\|_2^2, \quad \eta_1' = \frac{\eta_1}{1 - \eta_1} \]

**Case 3’**. It’s easy to see that \((Q_1^{-1} - I)\) converges from (16) due to \(\left\| I - Q_1 \right\|_2 < \mu_2 < \frac{1}{2}\). Take norm on both sides of \((Q_1^{-1} - I)\) and utilize the triangle inequality to get
\[ \left\| Q_1^{-1} - I \right\|_2 = \left\| \sum_{k \geq 1} (I - Q_1) k \right\|_2 \leq \sum_{k \geq 1} \left\| I - Q_1 \right\|_2^k < \sum_{k \geq 1} \mu_2^k = \frac{\mu_2}{1 - \mu_2} \]

The remaining proof process of **Case 3’** is the same as that of **Case 1’** except for \(\mu_2'\) instead of \(\mu_1'\) or that of **Case 2’** except for \(\mu_2'\) instead of \(\eta_1'\). At last, we obtain
\[ \hat{\alpha}_k(1 - \mu_2') \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_2') \left\| h \right\|_2^2, \quad \mu_2' = \frac{\mu_2}{1 - \mu_2} \]

**Remark 2.** In the Theorem 1-2, the condition \(\frac{\| E \|_2}{\| A \|_2} \leq \varepsilon A\) with \(\varepsilon A \ll 1\) can be taken place of \(E = \varepsilon A[9]\), in which \(E\) is a simple version of \(A\), the result are another correct forms theorem. Due to the paper volume, they are omitted here but their proofs are very simple that researchers can prove them and must yield quite perfect results.

The multi-perturbation CS system (6) can be viewed as an generalization of our proposed CS system (5) that the general conclusion of the (6) can come from that of (5). **Theorem 3** and **Theorem 4** give us the results.

**Theorem 3.** Assume that \(A \in R^{m \times n}\) is sensing matrix and \(E_1 \in R^{m \times n}\) is an unknown random matrix with \(m \ll n\). Let
\[ \eta_0 = \left\| \frac{m}{n} A A^T - I \right\|_2, \quad \hat{\eta} = \left\| I - \frac{m}{n} (A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T \right\|_2 \]

\[ \frac{\| E_1 \|_2}{\| A \|_2} \leq \varepsilon A, \sigma_1 \] is the largest singular value of matrix \(A\), \(\| A \| = \sigma_1\), and \(A\) satisfy RIP. An number \(s\) is an integer, \(s < n\), \(0 < \varepsilon A \ll 1\), then the relation of between \(\eta_0\) and \(\eta_0\) can be formulaled as
\[ \eta_0 - \frac{m}{n} (2s\varepsilon A \sigma_1 + s^2 \varepsilon^2 A \sigma_1) \leq \hat{\eta} \leq \eta_0 + \frac{m}{n} (2s\varepsilon A \sigma_1 + s^2 \varepsilon^2 A \sigma_1) \]

The proof of **Theorem 3** can be seen in Appendix.

**Theorem 4.** Assume that \(A \in R^{m \times n}\) is sensing matrix and \(E_1 \in R^{m \times n}\) is an unknown random matrix with \(m \ll n\). Let
\[ \eta_0 = \left\| \frac{m}{n} A A^T - I \right\|_2, \quad \hat{\eta} = \left\| I - \frac{m}{n} (A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T \right\|_2 \]

\[ \frac{\| E_1 \|_2}{\| A \|_2} \leq \varepsilon A, \hat{\eta} < \frac{1}{2}, \sigma_1 \] is the largest singular value of matrix \(A\), \(\| A \| = \sigma_1\), and \(A\) satisfy RIP. Suppose that \(s\) and \(n\) are integers, \(s < n\), \(0 < \varepsilon A \ll 1\). \(Q = \sigma A^2 + \sigma_0^2 (A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T, \gamma = \sigma I + \frac{s\varepsilon A + \sigma_0^2}{m} Q_1 = \frac{Q}{\gamma} \) then
\[ \hat{\alpha}_k(1 - \mu_3) \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_3) \left\| h \right\|_2^2, \quad \mu_3 = \frac{\mu_3}{1 - \mu_3} \]
\[ \hat{\alpha}_k(1 - \mu_3') \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_3') \left\| h \right\|_2^2, \quad \eta_3' = \frac{\eta_3}{1 - \eta_3} \]
\[ \hat{\alpha}_k(1 - \mu_4) \left\| h \right\|_2^2 \leq \left\| \hat{B}_Ah \right\|_2^2 \leq \hat{\beta}_k(1 + \mu_4) \left\| h \right\|_2^2, \quad \mu_4 = \frac{\mu_4}{1 - \mu_4} \]

The proof of **Theorem 4** can be seen in Appendix.

**Remark 3.** Though the bound \(0 < \hat{\eta} < 1\) is sufficient for the proof, the RIC for \(\hat{B}\) is positive in the restriction of \(\hat{\eta} < \frac{1}{2}\).

In the following, we focus on the coherence of \(\hat{B}\) after whitening \(A\) compared to \(A\) and \(\hat{A}\). \(A_i\) is used to denoted the ith column vector of a matrix \(A\).

At first, similar to the coherence of \(A\), the coherence of \(\hat{A} = A + E\) associated with \(A\)'s is first given and we show the coherence of \(\hat{A} = A + E\) related with \(A\)'s.
Definition 3. Let $A$ be a random matrix in CS, $\hat{A} = A + E$, then coherence of $\hat{A}$ can be formulated $\mu(\hat{A})$ as

$$\mu(\hat{A}) = \max_{1 \leq i < j \leq n} \frac{|\hat{A}_{i}^T \hat{A}_{j}|}{\|\hat{A}_{i}\|_2 \|\hat{A}_{j}\|_2}$$

That is, $\mu(\hat{A})$ is the largest absolute inner product between any columns $\hat{A}_{i}, \hat{A}_{j}, i \neq j$.

As mentioned above, $\|Q_{1} - I\|_2 \leq 1$ respectively in some special contexts with $\|E\|_2 \leq \epsilon A$ in Theorem 2. We can take advantage of $\|Q_{1} - I\|_2 < 1$ to prove Theorem 5. For lack of space, we only take $\|Q_{1} - I\|_2 < \mu_{B}, \|Q_{1}^{-1} - I\|_2 < \mu_{B}^{-1}$ as an example with $\eta < \frac{1}{2}$. The proofs of the rest of cases, including $\|Q_{1} - I\|_2 < \mu_{1}, \eta_{1}$, and $\|Q_{1}^{-1} - I\|_2 < \mu_{1}', \eta_{1}'$, is similar to that of Theorem 5, leave them to readers. AS for the general results $\mu(A + \sum_{i=1}^{n} E_{i}) = (A + \sum_{i=1}^{n} E_{i})(x + u) + e$, omit it too due to space constrains. The proving process of general coherence of $\hat{B} = Q_{1}^{-\frac{1}{2}}(A + \sum_{i=1}^{n} E_{i})$ is similar. Theorem 5 demonstrates that the relation between coherence of $\hat{B}$ and that of $\hat{A} = A + E$.

Theorem 5. Assume that $\mu_{2} < \frac{3}{4} \|Q_{1} - I\|_2 \leq \mu_{2}, \hat{B} = Q_{1}^{-\frac{1}{2}}\hat{A}$ with $\hat{A} = A + E$ then

$$\mu(\hat{B}) \leq \frac{(1 + \hat{\mu}_{2})}{(1 - \mu_{2})} \mu(\hat{A})$$

where $\hat{\mu}_{2} = (1 - \mu_{2})^{-\frac{1}{2}} - 1$, $\mu(\hat{B}) = \frac{|\hat{B}^T \hat{B}_{1}|}{\|\hat{B}_{1}\|_2 \|\hat{B}_{2}\|_2}$, $\hat{B}_{j}$ denote the jth column vector of whitening matrix $\hat{B}$, $\hat{A}_{j}$ denotes the jth column vector of $\hat{A}$, that is $\hat{A}_{j} = A_{j} + E_{j}$.

Proof. To prove the theorem, we should find out an upper bound of the numerator $|\hat{B}^T \hat{B}_{j}|$ of $\mu(\hat{B})$ and a lower bound of the denominator $\|\hat{B}_{1}\|_2$. For $i \neq j$ and by assume, we obtain

$$|\hat{B}^T \hat{B}_{j}| = |\hat{A}^T \hat{Q}_{1}^{-1} A_{j}| \leq |\hat{A}^T \hat{A}_{j}| + |\hat{A}^T (Q_{1}^{-1} - I) \hat{A}_{j}| \leq (1 + \hat{\mu}_{2}) |\hat{A}^T \hat{A}_{j}|$$

Next, we estimate lower bound $\|\hat{B}_{i}\|_2$ with restrictions $\|A_{i}\|$ and $\mu_{2}$. Similar to the proof of Theorem 2, $Q_{1}^{-\frac{1}{2}}$ can be expressed as a power series

$$Q_{1}^{-\frac{1}{2}} - I = \sum_{k \geq 1} c_{k} (I - Q_{1})$$

where $c_{k}$ is the coefficients in the Taylor expansion of $(1 - x)^{\frac{1}{2}}$. Both sides of the equality are taken norm obtaining

$$\|Q_{1}^{-\frac{1}{2}} - I\|_2 \leq \sum_{k \geq 1} c_{k} \|Q_{1} - I\|_2 \leq \sum_{k \geq 1} c_{k} \mu_{2} = (1 - \mu_{2})^{-\frac{1}{2}} - 1 \triangleq \hat{\mu}_{2}$$

Thus

$$\|\hat{B}_{i}\|_2 = \|Q_{1}^{-\frac{1}{2}} \hat{A}_{i}\|_2 \geq \|\hat{A}_{i}\|_2 - \|Q_{1}^{-\frac{1}{2}} - I\| \hat{A}_{i}\|_2 \geq (1 - \hat{\mu}_{2}) \|\hat{A}_{i}\|_2$$

where $\hat{\mu}_{2} = (1 - \mu_{2})^{-\frac{1}{2}} - 1$. Combine (17) with (18) to get the result. □

In [9], E is simply version of random matrix A such as $E = \varepsilon A$ with $0 < \varepsilon \ll 1$. The relation between $\mu(\hat{A})$ and $\mu(A)$ will be seen from Theorem 6 below in the case of $E = \varepsilon A$ with $0 < \varepsilon \ll 1$.

Theorem 6. Let $E = \varepsilon A$, $\hat{A} = A + E$, $0 < \varepsilon \ll 1$. The correlation of coherence between A and $\hat{A}$ proceed as

$$\mu(\hat{A}) \leq \frac{(1 + \varepsilon)^2}{(1 - \varepsilon)^2} \mu(A)$$

with $0 < \varepsilon \ll 1$

The proof of Theorem 6 is similar to Theorem 5, here we omitted it.

IV. CONCLUSION

We first propose a new CS system (5) and the completely perturbed model extends previous work by introducing a multiplicative noise $E$ and signal noise $u$ in addition to the usual additive noise $e$. We derived the whiten RIP, whiten coherence for $\hat{A} = A + E$ after whitening (5). Our main contribution show that the RIP, coherence for $\hat{A} = A + E$ was limited by the total noise. As a matter of fact, this paper proves that our proposed completely perturbed CS model (5) equals to a classic CS with only measurement noise. The only difference is the changed measurement matrix by incorporating nontrivial perturbation matrix $E$ to measurement matrix and nontrivial noise $u$ to signal $x$ to induce noise variance increased by a factor of $n/m$ that a tighter upper bound and lower bound of RIP constant is produced. As for coherence of deformed measurement matrix $\hat{A} = A + E$ in this model (5), the constant is nearly invariant essentially with $n/m \rightarrow 0$ of $n, m \rightarrow \infty$. 
Thanks to the features of our proposed CS model (5), there are many works to do. An obvious one is to search one or more optimal algorithms suitable for (5) to recover signal exactly. The related RIP of E in [8] further motivates us to think that E as a sensing perturbation matrix could form a perturbed CS model as \( y_1 = E(x + u) + e \). Thus, (5) may be consist of two similar systems \( y_2 = A(x + u) + e \) and \( y_1 = E(x + u) + e \). Similarly, our model may be divided another two models \( y_3 = (A + E)x + e, y_4 = (A + E)u + e \). Or three three basic parts \( y_1' = Ax + e, y_2' = Au + e, y_3' = E(x + u) + e \). If possible, what we can do to reduce or eliminate the influence of error CS system \( y_3' = E(x + u) + e ? \). Can we recover signal x from error system \( y_3' = E(x + u) + e \). And if can, how to do it? These open problems are worth considering and are to be waited for study in future work.

This paper only do some elementary research on our proposed CS and we hope that the idea and simple study will be helpful and has enlightenment to study and its wide application in the future. We hope A higher level compressed sensing model to be put forward and more and more people explore this areas in CS.

**APPENDIX A**

**PROOF OF THEOREM 3 AND THEOREM 4**

**Theorem 3.** Assume that \( A \in \mathbb{R}^{m \times n} \) is sensing matrix and \( E_i \in \mathbb{R}^{m \times n} \) is an unknown random matrix with \( m \ll n, \|E_i\|_2 \leq \varepsilon_A \). Let

\[
\eta_0 = \| \frac{m}{n} AA^T - I \|_2, \quad \tilde{\eta} = \| I - \frac{m}{n} (A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T \|_2
\]

\( \sigma_1 \) is the largest singular value of matrix \( A \), that is \( \|A\|_2 = \sigma_1 \). An number \( s \) is an integer, \( s < n, 0 < \varepsilon_A \ll 1 \). Then the relation of between \( \tilde{\eta} \) and \( \eta_0 \) can be formulated as

\[
\eta_0 - \frac{m}{n}(2\varepsilon_A \sigma_1 + s^2 \varepsilon_A^2 \sigma_1) \leq \tilde{\eta} \leq \eta_0 + \frac{m}{n}(2\varepsilon_A \sigma_1 + s^2 \varepsilon_A^2 \sigma_1)
\]

(19)

Proof: On the one hand

\[
\tilde{\eta} = \| I - \frac{m}{n}(A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T \|_2
\]

\[
= \| \frac{m}{n} AA^T - I + \frac{m}{n}(A \sum_{i=1}^{s} E_i^T + A^T \sum_{i=1}^{s} E_i + (\sum_{i=1}^{s} E_i)(\sum_{i=1}^{s} E_i)^T \|_2
\]

\[
\geq \| \frac{m}{n} AA^T - I \|_2 - \frac{m}{n} (\| A \|_2 \sum_{i=1}^{s} \| E_i^T \|_2 + \| A^T \|_2 \sum_{i=1}^{s} \| E_i \|_2 + (\sum_{i=1}^{s} \| E_i \|_2)(\sum_{i=1}^{s} \| E_i^T \|_2))
\]

\[
\geq \eta_0 - \frac{m}{n}(\| A \|_2 \varepsilon_A \sum_{i=1}^{s} \| A \|_2 + \| A^T \|_2 \sum_{i=1}^{s} \| \varepsilon_A \| A \|_2 + (\sum_{i=1}^{s} \| \varepsilon_A \| A \|_2)(\sum_{i=1}^{s} \| A^T \|_2)
\]

\[
= \eta_0 - \frac{m}{n}(2s\varepsilon_A \sigma_1^2 + s^2 \varepsilon_A^2 \sigma_1^2)
\]

(20)

The last equation holds because \( \| A^T \|_2 = \| A \|_2 = \sigma_1 \).

On the other hand

\[
\tilde{\eta} = \| I - \frac{m}{n}(A + \sum_{i=1}^{s} E_i)(A + \sum_{i=1}^{s} E_i)^T \|_2
\]

\[
= \| \frac{m}{n} AA^T - I + \frac{m}{n}(A \sum_{i=1}^{s} E_i^T + A^T \sum_{i=1}^{s} E_i + (\sum_{i=1}^{s} E_i)(\sum_{i=1}^{s} E_i)^T \|_2
\]

\[
\leq \| \frac{m}{n} AA^T - I \|_2 + \frac{m}{n} (\| A \|_2 \sum_{i=1}^{s} \| E_i^T \|_2 + \| A^T \|_2 \sum_{i=1}^{s} \| E_i \|_2 + (\sum_{i=1}^{s} \| E_i \|_2)(\sum_{i=1}^{s} \| E_i^T \|_2))
\]

\[
\leq \eta_0 + \frac{m}{n}(\| A \|_2 \varepsilon_A \sum_{i=1}^{s} \| A \|_2 + \| A^T \|_2 \sum_{i=1}^{s} \| \varepsilon_A \| A \|_2 + (\sum_{i=1}^{s} \| \varepsilon_A \| A \|_2)(\sum_{i=1}^{s} \| A^T \|_2)
\]

\[
= \eta_0 + \frac{m}{n}(2s\varepsilon_A \sigma_1^2 + s^2 \varepsilon_A^2 \sigma_1^2)
\]

(21)
The last equation holds because $\|A^T\|_2 = \|A\|_2 = \sigma_1$ As mentioned above, combine (20) with (21) to get (19), the conclusion is obtained. □

**Remark 4:** We can see (19):

$$\eta_0 - \frac{m}{n}(2s\epsilon_A\sigma_1 + s^2\epsilon_A^2\sigma_1) \leq \tilde{\eta} \leq \eta_0 + \frac{m}{n}(2s\epsilon_A\sigma_1 + s^2\epsilon_A^2\sigma_1)$$

Since $1 \leq s \ll n, m \ll n, s, m, n$ are positive integers, $\sigma_1$ is a constant, then $\frac{m}{n}(2s\epsilon_A\sigma_1^2 + s^2\epsilon_A^2\sigma_1^2) \rightarrow 0$ when $n \rightarrow \infty$ so that implies $\tilde{\eta} \rightarrow \eta_0$.

**Theorem 4.** Assume that $A \in R^{m \times n}$ is sensing matrix and $E_i \in R^{m \times n}$ is an unknown random matrix with $m \ll n$. Let

$$\eta_0 = \frac{m}{n}AA^T - I \|_2, \quad \tilde{\eta} = \| I - \frac{m}{n}(A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T \|_2$$

$$\|E_i\|_2 \leq \frac{\|E_i\|_2}{\|A\|_2} \leq \epsilon_A \tilde{\eta} \leq \frac{1}{2}\sigma_1$$ is the largest singular value of matrix $A$, that is $\|A\|_2 = \sigma_1$. Suppose that $s$ and $n$ are integers, $s < n, 0 < \epsilon_A \ll 1$. $Q = \sigma^2 I + \sigma^2_0(A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T, \gamma = \sigma I + \frac{m}{n}\sigma^2_0, Q_1 = \frac{Q}{\gamma}$, then

$$\hat{\alpha}_k(1 - \mu^2_k) \| h \|_2 \leq \hat{B}_k h \|_2 \leq \hat{\beta}_k(1 + \mu^2_k) \| h \|_2, \quad \mu^2_k = \frac{1}{1 - \mu^2_k}$$

$$\hat{\alpha}_k(1 - \eta^2_k) \| h \|_2 \leq \hat{B}_k h \|_2 \leq \hat{\beta}_k(1 + \eta^2_k) \| h \|_2, \quad \eta^2_k = \frac{\eta^2_k}{1 - \eta^2_k}$$

$$\hat{\alpha}_k(1 - \mu^2_k) \| h \|_2 \leq \hat{B}_k h \|_2 \leq \hat{\beta}_k(1 + \mu^2_k) \| h \|_2, \quad \mu^2_k = \frac{\mu^2_k}{1 - \mu^2_k}$$

**Remark 5.** Though the bound $0 < \tilde{\eta} < 1$ is sufficient for the proof, the RIC for $\hat{B}$ is positive in the restriction of $\tilde{\eta} < \frac{1}{2}$.

**Proof.** The proof depend on one fact that $Q_1$ is close to $I$ due to the definition of $\tilde{\eta}$. Assume that (13) can be written as $y = \hat{B}x + w$ where $\hat{B} = Q_1^{\frac{1}{2}}(A + \sum_{i=1}^s E_i)$ and $w = Q_1^{\frac{1}{2}}((A + \sum_{i=1}^s E_i)u + e)$. There are three different results of whitening on $A + \sum_{i=1}^s E_i$ due to the different proving process.

**Case 1.**

$$\| Q_1 - I \|_2 = \| \frac{Q}{\gamma} - I \|_2$$

$$= \frac{1}{\gamma} \| \sigma^2 I + \sigma^2_0(A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - \gamma I \|_2$$

$$= \frac{1}{\gamma} \| \sigma^2 I + \sigma^2_0(A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T(\sigma^2 + \frac{m}{n}\sigma^2_0)I \|_2$$

$$= \frac{n}{\gamma} \| \frac{m}{n}(A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - I \|_2$$

$$= \frac{n}{\gamma} \| \frac{m}{n}AA^T - I + \frac{m}{n}(A \sum_{i=1}^s E_i^T + A^T \sum_{i=1}^s E_i + \sum_{i=1}^s E_i)(\sum_{i=1}^s E_i^T) \|_2$$

$$\leq \frac{n}{\gamma} \| \frac{m}{n}AA^T - I \|_2 + \frac{m}{n}(\| A \|_2 \sum_{i=1}^s || E_i^T \|_2 + \| A^T \|_2 \sum_{i=1}^s || E_i \|_2 + \sum_{i=1}^s || E_i \|_2(\sum_{i=1}^s || E_i^T \|_2))$$

$$\leq \frac{n}{\gamma} \| \frac{m}{n}AA^T - I \|_2 + \frac{m}{n}(\| A \|_2 \sum_{i=1}^s \| E_i^T \|_2 + \| A^T \|_2 \sum_{i=1}^s \| E_i \|_2 + \sum_{i=1}^s \| E_i \|_2(\sum_{i=1}^s \| E_i^T \|_2))$$

$$= \frac{n}{\gamma} \| \frac{m}{n}E_A \sum_{i=1}^s \sigma_1^2 + \sum_{i=1}^s \sigma_1^2 + s^2\epsilon_A^2\sigma_1^2)$$

$$= \frac{n}{\gamma} \| \frac{m}{n}(2s\epsilon_A\sigma_1^2 + s^2\epsilon_A^2\sigma_1^2) \| \mu_3$$

(22)

The last equation holds because $\| A \|_2 = \sigma_1$ and $\| A \|_2 = \| A^T \|_2$.

From (22), since $0 < \epsilon_A \ll 1, m \ll n, \| A \|_2 = \sigma_1$ is a constant, that is

$$\frac{m}{n}(2s\epsilon_A\sigma_1^2 + s^2\epsilon_A^2\sigma_1^2) \rightarrow 0$$

when $n \rightarrow \infty$, therefore (22) $\rightarrow \frac{n}{\gamma} \| \frac{m}{n}(2s\epsilon_A\sigma_1^2 + s^2\epsilon_A^2\sigma_1^2) \| \mu_3 \leq \frac{1}{2}$. That is $\| Q_1 - I \|_2 < \mu_3 < \frac{1}{2}$ when $n \rightarrow \infty$. 

Case 2.

\[ \| Q_1 - I \|_2 = \| \frac{Q}{\gamma} - I \|_2 = \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - \gamma I \|_2 \]

\[ = \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - (\sigma^2 + \frac{n}{m} \sigma_0^2) I \|_2 \]

\[ = \frac{m \sigma_0^2}{\gamma} \| \frac{m}{n} (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - I \|_2 \]

\[ = \frac{m \sigma_0^2}{\sigma^2 + \frac{m}{n} \sigma_0^2} \tilde{\eta} \triangleq \eta _2 < \tilde{\eta} < \frac{1}{2} \]  \hspace{1cm} (23)

From (23) we get \( \| Q_1 - I \|_2 < 1 \)

Case 3.

\[ \| Q_1 - I \|_2 = \| \frac{Q}{\gamma} - I \|_2 = \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - \gamma I \|_2 \]

\[ = \frac{1}{\gamma} \| \sigma^2 I + \sigma_0^2 (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - (\sigma^2 + \frac{n}{m} \sigma_0^2) I \|_2 \]

\[ = \frac{m \sigma_0^2}{\gamma} \| \frac{m}{n} (A + \sum_{i=1}^s E_i)(A + \sum_{i=1}^s E_i)^T - I \|_2 \]

\[ \leq \frac{m \sigma_0^2}{\gamma} (\| - I \|_2 + \frac{m}{n} (\| A A^T \|_2 + \| A \|_2) \sum_{i=1}^s \| E_i \|_2 + \| A \|_2 \sum_{i=1}^s \| A T \|_2 + \| A \|_2 \sum_{i=1}^s \| \varepsilon A \| A \|_2 + \sum_{i=1}^s \| \varepsilon A \| A \|_2) \]

\[ + (\sum_{i=1}^s \varepsilon A \| A \|_2) (\sum_{i=1}^s \varepsilon A \| A \|_2) \]

\[ = \frac{m \sigma_0^2}{\gamma} (1 + \frac{m}{n} (\sigma_1^2 + \varepsilon A \sum_{i=1}^s \sigma_i^2 + \varepsilon_A \sum_{i=1}^s \sigma_i^2 + 2 \varepsilon^2 \varepsilon_A \sigma^2) \]

\[ = \frac{m \sigma_0^2}{\sigma^2 + \frac{m}{n} \sigma_0^2} (1 + \frac{m}{n} (\sigma_1^2 + 2 \varepsilon A \sigma_1^2 + s^2 \varepsilon^2 \varepsilon_A \sigma^2)) \triangleq \mu_4 \]  \hspace{1cm} (24)

The last equation holds because \( \| A \|_2 = \sigma_1 \) and \( \| A \|_2 = \| A^T \|_2 \).

From (24), since \( 0 < \varepsilon A < 1, m \ll n, \| A \|_2 = \sigma_1 \) is a constant, that is

\[ \frac{m}{n} (\sigma_1^2 + 2 \varepsilon A \sigma_1^2 + s^2 \varepsilon^2 \varepsilon_A \sigma^2) \rightarrow 0 \]

when \( n \rightarrow \infty \), thus \( (24) \rightarrow \frac{m \sigma_0^2}{\sigma^2 + \frac{m}{n} \sigma_0^2} < 1 \). That is \( \| Q_1 - I \|_2 < \mu_4 < 1 \). As mentioned above, \( \| Q_1 - I \|_2 < 1 \) under some condition. Using the above three cases (Case 1, Case 2, Case 3) as the condition, we can obtain three different results, denoted by Case 1’, Case 2’, Case 3’ respectively.

Case 1’. \( (Q_1^{-1} - I) \) can be expressed as follows

\[ Q_1^{-1} - I = I - Q_1 \]

\[ = \frac{I - Q_1}{Q_1} \]

\[ = \frac{I - Q_1}{(I - Q_1)^{-1}} \]

\[ = \sum_{k \geq 1}^\infty (I - Q_1)^k \]  \hspace{1cm} (25)

that converges from case 1 \( \| I - Q_1 \|_2 < \mu_3 < 1 \) where \( \| \cdot \|_2 \) is an operator norm. Take such norm on both side of the equality and utilize the triangle inequality to get

\[ \| Q_1^{-1} - I \|_2 = \| \sum_{k \geq 1} (I - Q_1)^k \|_2 \leq \sum_{k \geq 1} \| I - Q_1 \|_2^k \leq \sum_{k \geq 1} \mu_3^k = \frac{\mu_3}{1 - \mu_3} \triangleq \mu_3 \]

Let \( \Lambda \) be an index set of size \( k \), we have

\[ \| \tilde{B}_\Lambda h \|_2^2 = \| \tilde{A}_\Lambda h \|_2^2 = h^T \tilde{A}_\Lambda^T (Q_1^{-1} - I) \tilde{A}_\Lambda h, \quad \forall h \in \mathbb{R}^k \]
Since
\[ h^T \hat{A}_A^T (Q_1^{-1} - I) \hat{A}_A h \leq \| Q_1^{-1} - I \|_2 \| \hat{A}_A h \|_2^2 \leq \mu_3 \| \hat{A}_A h \|_2^2 \]
we obtain that
\[ \| \hat{B}_A h \|_2^2 - \| \hat{A}_A h \|_2^2 \leq \mu_3 \| \hat{A}_A h \|_2^2 \]
Remove the absolute value to get
\[ (1 - \mu_3^2) \| \hat{A}_A h \|_2^2 \leq \| \hat{B}_A h \|_2^2 \leq (1 + \mu_3^2) \| \hat{A}_A h \|_2^2 \]
Due to
\[ \hat{\alpha}_k \| h \|_2^2 \leq \| \hat{A} h \|_2^2 \leq \hat{\beta}_k \| h \|_2^2, \forall h \in R^n \]
we have
\[ \hat{\alpha}_k (1 - \mu_3^2) \| h \|_2^2 \leq \| \hat{B}_A h \|_2^2 \leq \hat{\beta}_k (1 + \mu_3^2) \| h \|_2^2, \mu_3 = \frac{\mu_3}{1 - \mu_3} \]

\textbf{Case 2}. Note that (25) converges since \( \| I - Q_1 \|_2 \leq \eta_2 < 1 \) from case 2 where \( \| \cdot \|_2 \) is an operator norm. Take such norm on both side of the equality (25) and utilize the triangle inequality to get
\[ \| Q_1^{-1} - I \|_2 = \| \sum_{k \geq 1} (I - Q_1)^k \|_2 \leq \sum_{k \geq 1} \| I - Q_1 \|_2^k \leq \sum_{k \geq 1} \eta_2^k = \frac{\eta_2}{1 - \eta_2} \leq \eta_2' \]
The remaining proving is the same as that of case 1' except for \( \eta_2' \) instead of \( \mu_3^2 \). At last we have
\[ \hat{\alpha}_k (1 - \eta_2^2) \| h \|_2^2 \leq \| \hat{B}_A h \|_2^2 \leq \hat{\beta}_k (1 + \eta_2^2) \| h \|_2^2, \eta_2' = \frac{\eta_2}{1 - \eta_2} \]

\textbf{Case 3}. Note that (25) converges since \( \| I - Q_1 \|_2 \leq \mu_4 < 1 \) from case 3, where \( \| \cdot \|_2 \) is an operator norm. Take such norm on both side of the equality (25) and utilize the triangle inequality to get
\[ \| Q_1^{-1} - I \|_2 = \| \sum_{k \geq 1} (I - Q_1)^k \|_2 \leq \sum_{k \geq 1} \| I - Q_1 \|_2^k \leq \sum_{k \geq 1} \mu_4^k = \frac{\mu_4}{1 - \mu_4} \leq \mu_4' \]
The remaining proving is the same as that of case 1' except for \( \mu_4' \) instead of \( \mu_3^2 \). At last we have
\[ \hat{\alpha}_k (1 - \mu_4^2) \| h \|_2^2 \leq \| \hat{B}_A h \|_2^2 \leq \hat{\beta}_k (1 + \mu_4^2) \| h \|_2^2, \mu_4' = \frac{\mu_4}{1 - \mu_4} \]

\textbf{ACKNOWLEDGMENT}

This work is partially supported by the National Natural Science Foundation of China (NSFC) A3 Foresight Program (No. 61411146001).

\textbf{REFERENCES}

[1] Li S, Gao F, Ge G, et al. Deterministic construction of compressed sensing matrices via algebraic curves[J]. Information Theory, IEEE Transactions on, 2012, 58(8): 5035-5041.
[2] Wang X, Zhao Z, Zhao N, et al. On the application of compressed sensing in communication networks[C].Communications and Networking in China (CHINACOM), 2010 5th International ICST Conference on. IEEE, 2010: 1-7.
[3] Malioutov D, Malloyev M. Boolean compressed sensing; LP relaxation for group testing[C].Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012: 3305-3308.
[4] Needell D, Vershynin R. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit[J]. Selected Topics in Signal Processing, IEEE Journal of, 2010, 4(2): 310-316.
[5] Donoho D L, Eldad M, Temlyakov V N. Stable recovery of sparse overcomplete representations in the presence of noise[J]. Information Theory, IEEE Transactions on, 2006, 52(1): 6-18.
[6] Candès E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. Communications on pure and applied mathematics, 2006, 59(8): 1207-1223.
[7] Candès E J, Tao T. Decoding by linear programming[J]. Information Theory, IEEE Transactions on, 2005, 51(12): 4203-4215.
[8] Arias-Castro E, Elad Y C. Noise folding in compressed sensing[J]. Signal Processing Letters, IEEE, 2011, 18(8): 478-481.
[9] Herman M A, Strohmer T. General deviants: An analysis of perturbations in compressed sensing[J]. Selected Topics in Signal Processing, IEEE Journal of, 2010, 4(2): 342-349.
[10] Ding J, Chen L, Gu Y. Performance analysis of orthogonal matching pursuit under general perturbations[C].Computing, Networking and Communications (ICNC), 2012 International Conference on. IEEE, 2012: 892-896.
[11] Ding J, Chen L, Gu Y. Perturbation analysis of orthogonal matching pursuit[J]. Signal Processing, IEEE Transactions on, 2013, 61(2): 398-410.
[12] Davenport M A, Duarte M F, Eldar Y C, et al. Introduction to compressed sensing[J]. Preprint, 2011, 93.
[13] E.J.Candes, J.Romberg,andT.Tao,Robust uncertainty principles: Exact signal reconstruction from highly incomplete requency information, IEEETrans.Inform.Theory,vol.52,no.2,pp.489-509, Feb.2006.
[14] Donoho D L. Compressed sensing[J]. Information Theory, IEEE Transactions on, 2006, 52(4): 1289-1306.
[15] Baraniuk R, Steeghs P. Compressive radar imaging[C]. Radar Conference, 2007 IEEE. IEEE, 2007: 128-133.
[16] Duarte M F, Davenport M A, Wakin M B, et al. Sparse signal detection from incoherent projections[C]. Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. IEEE, 2006, 3: III-III.

[17] Romberg J. Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming][J]. IEEE Signal Processing Magazine, 2008, 25(2): 14-20.

[18] Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit[J]. Information Theory, IEEE Transactions on, 2007, 53(12): 4655-4666.

[19] Needell D, Tropp J A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples[J]. Applied and Computational Harmonic Analysis, 2009, 26(3): 301-321.

[20] Foucart S, Rauhut H. A mathematical introduction to compressive sensing[M]. Berlin: Springer, 2013.

[21] Cai T T, Wang L, Xu G. New bounds for restricted isometry constants[J]. Information Theory, IEEE Transactions on, 2010, 56(9): 4388-4394.

[22] Davenport M A, Wakin M B. Analysis of orthogonal matching pursuit using the restricted isometry property[J]. Information Theory, IEEE Transactions on, 2010, 56(9): 4395-4401.

[23] S. Huang and J. Zhu. “Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP.” Inverse Problems, vol. 27, no. 3, Mar. 2011.

[24] A.K. Fletcher and S. Rangan, Orthogonal Matching Pursuit: A Brownian Motion Analysis, IEEE Trans. Signal Processing, vol. 60, no. 3, Mar. 2012, pp. 1010-1021.

[25] Mo Q, Shen Y. A remark on the restricted isometry property in orthogonal matching pursuit[J]. Information Theory, IEEE Transactions on, 2012, 58(6): 3654-3656.